

School of Computing

Adaptive Behavioural Cognition

Garry Briscoe

**“This thesis is presented as part of the requirements for
the award of the Degree of Doctor of Philosophy
of the
Curtin University of Technology”**

September 1997

Acknowledgements

The task of acknowledging all those who have helped in the completion of this thesis would require many pages. I can but thank them all collectively. However, a number of individuals merit special attention.

First and foremost, my supervisor, Terry Caelli. One would be hard pressed to imagine a better supervisor. While initially displaying slight skepticism of my broad ranging investigations, as well as some difference in basic epistemology, Terry has nevertheless been extremely encouraging and has allowed me the freedom to explore. His enthusiasm and drive are infectious, and he has collected about him a research team of loyal and committed students, of which I am proud to have been a member. Terry is also open to new ideas, and our debates and discussions were both stimulating and illuminating. He has also been a great friend, providing support of many kinds. I have greatly benefited from the experience, and I thank him sincerely.

Many thanks must also go to my second supervisor, Walter Bischof. His cheerfulness, encouragement and sound advice have been greatly appreciated.

The other members of Terry's team of students and researchers at Curtin University must also be thanked for their stimulating and enthusiastic discussions, both formally and as part of our frequent visits to the local coffee shop. There has been a real sense of academic community, with vigorous debate in a supportive environment. Special thanks must go to Adrian Pearce who has been both a good friend and colleague, and who has contributed greatly to a warm and joyful working environment.

I must also thank Beth for being so understanding of my heavy commitment to this thesis, and my son William for accepting a substantial drop in conditions while I returned to study after many years.

This work has been supported by a Priority Australian Postgraduate Research Award and an Australian Research Council Grant.

Abstract

Cognitive Science is at a crossroad. Since its inception, the prevailing paradigm in Cognitive Science (and associated fields such as Artificial Intelligence, Cognitive Psychology, and Linguistics), has been a formal, computer-based model of cognition—often termed the Symbol Processing System model (SPS) or *cognitivism*. This view, while still accepted by the majority of researchers, has been dogged by persistent and cutting criticism by various authors over many years. As well, the initial over-inflated promises made by the early practitioners within these fields have not come to fruition, and the initial enthusiasm has in many cases been reduced to frustration.

Many researchers have looked to the field of *connectionism* as a solution, and this discipline has found a new lease of life after a serious setback in the early 1970s. The major emphasis within this area has been on feed-forward neural networks (FFNN), but this paradigm also has its detractors.

In this thesis we critically evaluate both of these research programs, especially that of SPS. We propose a new model of human and animal cognition, termed *Adaptive Behavioral Cognition* (ABC), which integrates many current views on cognition, and provides a single-architecture, biologically-feasible theory that overcomes many of the problems associated with current models. As well as being an accurate description of the processes relevant to the new model, the term ABC is a none-too-subtle reference to the fact that we need to closely re-examine the aims and achievements of Cognitive Science and return to basic empirical findings in developing a theory of cognition.

The new model synthesises, unifies and links together many previously disjoint ideas and observations, from the neural level through to neurological structures and to observed behaviour. The claims that we make of the model are that it is biologically and neurologically consistent and reasonable, and that it has properties more closely associated with the actual brain than either the computational (cognitivist) approach, or the simplistic FFNN. Further, the model is internally consistent and self-similar, and is consistent with the observed neuroanatomical structures of the cortex. It also provides for massive parallelism, yet retains a serial component through its use of temporal

sequences.

The ABC proposal outlined in this thesis takes the view that the processes of the brain are to learn associated and temporally connected sequences, rather than ‘facts’ or ‘representations’, and that the learned behaviours resulting from the associated temporal sequences are the means of cognition, rather than computational operations on representations.

Contents

1	Introduction	1
1.1	Overview	1
1.1.1	Conceptualisation, Action, and Language	3
1.2	Chapter Layout	4
1.3	Current Theories of Cognition	6
1.3.1	Behaviourism	6
1.3.2	Cognitivism	8
1.4	Degenerating Research Programme	13
1.5	Connectionism	14
1.6	Summary	16
2	Learning and Generating Temporal Sequences	18
2.1	Introduction and Caveats	18
2.2	Temporal Learning	19
2.2.1	Previous Temporal Learning Results	20
2.3	Temporal Learning Model	30
2.3.1	Temporal Learning Algorithms	35

2.3.2	Mathematical Description of the Network	38
2.3.3	Kohonen Algorithm—Winner-Take-All	44
2.3.4	Extensions to Hebbian Learning	46
2.3.5	Differential Learning Periods	46
2.3.6	Refractory Period and Winner Exclusion	47
2.3.7	Input Data	48
2.4	Temporal Learning Experiments	50
2.4.1	Summary of Temporal Learning Results	50
2.4.2	Learning the Alphabet	52
2.4.3	Motor Control—Sequence Production	55
2.4.4	LAPS—Learning And Production of Sequences	58
2.4.5	Semantic Priming	60
2.5	Dynamic LAPS Structure—Predictive Knowledge	70
2.6	SOMA and LAPS Networks as Modules	73
2.7	Discussion of Temporal Learning Results	76
3	Adaptive Behavioral Cognition (ABC)—The Model	78
3.1	Introduction	78
3.2	Building the ABC Model	79
3.2.1	Sense Data and World Filters	80
3.2.2	Self-Organising Maps	82
3.2.3	Recurrent Loops	82
3.2.4	Intra-Modal and Inter-Modal Associative Layers	83

3.2.5	Equivalence of Representations	85
3.2.6	Model of Visual Cognition	86
3.2.7	Multi-modal Associations and Motor Control	88
3.2.8	Overall Model	92
3.3	Discussion of the ABC Model	95
3.3.1	Cross-Modal Linkages and Associationism	96
3.3.2	Neural-Based Not Symbol-Based	97
3.3.3	Recurrent Feedback	97
3.3.4	Sense Data and Filters—Sampling of the World	98
3.3.5	Perceptual Conceptualisation, Labelling, Language and Self-Talk	100
3.3.6	Motor Sequences	126
3.3.7	Parallel versus Serial	129
3.3.8	Self-Organisation	130
3.3.9	Hardware, Representations and Self-Organisation	132
3.3.10	Categorical Perception	136
3.3.11	“Soft-Wired” Reflex	140
3.3.12	Learning	142
3.3.13	Context	145
3.3.14	Common Sense	146
3.3.15	Errors and Unsupervised Learning	147
3.3.16	Representations	148
3.3.17	Continuity Across Phylogenetic Tree	149
3.3.18	Vector Concatenation	150

3.3.19	Topographical Connectionist Approach	151
3.4	Addresses Problems with Alternate Approaches	152
3.5	What Is Left Out?	164
3.6	Summary	167
4	Vision	168
4.1	Introduction	168
4.1.1	Basic Physiology and Anatomy of the Visual System	169
4.1.2	Eye Movements	177
4.2	Biological Considerations	181
4.2.1	Hierarchical Processing and Priming Effects	181
4.2.2	Adaption to Discordant Stimulation	183
4.2.3	Invariance	185
4.2.4	Saccades, Fixation and Spatial Fusion	186
4.2.5	The World As Outside Memory	190
4.2.6	Semantic Integration	196
4.2.7	Fill-in of Blind Spot and Scotomas	197
4.2.8	Global Influences on Perception	198
4.2.9	Neuroanatomy and Neurophysiology	202
4.3	Evolutionary Considerations	204
4.4	Active Vision	205
4.5	Scanpaths	208
4.5.1	Visual Imagery Scanpaths	212

4.6	ABC Temporal Experiments and Vision	213
4.7	ABC Model Avoids Problems of Traditional Models	224
4.7.1	Recognition	226
4.8	Further Discussion and Implications of the Model	227
4.8.1	Relational Learning	227
4.8.2	World Filters	228
4.8.3	Receptive Fields in Periphery	229
4.8.4	Treisman Figures	230
4.8.5	Integration of Local Features	232
4.8.6	Polysemous Figures	233
4.8.7	Reinforcement and Learning	233
4.8.8	Stabilised Retinal Images	236
4.8.9	Reading	236
4.9	Conclusion	238
5	Further Discussion and Implications	240
5.1	Language	240
5.1.1	Current Theories of Language	242
5.1.2	Chomsky and Language	244
5.1.3	Formal Learning Theory	253
5.1.4	Learning in Neural Networks—Theory	255
5.1.5	Statistical Language Learning	257
5.1.6	ABC and Language	258

5.2	Implicit and Explicit Learning	270
5.2.1	The Two-Store Model of Memory	273
5.2.2	Implicit Awareness and Brain Damage	274
5.2.3	Split-Brain Patients and Hemispheric Specialisation	289
5.2.4	Skill Acquisition, Language and Thinking	292
5.2.5	Automaticity	294
5.2.6	Knowing-How and Knowing-That	296
5.2.7	Perceptual Skills and Embodiment	297
5.2.8	The Use of Tools	299
5.3	Consciousness	301
5.3.1	Definitions of Consciousness	302
5.3.2	Limits to Conscious Experiences	305
5.3.3	Introspection and Verbal Reports	306
5.3.4	Conscious, Nonconscious and Unconscious	309
5.3.5	Conscious Awareness of Voluntary Acts	312
5.3.6	Sports Psychology	315
5.3.7	Problem Solving	317
5.3.8	Expert Knowledge	318
5.3.9	Other Models of Consciousness	318
5.3.10	ABC and Consciousness	319
5.3.11	Internal 'Feels' and Qualia	323
5.4	Common Sense and Everyday Cognition	325
5.4.1	Biology of Learning	335

5.4.2	The Notion of Theories	336
5.5	Comparison with Other Models and Related Work	340
5.5.1	Situated Cognition and Situated Action	340
5.5.2	Neural Darwinism	341
6	Summary and Ongoing Research	343
6.1	Future Research	344
6.2	Conclusion	345
A	Neuroanatomy, Neurophysiology and Neuropsychology	347
A.1	Introduction	347
A.2	Evolution of the Neocortex in Mammals	349
A.3	Thalamus	350
A.4	Topological Organisation of the Cortex	353
A.4.1	Plasticity and Self-Organisation	355
A.5	Association Cortices	357
A.5.1	Auditory Association Areas	359
A.5.2	Somatosensory Association Areas	360
A.5.3	Visual Association Areas	362
A.5.4	Frontal Association Areas	363
A.5.5	Multimodal Association Areas	365
A.5.6	Paralimbic Association Areas	368
A.5.7	Summary of Sensory Association Areas	369
A.6	Prefrontal Cortex	371

A.7	Motor Areas	377
A.8	Language Areas	378
A.9	Binding of Modalities	381
A.10	Discussion In Relation To the ABC Model	383
B	Temporal Learning Experiments in Detail	384
B.1	Temporal Learning—Finite State Automata	384
B.1.1	Stopwatch Example	385
B.1.2	Number Example	390
B.1.3	Bidirectional Link Example	392
B.1.4	Reber Example	394
B.1.5	Actions	395
B.1.6	Soft Programming	396
B.1.7	Discussion—FSAMap	396
B.2	Full Temporal Model—Experimental Results	397
B.2.1	Temporal XOR	397
B.2.2	Elman’s Consonant/Vowel Sequences	400
B.2.3	Bidirectional Link Revisited	403
B.2.4	Reber FSA Revisited	418
B.2.5	Counting and Memory	422
B.2.6	Embedded Sequences	427
B.2.7	English Sentences	435
B.2.8	Other Language Sentences	438

B.2.9	Embedded English Sentences	439
B.3	Hierarchical Temporal Sequences	462
B.3.1	Hierarchical FSAMaps	462
B.3.2	Hierarchical Temporal Sequences—SOMA and LAPS	480
B.4	Miscellaneous	480
B.4.1	Testing the Learning of the Alphabet	480
B.4.2	Learning Characters and Recognizing Words	484
B.4.3	Spelling	488
B.4.4	Computer Language	490
B.4.5	Loop	491
C	Implementation of the Full Simulation Model	493
C.1	Introduction	493
C.2	The Challenge	494
C.3	The Full Simulation Model	496
C.3.1	Visual Model—Receptive Fields	496
C.3.2	Visual Model—Wilson Modified Line-Element Model	497
C.3.3	Visual Model—Colour and Motion Filters	503
C.4	Simulation Methods	504
C.4.1	Data Flow Design	505
C.4.2	Action Script	506
C.4.3	GUI Windows	509
D	Computers and Computation	515

D.1	Binary (von Neuman) Computers	516
D.2	Representations	518
D.3	Rules and Computers	521
D.4	Assumptions of Cognitivism	523
D.5	Computation by Humans	531
D.6	Artificial Intelligence	533
D.7	ABC is Hardware-Only	534
E	Vision Background	535
E.1	Current Theories of Visual Recognition	535
E.1.1	Template Matching	537
E.1.2	Feature Models	539
E.1.3	Fourier Models	540
E.1.4	Structural Models	543
E.1.5	Neural Network Models	545
E.2	The Computational Theory of Marr	545
E.2.1	The $2\frac{1}{2}$ -D Sketch	546
E.2.2	Overarching Claims	549
E.3	Overall Problems with Existing Models	550
F	Endnotes	560
	Bibliography	586

List of Figures

2.1	Concatenation of Temporal Data.	21
2.2	Weighted Time Average of Temporal Data.	22
2.3	Time Average of Responses For Temporal Data.	22
2.4	Jordan's Recurrent Architecture	24
2.5	Jordan's Recurrent Architecture with Forward Model.	25
2.6	Elman Recurrent Learning FFNN.	26
2.7	Recurrent Kohonen Learning Arrangement of Scholtes.	27
2.8	Ritter & Kohonen Semantic Map.	30
2.9	Temporal Learning Structure.	32
2.10	Temporal Learning Components.	33
2.11	Hebbian Update of Motor Control Weights.	33
2.12	Actual Neuron Layout and Connections.	35
2.13	Mathematical Symbols Located.	40
2.14	Lateral Interactions.	44
2.15	"Visual" Representation of Characters.	53
2.16	Vector Representation of Characters.	53
2.17	Alphabet Mapping	54

2.18	Alphabet Transitions Mapping	55
2.19	Learning a Temporal Sequence.	56
2.20	Generating a Temporal Sequence.	57
2.21	LAPS—Learning and Production of Sequences.	59
2.22	Semantic Priming.	62
2.23	Badly Separated Semantic Map of Winning Nodes—SOM1.	65
2.24	Semantic Map of Winning Nodes—SOM1.	66
2.25	Semantic Map of Winning Nodes—SOM2.	67
2.26	Dynamic LAPS Network.	70
2.27	Dynamic LAPS Neuron Layout.	71
2.28	Dynamic LAPS as Module.	74
2.29	LAPS Unit Data-flow Diagram (Two Input Stream).	75
2.30	Double Direct Input LAPS Unit.	75
2.31	LAPS Unit Data-flow Diagram Symbol.	76
3.1	ABC Associative Layer.	83
3.2	Associative Layer and External World Filters.	84
3.3	Equivalence of Diagrammatic Forms.	85
3.4	ABC Vision Model.	86
3.5	Vision System Dissected.	87
3.6	Multi-Modal Associations and Motor Control.	90
3.7	Overall Cognition System.	92
3.8	Overall Cognition System as Extended LAPS Modules.	93

3.9	Connectivity of Frontal Cortex.	94
3.10	Attractor Basins and Bifurcation.	103
3.11	Vocal Sounds in External Medium.	122
3.12	Internal Self-Talk as Sound-Equivalent Vectors.	123
3.13	Postulated Self-Talk Areas and Connections.	123
3.14	Self-talk Model.	125
3.15	Voluntary Control of Muscles.	128
3.16	Hardware Kohonen Map Model.	133
3.17	SPS/Search versus ABC /Table-lookup	134
3.18	Categorical Perception.	138
3.19	“Hard” Reflex versus “Soft” Reflex.	141
3.20	Different Context Selects Different Behaviour.	145
4.1	Absorption Curves for the Three Cone Types in the Human Eye.	170
4.2	Distribution of Rods and Cones in the Retina.	171
4.3	Lateral Geniculate Nucleus.	173
4.4	Sub-cortical and Cortical Processing Streams in the Macaque.	174
4.5	Functional Segregation in the Primate Visual System.	175
4.6	Change in Orientation Selectivity in Area V1.	176
4.7	‘Wife and Mother-in-Law’ Polysemous Image	179
4.8	A Classical ‘Difficult Image’	180
4.9	Typical Structure of a Computational Vision System.	182
4.10	Visual Illusions Tested for Cultural Differences.	186

4.11 Internal Image Representation.	189
4.12 Trans-Saccadic Fusion Experiment	191
4.13 Perceptual Span in Reading	194
4.14 Bistable Motion Perception.	198
4.15 Virtual Occlusion.	199
4.16 Occluding Square.	200
4.17 Feedforward and Backprojection Connections	202
4.18 Saccade Scanpaths.	210
4.19 Simple Spatial Filters.	215
4.20 Saccade/Fixation Points.	216
4.21 Vision System Compared	218
4.22 Occluded Objects.	220
4.23 Occluded Objects with Mixed Filters at Joins.	220
4.24 Occlusion Filters.	221
4.25 Vision Network.	222
4.26 Vision Network with Learned Saccades.	223
4.27 Minsky-Papert Figures.	227
4.28 Minsky-Papert Figures and Temporal Learning	228
4.29 Receptive Fields.	230
4.30 Treisman Visual Search.	232
5.1 The Development of Language in the ABC Model.	259
5.2 Flows for Implicit and Explicit Learning.	272

5.3	Short-Term and Long-Term Memory.	274
5.4	Hemineglect.	283
5.5	Commissurotomy.	289
5.6	Gestalt Behaviour and Rules.	294
5.7	Reporting of Conscious Activity by Libet.	313
A.1	Brodmann's Mapping of Cortical Regions.	348
A.2	Schematic of Cortico-thalamic Connections.	351
A.3	Topological Mappings.	353
A.4	Cortical Plasticity.	356
A.5	Areas of Cortex of Rhesus Monkey.	357
A.6	Areas of Cortex of Rhesus Monkey.	358
A.7	Auditory Intrinsic Connections.	360
A.8	Visual Intrinsic Connections.	363
A.9	Architectonic Subdivisions of the Frontal Areas.	364
A.10	Multimodal Areas and Connections.	366
A.11	Premotor and Prefrontal Multimodal Areas.	368
B.1	Stopwatch FSTN.	385
B.2	Finite State Automata—FSAMap.	386
B.3	Training the FSAMap.	387
B.4	Running the FSAMap.	388
B.5	FSAMap Dissected.	389
B.6	Number FSA.	391

B.7 Bidirectional Link FSA.	393
B.8 Reber FSA.	394
B.9 Finite State Automata with Action Vector.	395
B.10 XOR Winning Nodes.	399
B.11 Consonant/Vowel FSA.	402
B.12 Consonant/Vowel Winning Nodes.	402
B.13 Bidirectional Link FSA Revisited.	404
B.14 Specific Path Information versus Generalisation.	404
B.15 Multiple-choice Transition Learning.	406
B.16 Generalisation of SOM Nodes.	408
B.17 Actual FSTN Learned By System.	409
B.18 Actual FSTN Learned By System—Extended Learning.	410
B.19 Generalisation of SOM Nodes—Extended Learning.	411
B.20 Actual FSTN Learned By System—Winning Node Links.	411
B.21 Actual FSTN Learned By System—Separated C Loop.	412
B.22 Bidirectional Loop—Generalisation Over Epochs.	414
B.23 Flow Diagram After 80 Epochs.	415
B.24 Bidirectional Loop—Generalisations Over Refractory Period.	417
B.25 Reber FSA Revisited.	418
B.26 Generalisation of Reber Sequence Winning Nodes.	420
B.27 Actual FSA Learned for the Reber Sequences.	421
B.28 Generalised Terminating Sequences At Each State.	422
B.29 Counting Via Temporal Memory.	423

B.30 Short Term Memory—Possible Reason For Limitations.	426
B.31 Possible Mechanism For Chunking.	426
B.32 Complex Finite-state Grammar With Embedded Clauses.	427
B.33 Complex Finite-state Grammar—Internal Sub-clause.	428
B.34 Complex Finite-state Grammar—Internal Component Winning Nodes. . .	429
B.35 Complex Finite-state Grammar—Linked Character Winning Nodes. . .	429
B.36 Complex Finite-state Grammar—Linked Transition Winning Nodes. . .	430
B.37 Learning Embedded Sequences Using Context.	432
B.38 Hierarchical Cluster Diagram—SRN.	433
B.39 Embedded English Sentences—‘Rules’.	440
B.40 Context and Embedded Sentences.	442
B.41 Elman Embedded Sentence Network.	443
B.42 English Embedded Sentences—Build-up with Restricted Components. .	447
B.43 English Embedded Sentences—All 12 Sentences.	448
B.44 English Embedded Sentences—Flow Paths for All 12 Sentences.	449
B.45 English Embedded Sentences—No Build-up.	450
B.46 English Embedded Sentences—Singular and Plural.	452
B.47 English Embedded Sentences—Singular and Plural.	454
B.48 English Embedded Sentences (4,4,2)	455
B.49 English Embedded Sentences (4,4,2) con’t	456
B.50 English Embedded Sentences—‘Semantics’.	458
B.51 English Embedded Sentences—‘Semantics’ (con’t).	459
B.52 Semantic Priming on Transition SOM.	460

B.53 Context Biased Recurrent Links.	461
B.54 Fixed Transitions in FSA.	462
B.55 Variable Transitions in FSA.	463
B.56 Parsing a ‘for’ Statement—One FSAMap.	463
B.57 Hierarchical FSAMaps.	464
B.58 Variable State Transitions.	465
B.59 Linked FSAMaps.	465
B.60 Context Sensitive Grammar $a^n b^n c^n$ —Three FSA Maps.	466
B.61 Parsing a ‘for’ Statement—two FSA Maps.	470
B.62 Partial FSTN for c-language.	472
B.63 Learning to ‘Recognize’—Characters into Words.	485
B.64 Loop Trace.	492
C.1 Correspondence Between Images and Sentences	495
C.2 Receptive Fields Used in Simulations	497
C.3 Plots of the Spatial Frequencies (1D)	499
C.4 Spatial Frequency 4.0 cpd	500
C.5 Schematic Diagram of Model Visual Filters	503
C.6 Connecting <i>inSites</i> and <i>outSites</i> of Modules	505
C.7 Loader Window	509
C.8 Experiment Window	510
C.9 Experiment Window During Run	511
C.10 Association Window	512

C.11 Association Window (con't)	512
C.12 Kohonen Window	513
C.13 Kohonen Window (con't)	513
C.14 Retina Window	514
C.15 Retina Window (con't)	514
D.1 Computer Binary Representation.	519
D.2 Representation of the Digit 'Three' on a Computer Screen.	520
D.3 Electromagnetic Interaction.	526
D.4 Gravitational Interaction.	527
D.5 Particles in a Gas.	527
E.1 $2\frac{1}{2}$ -D Sketch.	547
F.1 Lateral Inhibition	568
F.2 Alternate Vision Structure.	570
F.3 Minimal Parity FSA.	575

List of Tables

2.1	Semantic Map Sentences.	29
2.2	List of Symbols.	39
3.1	Exteroceptive Modalities and Sub-modalities.	81
3.2	The Form and Meaning of Infant Words.	112
3.3	Overextension and Restructuring of Words.	114
4.1	Vector Values for Line Filter	216
4.2	Vector Values for Colour Filter	217
4.3	Recognised Concept Scores.	222
B.1	State Transitions for Stopwatch FSA.	386
B.2	State and Input Vectors.	387
B.3	Error State Transitions.	393
B.4	English Sentence Components.	436
B.5	Simple English Sentences.	437
B.6	Phrase Structure Grammar for Embedded Sentences.	439
B.7	Elman Sentence Build-up.	444
B.8	Winning Node Word Counts.	446

B.9	Word Winning Nodes (4,4,2)	457
B.10	Count FSA—State Transition Table.	467
B.11	Control $a^n b^n c^n$ FSA—State Transition Table.	467
B.12	Control $a^n b^n c^n$ FSTN—Error Handling.	468
B.13	C-like Language Description.	470
B.14	Example Code.	471
B.15	State Transitions—Parse <i>print</i> and <i>read</i> Statements.	472
B.16	State Transitions—Parse <i>expression</i> .	473
B.17	State Transitions—Parse <i>for</i> , <i>if</i> and <i>while</i> Statements.	474
B.18	State Transitions—Parse Statement Block Brackets.	474
B.19	State Transitions—Context FSAMap.	475
B.20	Results Using Example Code (a).	476
B.21	Results Using Example Code (b).	477
B.22	Results Using Example Code (c).	478
B.23	Results Using Example Code (d).	479
C.1	Wilson Spatial-Frequency Sensitivity Constants	499
C.2	Nearest Neighbour Filter Separations	502
D.1	Differences Between the Brain and a von Neuman Computer	518

Chapter 1

Introduction

1.1 Overview

The dominant paradigm of Artificial Intelligence, Cognitive Science, Cognitive Psychology and even Linguistics remains the symbol processing system (SPS). This model treats the brain as a formal system, with intelligent behaviour being viewed as the syntactic manipulation (via some computational process) of linguistic-based symbols. The paradigm insists that the processes of cognition are essentially the same as those of a traditional (von Neuman) computer, and that the mechanism of cognition is computation.

More recently, connectionist approaches have been proposed as an alternative explanation of cognition. A number of models have been put forward, such as feed-forward neural nets (FFNN) and simple recurrent networks (SRN). Most attention to date has been directed toward the FFNN model.

However, a number of criticisms have been directed at both of these approaches. For the SPS model, these criticisms include the lack of a mechanism for symbol grounding, questionable biological feasibility, the requirement of extensive innate facilities, computational intractability and the general exclusion of learning capabilities. For the FFNN model, the criticisms include a non-biological learning mechanism and a requirement for innate semantic linkages.

This thesis presents a new biologically-consistent model of cognition that challenges these prevailing views. The proposed structure involves the cross-modal linkage of sensory inputs and the learning of temporal sequences, resulting in perceptual categorisation via the formation of (self-organising) attractors in multi-level topographical maps.

The cross-linkage occurs initially at sub-modality levels (for example, in the visual system, between mappings of inputs from spatial frequency and angle, colour and motion sub-modalities), and then eventually at the level of modalities (vision, auditory, and somatosensory). Attractors are formed on differential inputs, and these attractors define perceptual categories. Feedback is achieved both externally (via action/response within the environment) and internally at several levels.

The overall model allows a unified account of the processes involved in the emergence of 'symbols' and of language. The linguistic symbols, which play such a prominent role in the SPS paradigm, are relegated to a different (but no less important) role in the ABC model.

The model is behaviourally-based and consistent with the action-based views of Vygotsky (Vygotsky [1962] 1986, Cole, John-Steiner, Scribner & Souberman 1978) and Piaget (Piaget & Garcia 1991, Brief 1983), and the neurobiologically-based theories of Neural Darwinism (Edelman 1989a, Reeke Jr. & Sporns 1993, Reeke Jr. & Sporns 1991, Edelman, Reeke Jr., Gall, Tononi, Williams & Sporns 1992). The model is further unifying in that it also incorporates aspects of the newly expressed views of Situated Cognition (Clancey 1993, Greeno 1989).

The model takes a new approach to denoting (labelling), viewing it primarily as another form of "sensory" input that can take part in the forming of perceptual attractors by cross-associations. Language is then the learning and reproduction of temporal sequences of denotations for objects, events and associated linking concepts. The use of language allows us to bifurcate the world into finer details through the use of arbitrary labels for communication and action. In this sense, language is seen as a socially reinforced structure rather than an innate property of events, per se. The symbols that form the basis of the model are then seen not as primary, but as externally formed

and agreed-upon labels capable of forming associations with other perceptual inputs. In other words, language is seen as simply an (albeit specialised) input (and output) modality (audition), with writing partaking of the visual input modality.

The unified single-architecture neural model overcomes many of the problems associated with the current symbolic and feed-forward models. Such broad-reaching claims are not so much related to the explicit knowledge and anatomical enumeration of detailed processes, but rather, the conjecture that the fundamental adaptive and generative processes in cognition can be encapsulated in the proposed system.

Although a computational simulation of the model is presented, the proposed model is of itself non-computational (in that it incorporates only neuronal processes and does not require any ‘computer’ capabilities). The model is more akin to a soft reflex, in which the connections between inputs and outputs are not hard-wired but learned through associations.

In summary, we propose a new model of cognition that includes perceptual learning, categorisation, learning, self-organisation, and recognition in an integrated model that overcomes many of the problems of the existing paradigms. The new model is presented as a new theoretical framework for further experimentation.

1.1.1 Conceptualisation, Action, and Language

This thesis is about conceptualisation—how do animals and humans form concepts about the world that better enable them to survive and thrive, and about actions (behaviour)—how do animals and humans learn appropriate behaviour to accompany this conceptualisation. In a sense, the two are inseparable—concepts form behaviours, and behaviours forms concepts.

The initial stage of cognition (shared by humans and animals) is the formation of perceptual concepts. Perceptual conceptualisation is the formation of learned behaviours based not on linguistic terms, but on inputs from external sensory and internal bodily sources. We propose a mechanism for this low-level perceptual learning, and indicate how it relates to ‘non-conscious’ behaviours such as sensorimotor skills, intuition

and body-language. Perceptual concepts include attractors from multiple modalities—auditory, visual, somatosensory, and so on, as well as proprioceptive, limbic and other internal data—and are linked with corresponding and appropriate behaviour.

Once formed, perceptual concepts can be associated with denotations (labels) formed by additional attractors provided by behaviour. Primarily this is speech and audition, which provide sound denotations to associate with other perceptual concepts. Thus a linkage is suggested between perceptual categorisation, labelling and language, providing a solution to the so-called *Symbol Grounding Problem*.

The process of cognition is seen not as linguistic in nature, but a flow over concepts and actions. The learning and reproduction of temporal sequences is then seen as a major component of cognition, with recurrency at multiple levels.

Language in turn is the learning and reproduction of temporal sequences, linked to perceptual (and subsequently, also linguistic) concepts. However, the underlying basis is through perceptual categorisation. A proposal is given to extend the role of language to self-talk, a recurrent (internal) linkage of motor outputs which is equivalent to speech, but which is short-circuited within the brain. A neural mechanism is described.

This *self-talk* and similar recurrent links within other modalities form a new suggestion for the origin of thinking, and this is discussed in relation to consciousness.

In summary, the *Adaptive Behavioral Cognition* (ABC) model is a radical alternative to the current models of cognition that provides a fresh insight into the working of the brain, and overcomes many of the problems within existing models.

1.2 Chapter Layout

This chapter, starting in the following section, looks at a brief history of recent ideas about cognition. We then look briefly at the two current leading paradigms of cognition—cognitivism and connectionism.

One of the pillars of the ABC model is the learning of temporal sequences. Chapter 2 looks at temporal learning, including the reproduction of previously learned sequences.

In Chapter 3, we develop the ABC model and discuss various aspects of it. We also look at how the new model overcomes many of the problems facing the current paradigms.

Chapter 4 takes an in-depth look at one sensory modality—that of vision. A new theory for vision and object recognition is postulated, based on the overall ABC model.

The new model differs significantly from previous models, and in Chapter 5 we take up a further discussion of the model, and look at some of the implications that arise from it. This chapter includes a discussion on language and thinking, implicit and explicit learning, and consciousness.

Chapter 6 provides a summary and a discussion of future research.

In Appendix A we examine some of the current neurological data, and show that it provides good general support for the new model.

In Appendix B we provide details of a number of temporal learning experiments which were performed in order to establish the temporal learning capabilities of the ABC model. These experiments are in addition to those described in Chapter 2.

An on-going full-scale computer simulation of a major part of the ABC model is described in Appendix C. The test-bed is based on what came to be known as ‘Lakoff’s Challenge’, a reference to a well-known author of the original challenge paper (Feldman, Lakoff, Stolcke & Weber 1990). The challenge is described in Section C.2.

We discuss several issues regarding computers and cognition in Appendix D. We also look at issues concerning computation, representations and theories, and their role in cognition.

A number of background issues regarding vision are discussed in Appendix E. These issues relate to current theories of visual recognition (mainly in relation to the computer vision literature), and problems that these theories face.

Appendix F provides a series of endnotes that were omitted from the main text. These provide background material that is not essential to the main argument, as well as programming details for many of the temporal learning experiments.

1.3 Current Theories of Cognition

Before we begin a description of our proposed model of cognition, we need to examine the prevailing views. We need to briefly examine the assumptions which pervade the traditional symbol manipulation paradigm, and the criticisms which have been made against it. We also need to examine the other current paradigm, *connectionism*, which has been proposed as an alternative to cognitivism. However, as we shall see, it also faces a number of problems that prevent it being viewed as a comprehensive theory of cognition.

Of importance here is the history of Behaviourism and Cognitivism.^{1 †}

1.3.1 Behaviourism

Behaviourism was a reaction against the previous use of introspection as a methodology.

The behaviourists had two research criteria:

- any data should be publically observable, thus excluding introspection,
- only the *behaviour* of animals and humans should be studied, and mentalistic topics such as thinking, imagination, intentions, desires, plans, symbols, schemas, or other mental representations should be completely avoided.

Their view was that all psychological activity can be explained without the need to resort to such mentalistic entities. To the behaviourist, the most important factor in forming the behaviour of individuals was the environment. Individuals were seen as passive reflectors of various forces and factors in their environment.

The behaviourists developed general principles of conditioning and reinforcement to explain how such learning and shaping of particular behaviours might come about. They believed that the science of behaviour could account for anything an individual might do, including thinking, which was considered to be covert behaviour.

[†]This thesis uses, in addition to footnotes, a system of numbered *Endnotes*. These notes contain material of a background nature, and are indicated by superscript numbers within the text. The Endnotes are found in Appendix F beginning on page 560.

With leading researchers such as Ivan Pavlov, B.F. Skinner, E.L. Thorndike, and J.B. Watson, the behaviourist paradigm held sway until about the middle of the twentieth century. By disallowing irrefutable and subjective evidence from introspection, by rejecting vague and poorly defined concepts such as 'will' or 'purpose', and in trying to align a theory of human behaviour with their observations on animal behaviour, they were able to overcome the problems associated with introspectionism. But there was also a great deal of disillusionment, as many overzealous behaviourists rejected all discussion of important topics such as language, problem solving, planning, and imagination. The study of these 'unobserved' aspects of cognition was maligned and not tolerated. Although some behaviourists (such as B.F. Skinner with his *Radical Behaviourism*) allowed for these cognitive aspects, most did not, and so a ground-swell of opposition developed against the perceived sterility of behaviourism.

A strong debate had developed in the 50's and 60's around the need for any theory of human cognition to account for complex organised behaviours such as the use of language, the playing of games such as chess and tennis, or the playing of a musical instrument. In a very influential paper, Lashley (1951) discussed serially ordered behaviours such as these, and maintained that such complex behaviour could not possibly be explained by a series of simple associative chains between a stimulus and a response. For example, when playing a musical instrument, Lashley maintained that the time constraints "left no time for feedback, no time for the next tone to depend upon or in any way to reflect the course of the previous one". Further, slips of the tongue often anticipate words that only occur much later in a sequence. These events, Lashley claimed, cannot possibly be explained in terms of linear "A evokes B" chains, but must be planned and organised in advance.

Also at this time came Chomsky's so-called demolition of behaviourism, with his review of Skinner's book on language behaviour (Chomsky 1964). The subsequent over-reaction against behaviourism was palpable in many quarters, and it fell out of favour. Even today in some discussions, to even hint at support for the basic behavioural approach, or to allow one's discussion to mention behaviourism-like points, is to invite instant rebuke and derision.

The view taken in this thesis is that the behaviourist approach was rejected prematurely. Certainly the extreme view of not considering any form of cognitive process must be rejected, but we need to re-examine the fundamental tenets of behaviourism in the light of developments in the last 50 years, and the failure of the *cognitivist* approach which replaced it.

1.3.2 Cognitivism

In the late 1940s, with the advent of calculating machines and computers, a number of researchers were beginning to advocate a similarity between the working of the brain and the just emerging technology of the computer. Researchers such as John von Neumann, Turing and others pushed the idea of the computer as the model of cognition.

Also at about this time, an alternate paradigm for cognition gained support—the beginnings of *connectionism*. These two research programs proceeded in parallel until the 1960s, when connectionism suffered a serious setback that was to hold it back for twenty or so years.

The computer model became the dominant paradigm, and even today is still the major research program in Cognitive Science, Artificial Intelligence (including Computer Vision and Machine Learning), Cognitive Psychology and Linguistics. The model goes by a number of names—*cognitivism*, *symbol processing system*, *representationalism*, even *GOFAI* (Good Old-Fashioned AI). In this thesis, we will generally use the term *cognitivism*.

Cognitivism was a reaction against behaviourism, and whereas behaviourism excluded from its discussion the so-called processes of the ‘mind’, cognitivism concentrated on isolated mental processes to the exclusion of (bodily) behaviour.

Cognitivist entails a belief that (Dreyfus & Dreyfus 1988, page 310):

“the human brain and the digital computer, while totally different in structure and mechanism, [have] at a certain level of abstraction a common func-

tional description. At this level both the human brain and the appropriately programmed digital computer [can] be seen as two different instantiations of a single species of device—a device that [generates] intelligent behaviour by manipulating symbols by means of formal rules.”

The view holds that both minds and digital computers are physical-symbol systems. The *Physical Symbol System* hypothesis was put forward by Newell & Simon (1981):

A physical symbol system has the necessary and sufficient means for general intelligent action. . . . By ‘necessary’ we mean that any system that exhibits general intelligence will prove upon analysis to be a physical symbol system. By ‘sufficient’ we mean that any physical symbol system of sufficient size can be organised further to exhibit to exhibit general intelligence.

Cognitivism requires a formal *representation* of the world, and looks to logic to provide the linking mechanism. This view follows a long history of rationalist, reductionist tradition in philosophy, going back to at least the time of Plato. This tradition was enhanced by Descartes, who made the assumption that all understanding required the creation and manipulation of appropriate representations, and that these representations could be analysed into primitive elements (*naturas simplices*) such that all phenomena could be understood as complex combinations of these simple elements.

Other philosophers who made contributions include Hobbes (who made the suggestion that the elements were formal components related by purely syntactic operations, so that reasoning could be reduced to calculation), Leibniz (who posited that every human concept must be composed of complex combinations of ultimate simples, and if these concepts are to apply to the world, there must be simple features that these elements represent), through Frege and Russell. Wittgenstein contributed his apotheosis of the reductionist, rationalist tradition (“Tractatus Logico-Philosophicus”), which postulated a pure form of this syntactic, representational view of the relationship of the mind and a world composed of facts.

Artificial Intelligence (AI) is the continuation of this philosophical tradition. AI attempts to determine primitive elements and logical relationships that mirror the prim-

itive objects and their relationships within a world that is ready to be ‘carved at its joints’.[†]

This traditional reliance on a reduction to logical primitives means that practically all of Western philosophy supports the view that cognition is symbolic information-processing (Dreyfus & Dreyfus 1988).

1.3.2.1 Defining Statements and Assumptions

A number of authors have stated what they believe are the defining statements or simplifying assumptions made by researchers in the cognitivist domain. In this section, we list several of these statements with the view to returning to them later, following our discussion of the ABC model.

In his book on the history of Cognitive Science, *The Mind’s New Science*, Howard Gardiner states five aspects of cognitive science which he believes define the field (Gardner 1985, page 6):

- ... the belief that, in talking about human cognitive activities, it is necessary to speak about mental representations and to posit a level of analysis wholly separate from the biological or neurological, on the one hand, and the sociological or cultural, on the other.
- ... a faith that central to any understanding of the human mind is the electronic computer. Not only are computers indispensable for carrying out studies of various sorts, but, more critically, the computer also serves as the most viable model of how the human mind functions.
- ... the deliberate decision to de-emphasise certain factors which may be important for cognitive functioning but whose inclusion at this point would unnecessarily complicate the cognitive-scientific enterprise. These factors include the

[†]Note that the ABC model also relies on the use of ‘primitive’ or simple data inputs—vectors formed by sensory filters to the external world and to internal processes—but the ‘simples’ are certainly not of a linguistic or conceptual nature. Further, the inputs are not treated as ‘representations’ and cannot be regarded as separate from the brain.

influence of affective factors or emotions, the contribution of historical and cultural factors, and the role of the background context in which particular actions or thoughts occur.

- ... the faith that much is to be gained from interdisciplinary studies.
- ... the claim that a key ingredient in contemporary cognitive science is the agenda of issues, and the set of concerns, which have long exercised epistemologists in the Western philosophical tradition.

As we shall see in discussions in subsequent sections, the ABC model takes to task all of these assumptions and beliefs except the need for interdisciplinary study.

Dennett (1989) cites five simplifications made by AI practitioners, supposedly in order to advance and progress:

- Ignore both learning and development; attempt to model the “mature competence” first, postponing questions about how it could arise.
- Isolate the particular subcomponent or sub-sub-component, ignoring almost all problems about how it might be attached to the larger system.
- Limit the domain of operation of the modelled system or subsystem to a tiny corner of the real domain—try [to] solve a “toy problem”, hoping that subsequent scaling-up will be a straightforward extrapolation.
- Bridge various gaps in one’s model with frankly unrealistic or even deliberately “miraculous” stopgaps—“oracles”, or what [Dennett has] called “cognitive wheels”.
- Avoid the complexities of real-time, real-world coordination by ignoring robotics and specialising in what [Dennett calls] “bed-ridden” systems: systems that address the sorts of problems that can be presented via a narrow “verbal” channel, and whose solutions can be similarly narrowly conveyed to the world.

Costall & Still (1991, page 2), in their book attacking the cognitivist position, outline four serious limitations of the rules and representations required of cognitivism:

- Solipsism—how can the knower ever reach beyond internal representations to the reality they are supposed to represent?
- Development—how can a system of formal rules ever be flexible enough to capture the mutuality between a growing organism and its richly structured and changing environment?
- Relevance—how does anyone following a rule know when to apply that rule? (The so-called ‘frame’ problem.)
- Meaning—how do symbolic representations attain their semantic status?

Costall & Still (1991) summarise by noting that cognitivists all share the belief that:

... the world is not experienced directly, but only through representations which impose human meaning upon the lifeless world described by the physical sciences. Psychological explanation, therefore, is couched in terms of internal, mental representations of the world, and the rules governing their manipulation and transformation. Although psychologists and philosophers are often uneasy about the dualisms it implies—between mind and body, between organisms and environment—it still dominates the field; indeed, the cognitivist view has been recently bolstered by the metaphor of the mind as computer.

1.3.2.2 Symbol Processing Systems

The representational paradigm reaches its peak in the *physical symbol system* views of Simon and Newell (Newell 1980, Newell 1990, Newell & Simon 1988, Vera & Simon 1993a), and the SOAR project. Their work gives a formal statement to theoretically underpin work in AI, claiming that human intelligence entails a symbol system and a set of procedures to manipulate these symbols. In combination with Fodor’s *Language of Thought* hypothesis (Fodor 1976), AI reduces human thought (intelligence) to logic, rules, knowledge, and symbol processing.

Additionally, the *Knowledge Representation* hypothesis further cements the central role of language within the cognitivist agenda (Smith 1985):

“any process capable of reasoning intelligently about the world must consist in part of a field of structures, of a roughly linguistic sort, which in some fashion represent whatever knowledge and beliefs the process may be said to possess”.

This model also states that symbols may be manipulated based solely upon their ‘shape’ or ‘form’, without any regard to their meaning. That is, traditional AI is symbolic processing—the syntactic manipulation of symbols. In line with traditional Western analytic philosophy, AI has always made the assumption that the processes of cognition can be examined in isolation from the external world, and that somehow the symbols used can later be attached to objects in the world.

Thus the cognitivist view is logic based, and deals with higher cognitive functions, discrete (physical) systems, and calculi. The view holds that humans cognise the world through symbols, and symbols alone.

As well as representations, the cognitivist approach relies on *rules* to manipulate the representations. Thus an emphasis is placed on ‘rule-based systems’ such as Soar, Production Systems, Expert Systems and in linguistics, systems such as Chomsky’s *Transformational Grammar*. As well, the activity of *problem solving* is seen as a primary example of intelligence, and worthy of study. Learning is given very little attention, and it is only recently that the discipline of Machine Learning has become active.

1.4 Degenerating Research Programme

We concur with many other researchers that classical cognitivism represent an example of what Lakatos (1978) has called a degenerating research programme. These critics come from diverse fields and include Dreyfus (1992), Winograd & Flores (1986), Greeno & Moore (1993), Agre (1993), Suchman (1993), Clancey (1993), and the volume by Costall & Still (1991). Others include Lakoff (1986), Bruner (1990), Bickhard &

Richie (1983), Coulter (1983), Coulter (1989), de Gelder (1985), Dietrich (1990), Gibson (1979), Hamlyn (1990), Heil (1981), Heritage (1984), Rogoff (1984), Lave (1988), Malcolm (1971), Middleton & Edwards (1990), Prilleltensky (1990), Russell (1984), Sampson (1981), Searle (1989), Shanon (1988), Skinner (1977), Skinner (1985), Williams (1985) and Woolgar (1987).

The early promise of cognitivism has not continued. Rather than fulfilling Minsky's prediction that "within a generation the problem of creating "artificial intelligence" will be substantially solved' (Minsky 1967), AI is still an open issue.

Cognitive Science, being currently more of a *descriptive* science and hence not needing as much of the hard-edged practical demonstration required of AI, has not been as much criticised. As well, Cognitive Science has hedged its bets by partially opening up its doors to connectionism, but symbol-based cognitivism remains the major paradigm in Cognitive Science, Cognitive Psychology, and Linguistics.

Despite an immense research effort worldwide, progress has been slow, and major theoretical, philosophical and practical problems remain unanswered. Cognitivism has been a failure, and needs to be replaced as a research paradigm within Cognitive Science. Cognitivism would be more appropriately renamed *advanced computer programming*.²

1.5 Connectionism

Many AI and Cognitive Science researchers have taken up the connectionist research programme. Following a serious setback in the 1960s, caused in part by an overzealous critique of the power of perceptrons by Minsky & Papert (1988), the connectionist approach is now undergoing a resurgence in popularity.

Most current research in connectionism concerns the feed-forward neural net (FFNN) methodology (Rumelhart, Hinton & Williams 1986, Stone 1986, McClelland & Rumelhart 1988, Hinton 1989, Greenwood 1991, Hecht-Nielsen 1987, Khanna 1990), but other neural structures studied include simple recurrent nets (SRN) (Rumelhart et al. 1986, Williams & Zipser 1989, Pineda 1989, Elman 1990), cascade correlation (Fahlman & Lebiere 1990, Fahlman 1991a), adaptive resonance theory (ART) (Carpenter &

Grossberg 1988, Carpenter & Grossberg 1992a, Carpenter, Grossberg & Reynolds 1991, Carpenter & Grossberg 1991) and Kohonen self-organising maps (Kohonen 1984, Kohonen 1982, Kohonen 1990, Ritter & Kohonen 1989, Kangas, Kohonen, Laaksonen, Simula & Venta 1989, Clippingdale & Wilson 1996). Review and general references are found in Hertz, Krogh & Palmer (1991), Rumelhart & McClelland (1986), McClelland & Rumelhart (1986) and Anderson & Rosenfeld (1988).

There are a number of criticisms of the current status of connectionism. This thesis is strongly in the connectionist camp, but we believe that current research has too narrow a focus and concentrates too heavily on models that are more engineering than biological, and more task-related than teleological to qualify as a true theory of cognition.

The FFNN model is generally regarded as a *software* method, one that can be programmed on a serial computer. Our view is that biological systems do not rely on software, but are *hardware*-only systems. Most people have looked at the implementation of neural net software, with little thought to how this might translate into hardware methods.

The FFNN model uses a learning rule that is of questionable biological validity. The *back-propagation* method relies on finding an *error* difference between the current output of the network and the *expected* output. This error is then back-propagated through the network to update the weights so that the same inputs at a subsequent time will produce a *closer* approximation of the required output. The problem with this scheme from a biological point of view is that there is no *expected* output—an end result is not known. In biological systems, there is no such concept as a desired end-point or the minimisation of errors—the world is simply too dynamic. The brain of a creature will simply self-organise to whatever the incoming statistics allow. The creature does not know a priori what is required, otherwise why *re-learn* it. For the FFNN model, the *programmer* may know the required output, but this assumption is biologically infeasible. Further, a biological (neural) mechanism has not been proposed for the back-propagation update rule.

While learning is thought to be a major component of connectionism, most proponents

of the FFNN model do not take learning seriously. In most cases, researchers propose that the input to the network be composed of *concepts* which are known a priori, and so the output is limited to combinations of these existing conceptualisations. Little effort is spent in determining just where the concepts come from in the first place. *Local* representations generally require some form of prior conceptualisation. A connectionist model of the brain should not assume innate *concepts*, but suggest a means whereby concepts may be formed based on external sensory inputs and internal body-related sources.

A general theory of cognition should not ‘design a net’ as in some localist networks, but rather have a general structure that is able to learn for itself—otherwise the net will only exhibit the ‘intelligence’ built into it by the designer. Intelligence is determined by motivations and purposes within an organism, as well as successful behaviours picked up by the organism from an ongoing external world and culture.

1.6 Summary

Much of current AI and Cognitive Science appears to be direction-less, and very little of real substance has been achieved following the inflated promises of the past. To date, there is no general, complete and comprehensive theory linking the neural substrate with the behaviour of humans and other animals.

This thesis is an attempt at such a synthesis; an attempt to unify and link many previously disjoint ideas and observations—from the neural level through neurophysiological structures to behaviour. It of necessity covers a number of disciplines, and as it also deals with the very essence of cognition—language, concepts, words, meaning and world views—we need to be extremely careful with our use of terms[†].

During the development of the thesis, priority has been given to maintaining full bi-

[†]One term I have tried to avoid throughout this thesis is *mind* as it leads to the sort of ‘category error’ described by Ryle (1949). *Mind* implies a separate entity for ‘mental processes’, separate from the brain. I prefer to use *cognition* which denotes the processes of the brain, consistent with the behavioural cognition tone of the thesis.

ological feasibility. An attempt has been made to include nothing that could not be described in terms of neural structures. The question asked throughout was—given the available knowledge of neural components as building blocks, how could these be combined to achieve a system with the observed behaviour and capabilities of humans and animals. The theory also had to be feasible from an evolutionary point of view and phylogenetically consistent—that is, could it explain the similarities and differences of cognition in animals and humans, especially as regards language.³

Chapter 2

Learning and Generating Temporal Sequences

2.1 Introduction and Caveats

The importance of fully simulating and testing a theoretical cognitive model must be strongly emphasised. It is often not until the development of the simulation is under way, that many implementation and other difficulties are discovered. These are often overlooked in a theoretical approach. This chapter discusses the software simulation of various components of the overall ABC model.

In spite of this requirement to fully test a cognitive model, one also needs to be reminded of the great difficulty in undertaking such a simulation, in any real sense, because of the severe limitations of current computer equipment. Despite tremendous gains in recent years in both computer capacity and speed, the sheer enormity of the number of neurons in the brain of even the lowest creatures means that a full simulation is beyond the capacity of current computers. The serial processing architecture of current computers is a severe limitation in trying to simulate the massively parallel processes of cognition.

The brain of a human is estimated to contain some 10^{12} neurons, with perhaps 10^{15} synaptic connections (Kandel & Schwartz 1983). Each neuron forms about 1,000 synaptic connections, and receives many more connections in its dendritic tree. As is pointed

out in Kandel & Schwartz, this is a huge number—*there are more synapses in the human brain than there are stars in the galaxy.*

In the current simulations we are limited to perhaps a few thousand neurons at most. A true test of any cognitive model will only be realized when the massively parallel processes found in the brain are available as hardware components. This point is further considered in the discussion of results in Section 2.7. So we are forced to begin with relatively simple systems.

In this chapter, we report on a number of temporal experiments that were performed using the ABC model. A number of other experiments are also discussed in Appendix B. Full program details of all temporal learning examples (including parameter settings) are given in the Endnotes section beginning on page 560. Endnotes are indicated by superscripted numbers within the text.

2.2 Temporal Learning

The learning and reproduction of temporal sequences is an extremely important component of human and animal behaviour. As well as the motor control involved in routine behaviour such as walking, running, talking, tool use and so on, humans have an apparently remarkable capacity for learning and reproducing temporal sequences.

A classical pianist is able to learn and reproduce long and complicated sequences of hand movements; gymnasts and ballet dancers are able to display fine muscle control for extended periods; poets and musicians can reproduce at will, sequences of many thousands of words or intonations.

And once learned, a behavioural sequence may be retained for many decades. I can still remember a somewhat silly series of nonsense words taught to me by a chemistry teacher some thirty-five years ago—a sequence designed to aid in the recall of the first few elements of the periodic table.[†] Although I have had no use of this sequence since that time, it still remains with me.

[†]“Hi, Hello Little Beryl Brown Chewing Nuts On Friday, Ney, Namgal Sips Chlorine, Arh” corresponding to H, He, Li, Be, B, C, N, O, F, Ne, Na, Mg, Al, Si, P, S, Cl, Ar.

How is it that a few bars of music from a popular song is enough to prompt us into singing along with immediate recall of the music and words? How is it that with practice we are able to sight a small ball travelling towards us in excess of one hundred kilometres per hour, position our body in such a way that a tool held in our hand can strike the ball with such precise timing that we produce a fore-hand winner down the line? And how is it that we can learn a language?

Repetition and practice—learning—is certainly an important element in these endeavours. Unless one takes the view that these behaviours are innate and only need to be activated in some manner (discussed further in Section 3.4, especially in relation to language learning), then the learning, and in some cases the subsequent re-generation of temporal sequences, is seen as an important component of cognition.

This thesis rejects the inherently sequential cognitivist model of cognition. The architecture of the brain is one of massive parallelism of interconnected neurons, and one of the challenges of Cognitive Science is to find an alternative understanding of how this parallelism can account for the observed serial order of much of cognition. Further, the temporal sequences found in behaviour must be stored in the brain in some form, and not simply come about as the result of external stimuli as suggested by some over-zealous behaviourists. The challenge is to find the mechanisms of this storage of temporal sequences, how sequences may be learned, and how they can be used and reproduced in subsequent behaviour.^{4 5 6}

2.2.1 Previous Temporal Learning Results

There have been a number of previous studies on the learning of temporal sequences using neural networks, many in connection with speech recognition and language (Jordan 1986, Elman 1990, Tavan, Grubmuller & Kuhnel 1990, Kangas 1991).

Various approaches have been used to incorporate the temporal component, both explicitly and implicitly. One approach, which incorporates time explicitly by a transformation into a spatial dimension, has been explored by Waibel, Hanazawa, Hinton, Shikano & Lang (1989), Kangas (1989), Lang, Waibel & Hinton (1990), and Kangas (1991).

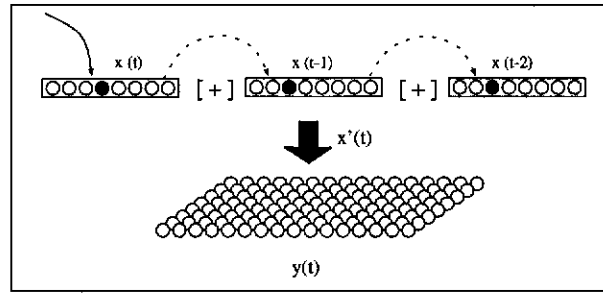


Figure 2.1: Concatenation of Temporal Data.

The work of Kangas (1989) is typical, and will be described here briefly. The method utilises a Self-Organising Map (SOM), and uses a number of methods to incorporate the element of time.

Figure 2.1 illustrates a *pattern concatenation* model,[†] in which the time-dependent input vectors are stored in a number of ‘registers’, and concatenated before being used with the Self-Organising Map. The registers are shifted after each cycle. This method is also known as the Time Delayed Neural Network (TDNN) model (Waibel et al. 1989).

An alternate approach, in which the input vectors are averaged over time, is illustrated in Figure 2.2. The weighting gives a backwards exponentially weighted sum $x'(t)$ of the original input values $x(t)$ as described in equation 2.1, where w is a weighting factor.

$$x'(t) = w \cdot x(t) + (1 - w) \cdot x'(t - 1) \quad (2.1)$$

The weighting factor w determines the extent of the averaging over time. A value of the parameter close to unity will diminish the importance of previous input vectors, whereas a small value will result in a time average over a longer time period, with less importance placed on the latest input vectors.

The third alternative considered in Kangas (1989) is illustrated in Figure 2.3. Here two Self-Organising Maps are used, with the original input vectors mapped onto the first map, and the averaged responses of this first map then mapped, in turn, onto the

[†]In Figure 2.1, the [+] notation denotes vector concatenation, while the solid arrow denotes a mapping of a vector to a SOM surface.

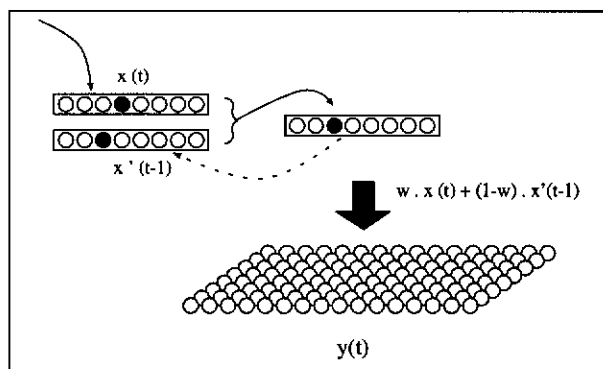


Figure 2.2: Weighted Time Average of Temporal Data.

second map.

The averaging of the responses of the first map is again given as a backwards exponentially weighted average $x'(t)$ of the outputs, with β a weighting factor.

$$x'(t) = \beta \cdot y(t) + (1 - \beta) \cdot x'(t - 1) \quad (2.2)$$

The classification of the outputs of the second map is thus not based on the current input vector, but rather on the time average of the responses that previous vectors produce on the first map.

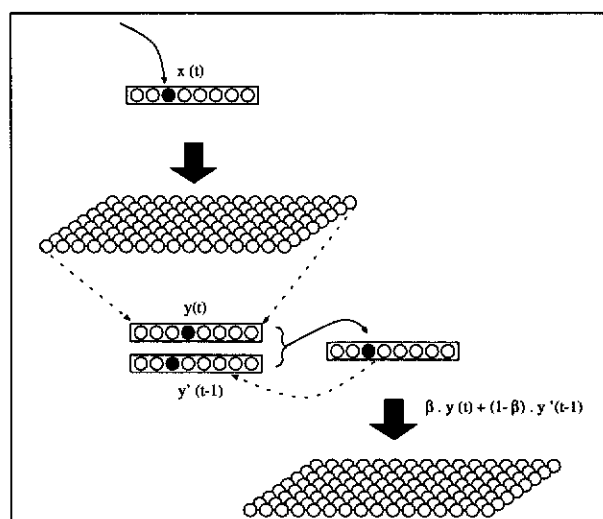


Figure 2.3: Time Average of Responses For Temporal Data.

Kangas obtained reasonable results with these arrangements, getting improved classification results on a temporal task over that of a standard SOM. Classification accuracy improved from some 89% (72% with noise) for the non-temporal learning case, to 97% (96%) in the case of averaging the data, 89% to 100% (72% to 99%) in the case of concatenation of from 1 to 9 pattern vectors, and 100% (98%) in the case of averaging of the responses.

Kangas (1991) also used these methods to obtain improved phoneme recognition.

Despite some reasonable results, these methods are generally unsatisfactory as the basis for a cognitive model; the concatenation method requires a buffering of the previous input data through a series of shift registers, and it is not clear that biological systems contain such components; further, as pointed out by Elman (1990), this method makes no distinction between relative temporal position (in the concatenated vector) and absolute temporal position; and finally, the averaging methods require a somewhat arbitrary averaging mechanism, with no indication as to their biological realisation.

Recurrent neural connections were introduced by Jordan (1986) (see also Jordan & Rosenbaum 1989, Jordan 1989b). Jordan was concerned with the parallelism of speech production, and the relationship between these parallel properties and the overarching sequential nature of speech. For example, nearby phonemes can overlap in time, either anticipating future phonemes or preserving the phoneme effect while another begins. This *co-articulation* is ubiquitous in utterances, and provide difficulties for a non-parallel solution.

The proposed network, shown in Figure 2.4, involved a feed-forward neural net with the output units recurrently connected to an extra set of input units—nodes that Jordan termed *state* units.

Suppose that there is a sequence of *actions* $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_s$ that are to be produced, in order, given a particular *plan* \mathbf{p} . The plan is assumed to remain constant during the production of the sequence, and is assumed to emanate from a higher level of the system. Both the actions and the plan are vectors, and different plan values will produce different action sequences.

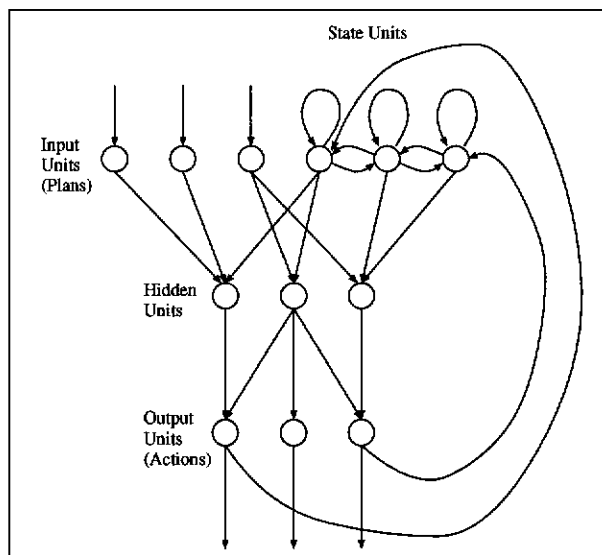


Figure 2.4: Jordan's Recurrent Architecture (not all connections are shown).

At each moment in time, the output vector is determined by

$$x_n = f(s_n, p) \quad (2.3)$$

and the state transitions are given by

$$s_{n+1} = g(s_n, p) \quad (2.4)$$

where the function f is to be learned by the network, and the next-state function g essentially constructed by Jordan.

The plan units vector and the state units vector are concatenated to form the input units to the multi-layer network. The state units are connected via recurrent loops from the output nodes, as well as recurrent loops within the state layer itself. This allows the next-state function to depend on the previous state as well as on the previous output. To ensure that state vectors at nearby points in time are similar, and that the system be able to represent repeated and similar sequences, Jordan chose to implement the next-state connections by assigning a weight of one to the connection from the output units, and to assign a weight of μ to a recurrent loop from each state node to itself. This is then an exponentially weighted average of past outputs in which all prior outputs are represented, but with reduced importance.

$$\begin{aligned}
 s_n &= \mu s_{n-1} + x_{n-1} \\
 &= \sum_{r=1}^{n-1} \mu^{r-1} x_{n-r}
 \end{aligned}
 \tag{2.5}$$

During learning, plan and state values are presented to the network and an output function is learned using back-propagation. Jordan considered learning the next-state function g using a modified back-propagation algorithm appropriate to recurrent networks, but rejected the idea as he felt there was little to be gained—he felt that all the hard work is in learning the output function.

Jordan also considered the problem of context sensitivity in which the same action can have different forms depending on the context (Jordan & Rosenbaum 1989 page 755, Jordan 1989a). The network used is shown in Figure 2.5. In addition to the structure of Figure 2.4, two additional layers of nodes form another recurrent loop. The outputs of the network are the Articulatory Units (as in the previous model), and the purpose of the additional recurrent loop is to provide an adjustment to the required sequence in task space. Errors generated at the task units between the actual and required

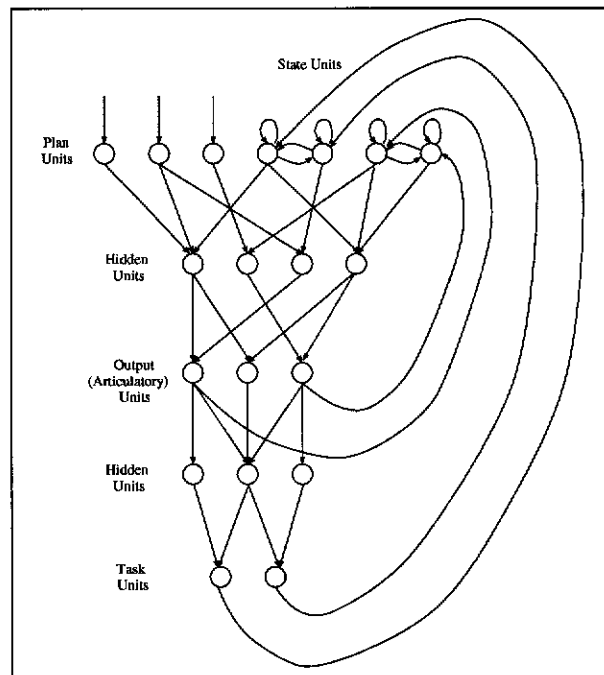


Figure 2.5: Jordan's Recurrent Architecture with Forward Model.

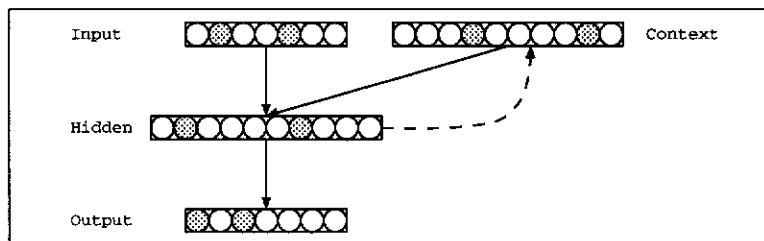


Figure 2.6: Elman Recurrent Learning FFNN.

task vectors are backpropagated to adjust the weights from the output units, and thus modify the contextual outer recurrent loop.

Elman (1990) conducted a series of experiments in temporal learning that have become quite influential in various fields. A number of these experiments have been duplicated and extended in this thesis using the temporal model proposed as part of the ABC model. Various differences and improvements over Elman’s results have been found, and these results will be described in subsequent sections. We will here restrict ourselves to a description of the model used by Elman, which is shown in Figure 2.6.

In this structure, a recurrent loop is again used to implicitly include the element of time. The overall network is a feed forward neural net that is trained using the backpropagation algorithm. An extra set of input nodes duplicates the hidden layer nodes that were evaluated on the previous cycle. Thus the dotted line on Figure 2.6 can be regarded as a recurrent link with fixed weights of 1.0, and the solid lines as trainable weights.

The input nodes which duplicate the hidden layer may be thought of as *context* units—they have to learn to map both the input values, and the previous internal state, to the appropriate output values.

Suppose initially that the context units are assigned a value of 0.5[†], and at time t an input vector is assigned to the input nodes. The input and context vectors are concatenated, and the combined vector treated as the input to the feed forward net. The output required for this input will enable the error to be backpropagated to update the weights at all layers.

[†]The activation function used bound values between 0.0 and 1.0.

At the next time cycle, $t + 1$, the values of the hidden layer nodes are directly copied to the context nodes, and the feed-forward learning process repeated. As with the usual FFNN learning methodology, the process is iterated until the outputs are learned to some required accuracy.

The network used by Elman has been duplicated in various other studies of temporal learning (Elman 1991, Elman 1993, Cleeremans, Servan-Schreiber & McClelland 1989, Cleeremans 1993, Servan-Schreiber, Cleeremans & McClelland 1991, Sharkey & Sharkey 1996), and the results obtained have opened up a whole new area of study into dynamical systems as applied to temporal learning, and in particular linguistic behaviours. However, there are a number of problems with the FFNN model which also apply to Elman's model, and these issues are discussed in Section 2.7.

Scholtes (1991) examined the use of multi-layer SOM surfaces in natural language learning. The arrangement he used is shown in Figure 2.7. In this arrangement, input vectors randomly assigned to incoming words ($x_s^{(1)}(t)$) are concatenated with a recurrent component ($x_r^{(1)}(t)$), and mapped onto the initial SOM surface. This mapping forms a topological map of the incoming 'symbols'.

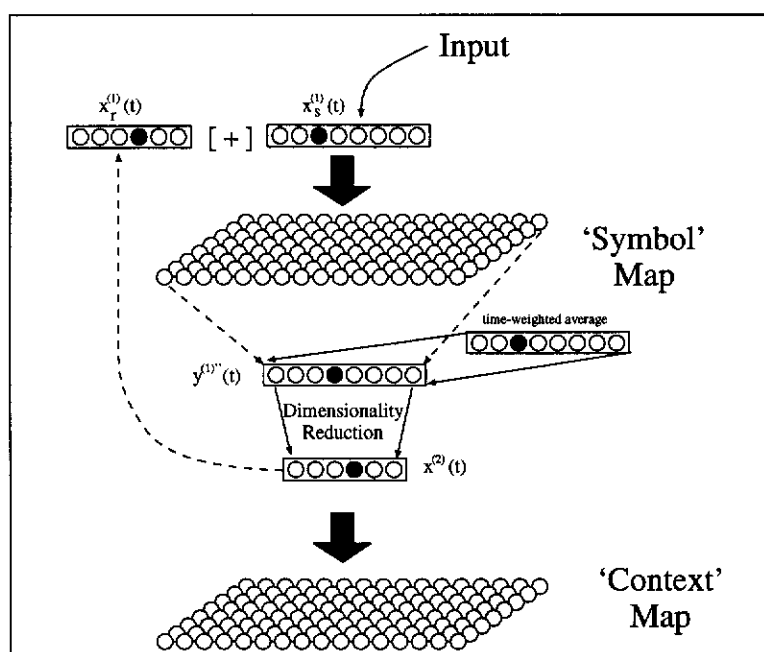


Figure 2.7: Recurrent Kohonen Learning Arrangement of Scholtes.

The output from the initial SOM surface is first normalised, then combined with the output from the initial SOM on the previous time period (which presumably has been stored off-line); that is

$$y^{(1)''}(t) = \frac{\omega \cdot y^{(1)'}(t) + (\omega - 1) \cdot y^{(1)'}(t - 1)}{\omega} \quad (2.6)$$

where $y^{(1)'}(t)$ is the normalised output of the initial SOM surface.

The time-averaged output is then reduced in dimension via a heuristic method, and the result both mapped to the second SOM surface and returned as the recurrent input component for the next cycle.

Scholtes was able to obtain a *semantic map* on the second SOM layer, and was able to show that temporal learning did occur. However, the results obtained were somewhat limited, with large computational requirements due to the slow convergence of the model. The method is also somewhat heuristic, and Scholtes gave no justification for the dimensionality reduction.

It is perhaps appropriate at this stage to describe the results of Ritter and Kohonen (1989) in the development of self-organising *semantic maps*. Simple sentences, using the words of Table 2.1 (a), were constructed from the patterns of Table 2.1 (b). Example sentences are shown in Table 2.1 (c).

A representation of these simple sentences was mapped onto a single SOM surface. This was done by generating a test set of random sentences, and then forming a vector for each word defined by its immediate predecessor and successor (the sentences are taken to be contiguous, and individual vector components are randomly assigned to each word). This extended vector then contains a measure of the particular *context* of the word. As each word appears in multiple contexts, an average over 10,000 sentences of all code vectors of predecessor/successor pairs surrounding each word was generated. This single vector 'context' representation for each word (suitably normalised) was then concatenated with the vector component used to represent the word itself. This 'word' vector component was scaled relative to the context component so as to maintain a relative influence of the symbol part in comparison to the context part. The full three

Bob/Jim/Mary	1	1-5-12	1-9-2	2-5-14	Mary likes meat
horse/dog/cat	2	1-5-13	1-9-3	2-9-1	Jim speaks well
beer/water	3	1-5-14	1-9-4	2-9-2	Mary likes Jim
meat/bread	4	1-6-12	1-10-3	2-9-3	Jim eats often
runs/walks	5	1-6-13	1-11-4	2-9-4	Mary buys meat
works/speaks	6	1-6-14	1-10-12	2-10-3	dog drinks fast
visits/phones	7	1-6-15	1-10-13	2-10-12	horse hates meat
buys/sells	8	1-7-14	1-10-14	2-10-13	Jim eats seldom
likes/hates	9	1-8-12	1-11-12	2-10-14	Bob buys meat
drinks/eats	10	1-8-2	1-11-13	1-11-4	cat walks slowly
much/little	11	1-8-3	1-11-14	1-11-12	Jim eats bread
fast/slowly	12	1-8-4	2-5-12	2-11-13	cat hates Jim
often/seldom	13	1-9-1	2-5-13	2-11-14	Bob sells beer
well/poorly	14				...
(a) Words.		(b) Sentence patterns.			(c) Example Sentences.

Table 2.1: Semantic Map Sentences.

component vector is then mapped to a SOM surface.

Ritter and Kohonen found that the words were organised on the SOM surface in a way that reflected both their *grammatical* and *semantic* relationships. Words of the same type (for example, nouns, verbs, and adverbs) were segregated into separate domains on the map. Further, each of these domains was further subdivided according to semantic similarities (for example, proper names were grouped together, as were animals and food items). A typical semantic map as found by Ritter and Kohonen is shown in Figure 2.8.

These results were significant in that they demonstrated for the first time that a semantic and/or a syntactic interpretation may be drawn from the relative mappings of vectors onto a SOM surface. The *interpretation* results from the *source* of the actual vectors mapped. In the Ritter and Kohonen experiment, the actual source of the vectors was rather arbitrary (averaged context, biased relevance of symbol and contextual components, etc) and not well justified, but the point remains that the vectors did contain semantic and syntactic information, and thus were mapped onto the SOM surface based upon this information. It demonstrated the potential of a self-organised process to form abstract maps.

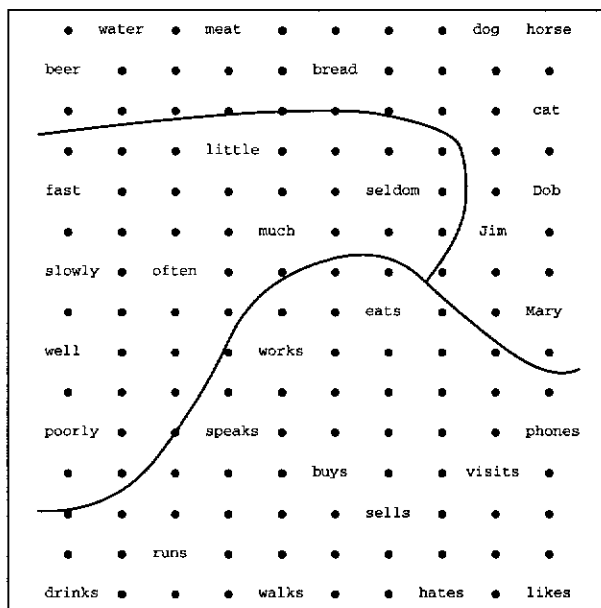


Figure 2.8: Ritter & Kohonen Semantic Map.

In this chapter we seek a much more dynamic approach to temporal learning that is able to both learn a sequence, and to then reproduce that learned sequence at some later time. This production aspect is important—it is often not much use being able to learn a motor action alone—we may need to be able to reproduce that learned action under similar conditions.

2.3 Temporal Learning Model

This section contains a description of the temporal learning model proposed in this thesis. The model forms an important component of the overall ABC model of cognition. We first describe the model in some detail, then move on to describe a number of experimental simulations that were performed using the model.

The model we propose brings together two important developments in artificial neural networks; recurrence and self-organising maps (SOM). Although some have tentatively combined these two elements before (for example, Scholtes 1991), their combination has not been given sufficient attention, in our opinion, and research has tended to concentrate on the simple recurrent networks (SRN) initiated by Elman (1990).

Topological maps appear to be ubiquitous in the cerebral cortex of the higher animals (Kohonen 1990), especially in the primary sensory areas. For example, the visual cortex contains maps of spatial frequency and orientation, and of colour (Van Essen 1985, Chino, Kaas, Smith III, Langston & Cheng 1992, Swindale 1992), the auditory cortex includes tonotopic maps (Brugge & Reale 1985, Reale & Imig 1980) and other auditory maps (Knudsen 1984, Merzenich, Jenkins & Middlebrooks 1984, Suga & O'Neill 1979, Suga 1990). The body surface is mapped in the somatosensory map (Kaas, Nelson, Sur, Lin & Merzenich 1979, Kaas, Merzenich & Killackey 1983) and the corresponding motor map records voluntary muscle control actions (Murphy, Kwan, MacKay & Wong 1977). Other higher-order maps include areas that seem to be organised according to categories and semantic values (Caramazza 1988, Goodglass, Wingfield, Hyde & Theurkauf 1986, McCarthy & Warrington 1988, McKenna & Warrington 1978, Warrington 1975, Warrington & McCarthy 1983, Warrington & McCarthy 1987, Warrington & Shallice 1984). See also Kaas (1995) for a discussion on the reorganisation and plasticity of sensory and motor maps. In this light, it would appear that the use of topological maps in a cognitive model is indeed justified, if not required.

As well, the arrangement proposed lends itself well to a hardware implementation which, as stated elsewhere in this thesis, is what we maintain is required before full-scale and realistic cognitive behaviour can be realised.

To learn a temporal sequence, consider the arrangement shown in Figure 2.9. Here two self-organising maps (we use the Kohonen mapping algorithm in simulations) are combined together, with the second having a recurrent connection back to itself. This arrangement is similar in principle to that used by Elman (Figure 2.6), but uses self-organised maps and Hebbian learning rather than a feed-forward net with backpropagation. The arrangement is first described in some detail.

Figure 2.10 details the components of the basic temporal learning network. The first component (a) is a standard self-organising mapping (SOM) from an input vector onto the SOM surface. The mapping is indicated by the solid arrow. Rather than being considered as separate entities, the input vectors are to be regarded as a sequence of

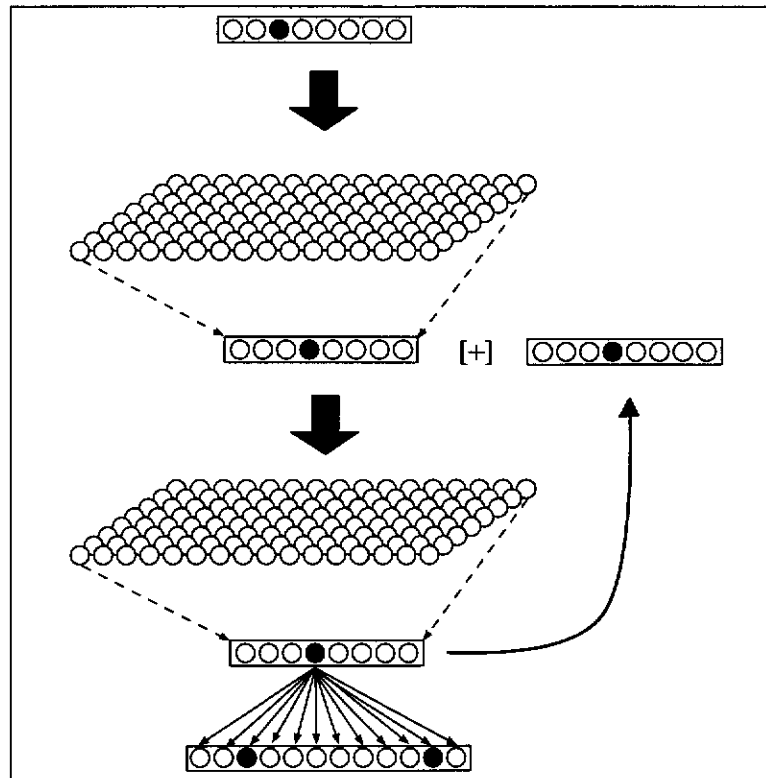


Figure 2.9: Temporal Learning Structure.

vectors v_t , $t = 0, \dots, n$. The aim of the learning algorithm is to find some temporal relationship between the v_t vectors such that $v_t = v_t(v_{t-1}, v_{t-2}, v_{t-3}, \dots, v_{t-p})$ where $p \leq n$.

The output from the initial SOM is ‘vectorised’ as indicated as Figure 2.10 (b). The process of ‘vectorisation’ is merely a method of obtaining a vector from an array by some process of linearising the matrix. The rows (or columns) of the array are simply stacked end-on-end to produce a vector. If the map surface is say, 10 by 10, then the length of the resulting vector will be 100. Note that the order of the vectorisation process is not relevant as the algorithm is order independent.

The vector that is mapped onto the second SOM surface is made by forming a vector concatenation of two other vectors—the ‘vectorised’ output from the first map, and the ‘vectorised’ output from the second map on the previous epoch—Figure 2.10 (c). That is, if we assume time-locking of steps in the process, and the time of the current input to the first SOM surface is given by t_q , then the relevant output from the first SOM

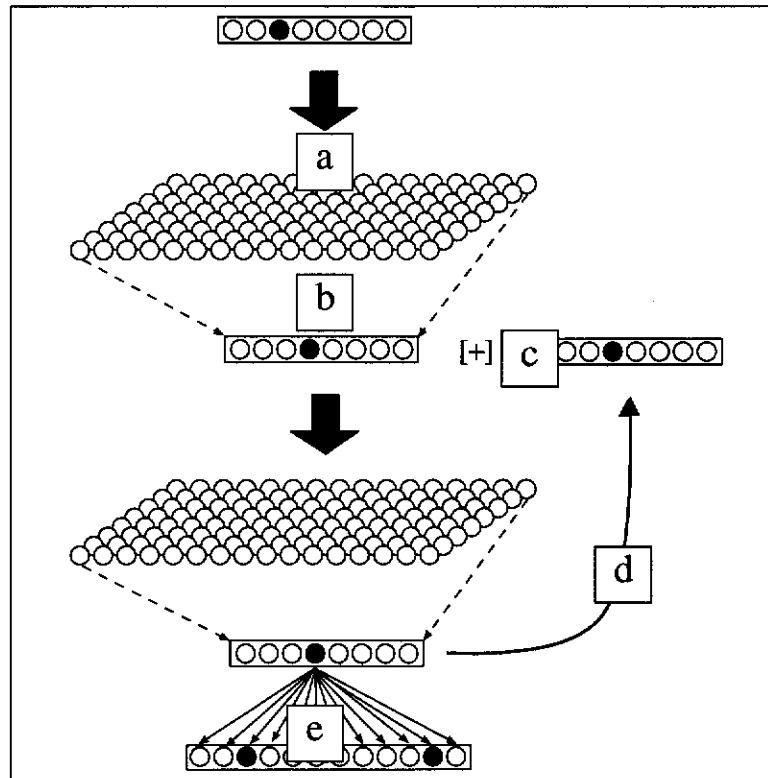


Figure 2.10: Temporal Learning Components.

surface will have been formed at t_{q-1} , and the recurrent output vector from the second map will have been formed at t_{q-2} . The new total vector forms the input to the next stage—mapping to the second SOM surface.

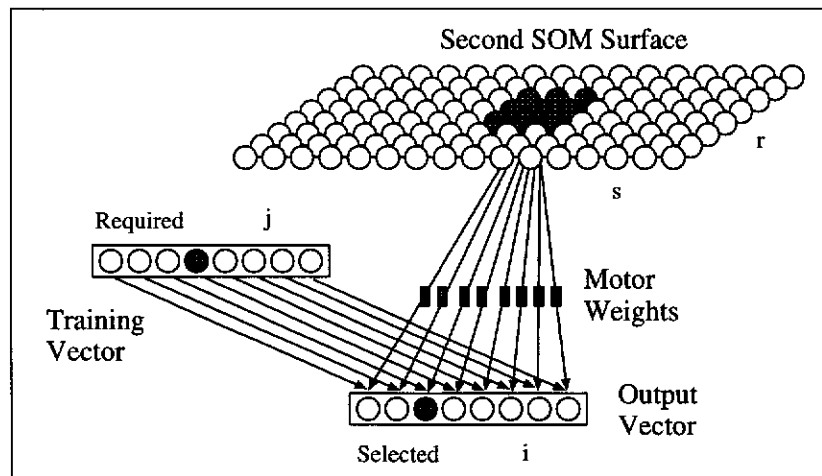


Figure 2.11: Hebbian Update of Motor Control Weights.

The ‘vectorised’ output of second SOM surface is recurrently linked back to its inputs via the link shown in Figure 2.10 (d).

The final component of this initial temporal configuration is a weighted, fully-connected layer that is updated via Hebbian learning—Figure 2.10 (e). This component learns each output vector that is to be associated with the corresponding input vectors.

The network as shown in Figure 2.10 is useful diagrammatically in that it makes it easier to understand the overall processes. In fact, the vectors as shown in Figure 2.10 are not really necessary. For example, in the biological version of the system, the outputs of the first SOM connect directly to those of the second, without the intermediate vector of neurons as shown in Figure 2.10. The actual connections are shown in Figure 2.12 alongside the corresponding diagrammatic form. †

The ABC system is based on ‘matrix’ or ‘array’ transformations—essentially transforming one SOM surface into the next. However, it is sometimes easier to conceptualise the model if, at times, we think of these arrays as vectors. This enables us to better fit the ideas expressed about the ABC model into the current literature on artificial neural networks (ANN).

The SOM surface weights are updated according to the standard Kohonen algorithm (see, for example, Kohonen 1982, Kohonen 1984, Kangas et al. 1989, Kohonen 1990, Ritter, Martinetz & Schulten 1992). The weights between the outputs of the final SOM and the output vector (which we shall refer to as *motor* weights or *motor action* weights for reasons that will soon become apparent) are modified via Hebbian learning to reproduce the required ‡ output vectors. The Hebbian update of the motor weights is illustrated in Figure 2.11. This figure shows the actual output vector connected directly to the Kohonen surface and not via an intermediate vector as in Figure 2.10

†Of course, in the computational simulation of this process, temporary vectors representing the additional vectors were used for convenience.

‡We made the point earlier that there is no such thing as ‘required’ behaviour for biological entities, other than in the general sense of needing to find food, shelter and so on. As we see later, ‘required’ here means that the model learns to reproduce certain vectors that are supplied internally, a process of self-supervision. In this sense, the internally supplied vectors are ‘required’ to be learned by some mechanism.

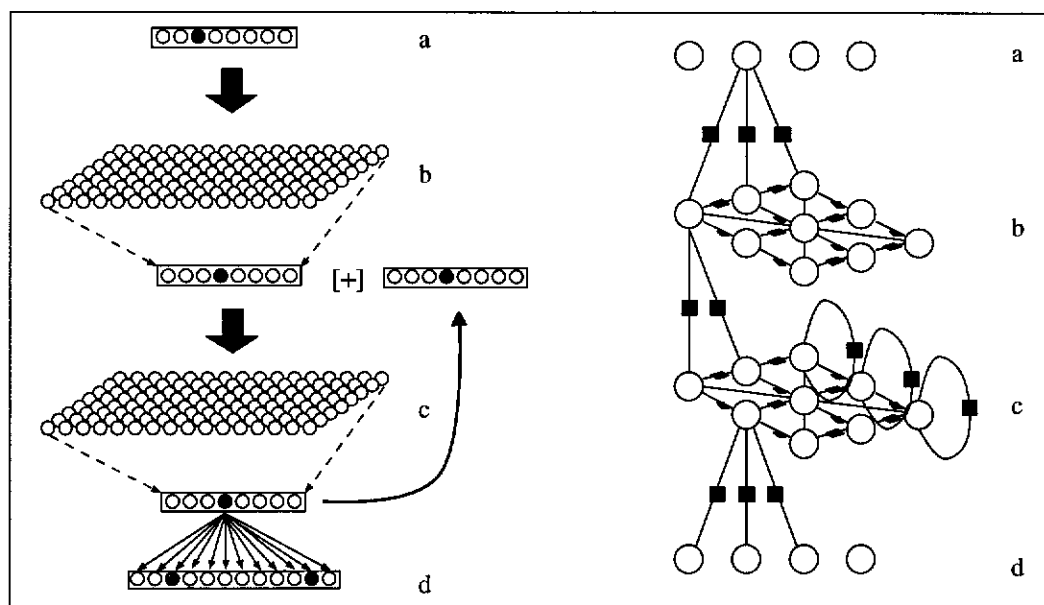


Figure 2.12: Actual Neuron Layout and Connections.

(e).

The connections shown from the training vector to the output vector are not weighted, and the update algorithm is described in Algorithm 1 for the case of reinforced learning, and Algorithm 2 for the case of self-supervised learning. [†]

The actual neural connections within the model are shown in Figure 2.12. These are indicated in comparison with the diagrammatic representation described previously. The connections within each self-organising layer are only representative, and may even require another layer for some implementations of the self-organisation method using physical neurons. However, the figure does give an indication of the neural connections between the components.

2.3.1 Temporal Learning Algorithms

The overall method is explained in the previous section. Here, a number of algorithms for updating the motor weights are outlined.

The first algorithm is used if a *local* representation is required for the output vector;

[†]The term self-supervised is explained more fully in Section B.2.3.1.

that is, for each output vector, only *one* element is to be non-zero. This method is termed *supervised* Hebbian learning in line with previous practice. However, within the ABC model, the differences between the traditional supervised and unsupervised learning modes become less pronounced, as will be made clear in subsequent sections.

The Supervised algorithm only updates those weights that are connected to the winning node and its active neighbourhood (as per the Kohonen algorithm). It calculates the maximum output vector element m_{sel} from

$$m_{sel} = k_{max} = \arg \left(\max \sum_{i,j} O_{i,j}^{(2)} \times w_{k,i,j} \right) \quad (2.7)$$

Algorithm 1: Supervised Hebbian Learning of Motor Action.

```

for all cells  $r$  within SOM surface do
  if cell  $r$  within winning neighbourhood then
    // SOM cell firing
    select winning motor action  $m_{sel}$ 
    if selected motor action  $m_{sel}$  = required motor action  $m_{req}$  then
      // correct selection of motor action
      for all  $i$  do
        if  $i = sel$  then
          increase  $w_{ri}$  (large gain)
        else
          decrease  $w_{ri}$  (small gain)
        end if
      end for
    else if selected motor action  $m_{sel} \neq$  required motor action  $m_{req}$  then
      // incorrect selection of motor action
      for all  $i$  do
        if  $i = i_{sel}$  then
          decrease  $w_{ri}$  (large gain)
        else if  $i = i_{req}$  then
          increase  $w_{ri}$  (large gain)
        else
          decrease  $w_{ri}$  (small gain)
        end if
      end for
    end if
  else if cell  $r$  not within winning radius then
    // SOM cell not firing
    no update required
  end if
end for

```

where $O_{i,j}^{(2)}$ is the (i, j) element of the output matrix from the second SOM surface, and $w_{k,i,j}$ is the motor weight between this same element and the k th element of the output vector. If the selected motor vector element is the same as the required element as specified in the supervised teacher vector, the weights are further enhanced. Otherwise, the selected element weights must be decreased, while those to the required element are increased.

The second algorithm is appropriate if a *distributed* representation is used on the teacher vector. The method is termed *self-supervised* as discussed in Section B.2.3. The algorithm is appropriate if the teacher vector elements are essentially *binary*—that is, although the elements are real numbers, they only take on values of 0.0 or 1.0. The output vector element values may take on intermediate values during learning.

A variation of Algorithm 2 for the case of non-binary teacher signals is shown in Algorithm 3. A discussion of Hebbian learning is given in Section 2.3.4.

In general then, the first SOM surface forms a topological map of the input vectors (that is, of the input words, characters, spatial frequencies, and so on, depending upon the application), with similar input vectors then being found closer together. It therefore learns to order the input vectors into some topological order.

Algorithm 2: Self-supervised Hebbian Learning of Motor Action.

```

for all cells  $r$  within SOM surface do
  if cell  $r$  within winning neighbourhood then
    // SOM cell firing
    for all  $i$  do
      if training signal $_i$  > threshold then
        // training node firing
        increase  $w_{ri}$ 
      else
        // training node not firing
        decrease  $w_{ri}$ 
      end if
    end for
  else if cell  $r$  not within winning radius then
    // SOM cell not firing
    no update required
  end if
end for

```

Algorithm 3: Self-supervised Hebbian Learning of Motor Action—2.

```

for all cells  $r$  within SOM surface do
  if cell  $r$  within winning neighbourhood then
    // SOM cell firing
    for all  $i$  do
      if  $\text{training signal}_i > \text{output}_r$  then
        // training node firing
        increase  $w_{ri}$ 
      else
        // training node not firing
        decrease  $w_{ri}$ 
      end if
    end for
  else if cell  $r$  not within winning radius then
    // SOM cell not firing
    no update required
  end if
end for

```

The input vector to the second map is made up of two components: a topological ordering of the input vectors, and a recurrent component being a topological ordering of the context transitions.

It could be argued that the first SOM mapping is somewhat redundant, as the net result is simply to produce another vector to replace the input vector. However, the initial layer is necessary to provide a degree of ‘fuzziness’ in the learning of the input vectors, especially in the case of temporal sequence production as described in Section 2.4.3. Moreover, the self-organisation and topological organisation produced on the first SOM allows the second SOM to undergo certain generalisations.

2.3.2 Mathematical Description of the Network

In this section we provide a formal description of the ABC temporal model, but make no attempt to examine the dynamic behaviour of the system. The reasons for this are several:

- the overall complication of the temporal model means that the task would require undue attention,

- the use of the Kohonen algorithm on each of the SOM surfaces—this algorithm has not yet been analysed in an analytic form, except in the 1D case (Cottrell, Fort & Pagès 1994, Fort & Pagès 1996, Flanagan 1996, Csabai & Geszti 1992). An alternate analytic self-organising method might have been substituted, but this was not attempted.

The ABC temporal model may be described in terms of the scalar variables in Table 2.2 as follows:

e_k	: external input vector element k
n_k	: number of external input vector elements
$i_k^{(s)}$: SOM surface s input vector element k
$y_{i,j}^{(s)}$: internal state of neuron (i, j) on SOM surface s
$x_{i,j}^{(s)}$: output of neuron (i, j) on SOM surface s
$u_{i,j}^{(s)}$: winning node SOM surface s
$\omega_{i,j,k}^{(s)}$: SOM surface s weights for node (i, j) element k
$\omega_{i,j,k}^{(2r)}$: recurrent weights for node (i, j) element k SOM surface 2
$z_q^{(s)}$: SOM surface s linearised output vector element q
$\omega_{i,j,k}^m$: motor weights SOM surface s node (i, j) to motor element k
o_p	: output vector element p
$l_{i,j}^{(s)}(t, u_{i,j}^{(s)})$: learning rate function SOM surface s
$l_+^{(s)}(t)$: learning rate within winning radius SOM surface s
$l_-^{(s)}(t)$: learning rate outside winning radius SOM surface s
$r_{i,j}^{(s)}(t, u_{i,j}^{(s)})$: winning radius inclusion function SOM surface s
$r_+^{(s)}(t)$: winning radius SOM surface s
$n_{(s)}$: number of rows/columns SOM surface s (assume square array)
d_s	: dimension of input vector SOM surface s
n_p	: number of motor output vector elements
\oplus	: vector concatenation
$ x $: absolute value of x
$\ (i, j)\ $: distance function
$\sigma^2/2$: variance of Gaussian

Table 2.2: List of Symbols.

First the initial SOM surface satisfies the relations:

$$i_k^{(1)}(t) = e_k(t) \tag{2.8}$$

$$d_1 = n_k \tag{2.9}$$

$$y_{i,j}^{(1)}(t) = \frac{\sum_k |i_k^{(1)}(t) - \omega_{i,j,k}^{(1)}(t)|}{d_1} \tag{2.10}$$

$$x_{i,j}^{(1)}(t) = \exp\left(-\frac{y_{i,j}^{(1)}(t) \times y_{i,j}^{(1)}(t)}{\sigma^2}\right) \tag{2.11}$$

$$u_{i,j}^{(1)}(t) = \max x_{i,j}^{(1)}(t) \tag{2.12}$$

$$z_q^{(1)}(t) = x_{(i+n_{(1)} \times j)}^{(1)}(t) \tag{2.13}$$

$$\omega_{i,j,k}^{(1)}(t) = \omega_{i,j,k}^{(1)} + l^{(1)}(t, u_{i,j}^{(1)}) \times |i_k^{(1)}(t) - \omega_{i,j,k}^{(1)}| \tag{2.14}$$

The second SOM surface satisfies the relations:

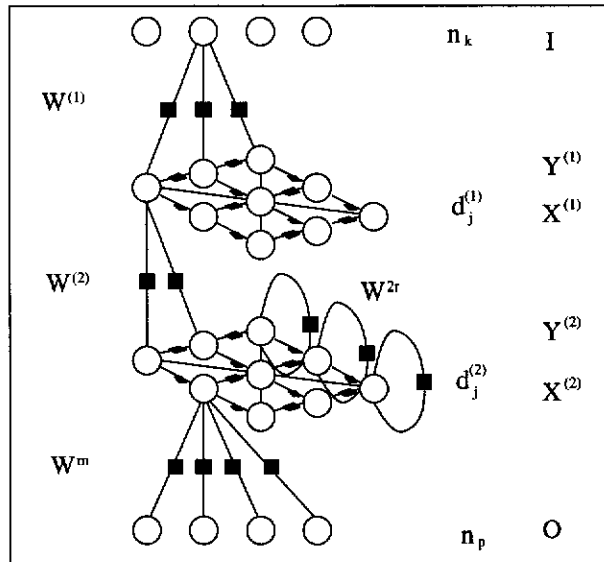


Figure 2.13: Mathematical Symbols Located.

$$i_u^{(2)}(t) = z_q^{(1)}(t) \oplus z_v^{(2)}(t-1) \quad (2.15)$$

$$d_2 = n_{(1)} \times n_{(1)} + n_{(2)} \times n_{(2)} \quad (2.16)$$

$$y_{i,j}^{(2)}(t) = \frac{\sum_u |i_u^{(2)}(t) - \omega_{i,j,u}^{(2)}(t)|}{d_2} \quad (2.17)$$

$$x_{i,j}^{(2)}(t) = \exp\left(-\frac{y_{i,j}^{(2)}(t) \times y_{i,j}^{(2)}(t)}{\sigma^2}\right) \quad (2.18)$$

$$u_{i,j}^{(2)}(t) = \max x_{i,j}^{(2)}(t) \quad (2.19)$$

$$z_v^{(2)}(t) = x_{(i+n_{(2)} \times j)}^{(2)}(t) \quad (2.20)$$

$$\omega_{i,j,u}^{(2)}(t) = \omega_{i,j,u}^{(2)} + l^{(2)}(t, u_{i,j}^{(2)}) \times |i_u^{(2)}(t) - \omega_{i,j,u}^{(2)}(t)| \quad (2.21)$$

$$\omega_{i,j,u}^{(2r)}(t) = 1 \quad (2.22)$$

where (as per Table 2.2), $z_v^{(2)}(t)$ is the v th element of the linearised output of the second SOM surface.

The recurrent weights are set to unity:

$$\omega_{i,j,u}^{(2r)}(t) = 1 \quad (2.23)$$

The output motor vector is calculated from:

$$o_p(t) = \frac{\sum_{i,j} r_{i,j}^{(2)}(t, u_{i,j}^{(2)}) \times x_{i,j}^{(2)}(t) \times \omega_{i,j,p}^m}{\sum_{i,j} r_{i,j}^{(2)}(t, u_{i,j}^{(2)})} \quad (2.24)$$

The winning radius inclusion function determines whether a particular node on the surface is within the winning radius or not. It is a function of the winning node, and of time.

$$r_{i,j}^{(s)}(t, u_{i,j}^{(s)}) = \begin{cases} 1 & \text{if } \|u_{i,j}^{(s)} - (i, j)\| \leq r_+^{(s)}(t) \\ 0 & \text{otherwise} \end{cases} \quad (2.25)$$

That is, the node is included in the winning radius if $r_{i,j}^{(s)} = 1$.

The distance function $\|a_{i,j} - b_{r,s}\|$ gives the Manhattan distances between two nodes a and b at indices (i, j) and (r, s) respectively. That is,

$$\|a_{i,j} - b_{r,s}\| \leq r_c \equiv \begin{cases} |i - r| \leq r_c & \text{and} \\ |j - s| \leq r_c \end{cases} \quad (2.26)$$

The learning rate of the Kohonen algorithm is a decreasing function of time.

$$r_{max} \geq r_+^{(s)}(t) \geq r_+^{(s)}(t+1) \geq 0 \quad (2.27)$$

The weight update for each SOM surface is dependent on the winning radius.

$$l_{i,j}^{(s)}(t, u_{i,j}^{(s)}) = \begin{cases} l_+^{(s)}(t) & \text{if } r_{i,j}^{(s)}(t, u_{i,j}^{(s)}) = 1 \\ l_-^{(s)}(t) & \text{otherwise} \end{cases} \quad (2.28)$$

Both the winning and the losing learn rates are decreasing functions of time.

$$l_+^{max} \geq l_+^{(s)}(t) \geq l_+^{(s)}(t+1) \geq 0 \quad (2.29)$$

$$l_-^{max} \geq l_-^{(s)}(t) \geq l_-^{(s)}(t+1) \geq 0 \quad (2.30)$$

$$(2.31)$$

Further, the winner learning rate is greater than the loser learning rate.

$$l_+^{(s)}(t) > l_-^{(s)}(t) \quad (2.32)$$

The update to the motor weights covers two cases—local and distributed representations (as discussed in Algorithms 1 and 2. For the distributed representation (appropriate for the self-supervised method discussed in Section B.2.3)

$$\omega_{i,j,k}^m(t+1) = r_{i,j}^{(2)} \times \omega_{i,j,k}^m(t) \times (1 + \delta(t_k - c_0) \times g) \quad (2.33)$$

where

$$\delta(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{otherwise} \end{cases} \quad (2.34)$$

c_0 is a cutoff, and g the gain.

The first term on the right-hand side ($r_{i,j}^{(2)}$) restricts the calculation to only those nodes which are within the winning radius of the second SOM surface. The function $\delta(x)$ forces the weight to the particular element to be increased or decreased depending on the value of the corresponding teacher/reinforcement vector element t_k .

For the local representation case

$$\omega_{i,j,k}^m(t+1) = r_{i,j}^{(2)} \times \omega_{i,j,k}^m(t) \times (1 + \mu(k, k_{sel}, k_{req})) \quad (2.35)$$

where

$$\mu(k, k_{sel}, k_{req}) = \begin{cases} \text{if } k_{sel} = k_{req} & \begin{cases} g_L & \text{if } k = k_{sel} \\ -g_S & \text{if } k \neq k_{sel} \end{cases} \\ \text{if } k_{sel} \neq k_{req} & \begin{cases} -g_L & \text{if } k = k_{sel} \\ g_L & \text{if } k = k_{req} \\ -g_S & \text{otherwise} \end{cases} \end{cases} \quad (2.36)$$

The first bracketed component represents the case for which the system has correctly selected the required motor response. Thus the selected element is enhanced by a large gain (g_L) while all other elements are reduced by a small gain (g_S). The second

case is that in which the system has incorrectly selected the required motor action element. Here the selected action weight is suppressed and the required action weight is increased, both by g_L , while all other weight components are suppressed by g_S .

For further formal analysis of the Kohonen algorithm, see Fort & Pagès (1996), Flanagan (1996), and Cottrell et al. (1994). Also see Amari (1990), Ritter & Schulten (1988), Zhang (1989), Tanaka (1990), and Takeuchi & Amari (1979).

The analysis of discrete-time recurrent neural networks has been undertaken by, among others, Amari (1990), Chen & Aihara (1997), and Doya (1994).

2.3.3 Kohonen Algorithm—Winner-Take-All

It was initially assumed that the output nodes emanating from each SOM surface should conform to a winner-take-all (WTA) scheme—that is, the output of all nodes other than those within the winning radius should be set to zero as a result of inhibitory connections within the SOM layer.

However, this mechanism proved to be problematic as the resultant vectors (to be learned by the following map) proved to be too sparse. Although the method was shown to work, learning times and accuracy suffered as a consequence of the sparse vectors. The system also showed large sensitivity to various parameters.

In a discussion of lateral interactions between cells, Kohonen (1982) suggested that

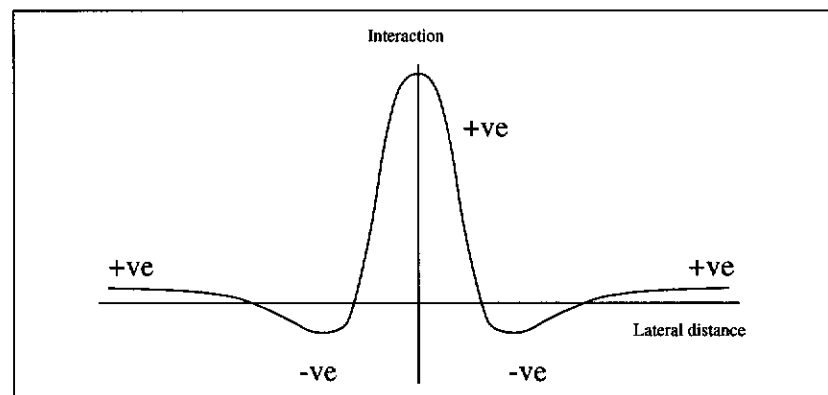


Figure 2.14: Lateral Interactions.

there is both anatomical and physiological evidence for an interactive profile as shown in Figure 2.14. This is the standard ‘Mexican hat’ shape of lateral interactions. The central short-range lateral excitatory region has a radius of between 50 to 100 μm in primates. The surrounding inhibitory region has a radius of 200 to 500 μm , while the weaker excitatory region which surrounds the inhibitory pen-umbra may have a radius of up to several centimetres.

While packing densities of neurons within the cerebral cortex vary depending upon the layer or region involved, we can make a rough calculation by considering that of the granular layer. This is given by Carpenter (1991, page 228) as between 3 and 7 million cells per cubic millimetre, which is extremely tight packing. This gives an inter-neural separation of roughly 50 μm . Thus a very rough calculation suggests that the excitatory interaction extends only to a one or two neuron radius, whereas the inhibitory interaction extends to a radius of perhaps 4 to 10 neurons. The weak excitatory interaction may extend over a radius of several hundred neurons.

Thus most neurons on a large Kohonen sheet will tend to be weakly excited by the winning node, rather than inhibited as required by WTA.

This rather loose analysis was used as justification for a learning and processing method that worked extremely well. Through experimentation it was found that a very efficient method was to use the winning radius associated with the standard Kohonen algorithm when updating the appropriate weights, but to ignore any excitation or inhibition in calculating the outputs from each SOM surface. Thus the output of each Kohonen node was simply a function of its activity—which was in turn a function of the difference between its associated vector and the shared map input vector.

This allowed the learning mechanism in the following map to access a full compliment of vector elements rather than the sparse vectors as previously. The method proved to be efficient and accurate. Future research will re-examine this issue of lateral interactions within the model.

2.3.4 Extensions to Hebbian Learning

The original Hebb rule only allowed for an *increase* in synaptic weights. This rule was unstable, and the synaptic weights could only be driven to their maximal value.

The algorithm here uses an extension to the standard Hebbian learning in that, as well as the weights being increased for temporally synchronised pre- and post-synaptic activity, weights are also *decreased* when the pre- and post-synaptic neurons are not temporally correlated.

A generalised Hebbian rule may then be stated as (Shouval & Perrone 1995):

$$\Delta\omega_i = -\eta(x_i - x_0)(y - y_0) \quad (2.37)$$

where ω_i is the synaptic weight between the pre-synaptic neuron i and the post-synaptic neuron. x_i and y are the pre- and post-synaptic neuron activities, and η , x_0 and y_0 are constants.

2.3.5 Differential Learning Periods

The learning of a mapping onto each SOM surface requires that an appropriate learning rate and neighbourhood radius be altered over a number of epochs.

To enable stability of the algorithm, different periods of learning are used in updating the weights of each SOM surface, and for the motor action weights. For example, the learning on the first map (SOM1) may extend for 100 epochs, after which time no further learning takes place for these weights (learning rate = 0, learning radius = 0). This means that the input vectors for this map must have been separated and distributed over the surface neurons within this period. The second map (SOM2) will need to learn for a longer period, say for 200 epochs. The stability of the output from SOM1 after epoch 100 ensures that SOM2 is able to learn to separate and distribute the combined vectors over the surface neurons at least from epoch 100 on. The motor action weights, in turn, may need to be trained over say 300 epochs, as the stability of

the output from SOM2 is not found until say epoch 150 onwards.

It was typically found that if SOM1 was trained for 100 epochs, SOM2 for 200 epochs, and the motor action weights for 300 epochs, then the system was able to learn the data sequences as required. Increasing the number of epochs did not alter the final learned result much. Of course the optimal number of epochs would vary according to the actual temporal learning problem, but it was not considered relevant to obtain these optimal training times. Instead, a standard training regime of 100, 200 and 300 epochs was generally used. ⁷

There is some biological evidence for this differentiation of learning periods. For example, humans seem to learn phonemes before they acquire words. Bjorklund (1989, page 82) cites research by Colombo (1986) and Walley, Pisoni & Aslin (1981) in which young infants learn to make the sound discriminations of their language culture. The infants both lose some initial discriminations that they are able to make but which are not in the adult language, and as well gain some discriminations that they were initially unable to make but that are able to be made by the adults. This separation of (at least some) phonetic contrasts into cultural groups occurs in 6- to 8-month-old infants, and is very suggestive of a self-organisation of the input sounds onto an initial 'phoneme' map.

In contrast, although most 10- to 12-month-old children have started uttering individual words, it is not until 18 months or so that they put two words together into a sentence (Bjorklund 1989, page 108). The stability of the sound discrimination mapping appears to be in place before word discriminations and associations are made.

2.3.6 Refractory Period and Winner Exclusion

To remember a sequence of vectors, the recurrent SOM in particular must select a *unique node* for each *unique component* of the sequence (and in general for all other unique components in all other sequences) in order to be able to follow the correct path through the maps. Failure to do so gives *cross-talk*—the same winning node being used for different inputs.

To prevent the SOM surface nodes from being used more than once during a learning epoch, (and thus giving cross-talk and errors), two methods were used.

All neurons in the brain exhibit a *refractory period*—a period following the firing of the cell during which it is unable to recover and fire again (Milton, Mundel, an der Heiden, Spire & Cowan 1995). There is an *absolute* refractory period, during which time no amount of voltage can re-excite the node (this period lasts for approximately 1-3 msec), and a following *relative* refractory period during which the neuron can be made to fire but only with a higher input than when the neuron is at rest. This period lasts from 5-200 msec. This inability of a neuron to fire could be expected to have a nontrivial influence on the network dynamics as the firing threshold depends upon the time elapsed since the neuron fired—in other words, it records some information about the neuron's temporal history.

To simulate a refractory period, a simple method was used. If, in processing a SOM surface, a particular node is chosen as a winning node, then that node is not able to be used again as a winning node during the learning of the *next input* (but not the whole epoch). After waiting for one (or more) input periods, (during which time the node's output is set to zero), it can then recover and take part in subsequent processing. Thus, although a node may not be a winner for two (or more) successive inputs, it can be a multiple winner over one epoch.

An extension to this scheme gave a second method—an *exclusion* method. Once a node has been selected as a winner during one epoch, it can no longer be selected as a winner for the duration of that *epoch*. This method ensures that all input vectors select unique winning nodes on each epoch.⁸

2.3.7 Input Data

In most cases, the input data strings were formed into one *contiguous* stream of 'symbols' (representing, say, characters or words) with the final character (word) wrapped around to precede the first on the next pass.

For example, if the sequence to be learned is

```
cat c a t  
dog d o g
```

then the continuous string presented to the system is

```
cat c a t dog d o g cat c a t . . .
```

In most cases, the input symbols (words of a sequences or characters within words, as the case may be) were arbitrarily assigned random (real-binary) vectors. These vectors were selected using a pseudo-random number generator, while ensuring at the same time that the Hamming distances between all vectors in the full input symbol set exceeded a specified number. For example, if the length of the input vector was set at 10, each chosen vector was usually required to have a Hamming distance of at least 3 from all other vectors in the set selected so far before it could be selected and added to the set. This was to ensure that spurious similarities in the random vectors were minimised, and to ensure that the input data vectors are able to be spread across the initial SOM in a reasonably even distribution.

While many of the examples in this chapter deal with language-based inputs such as characters or words, it must be emphasised that in the full ABC model, the vector input values are derived from exteroceptive (sensory) filters, as well as internal proprioceptive and interoceptive sources. These vector values are not semantically interpretable in the sense that words and characters are, and the ABC model takes a strong stand on the view that cognition is not semantically (or symbol) based. However, the same principles of temporal learning apply—a series of consistent inputs will be learned as a temporal sequence, and the sequence will provide the basis for later recognition (and reproduction) of the originating object, concept or event.

This point is emphasised in Chapter 4 in which we discuss a new theory of vision based on the ABC model. In this case the inputs are provided by various visual filters including spatial frequencies and angles, colour and motion, and the consistency is provided by a learned saccade sequence (scanpath).

Another point to emphasise is that *similar* inputs will fall on near neighbour locations on the SOM surfaces, providing a form of generalisation of the input data.

2.4 Temporal Learning Experiments

Most of the temporal learning experiments performed with the ABC temporal learning model are described and discussed in Appendix B. In this section we first provide a summary of the results discussed in that appendix, and then look at a small number of key experiments which illuminate the method, and which provide important concepts which are required for later discussions.

2.4.1 Summary of Temporal Learning Results

We present a summary of the results discussed in Appendix B. These results indicate that the temporal learning capabilities of the ABC model are extremely good. The results in fact suggest a mechanism that is better than the SRN mechanism for a number of reasons, some of which are discussed here while others are mentioned in the appendix.

reduced model & explicit training

We explore a reduced temporal model in Section B.1 on page 384, and show that it is able to be trained, using explicit state transitions, to learn various Finite State Automata (FSA) grammars.

‘exact’ learning

The additional generalisation on the hidden layers of the ABC temporal model means that it is able to learn the FSA grammars *exactly*, whereas the SRN model of Elman is usually only able to learn the grammar to a certain statistical approximation.

The model forces the learning mechanism to select a particular winning node, and this form of generalisation ensures that the same node is used for subsequent identical characters. This means that the ABC model appears to have much greater **compositionality and systematicity** than the SRN model. [†]

[†]Within linguistics, compositionality usually means that atomic symbols can be combined and molecular representations can be decomposed according to a formal syntax, and systematicity means that the atomic and the molecular symbols and the rules of syntax can be systematically assigned a meaning.

Various FSA networks were learned, including the temporal XOR and other sequences initially introduced by Elman (1990, 1993). The discussion begins on page 397 with Section B.2.

infinite clauses

The ABC model appears to be able to learn infinite loops in FSA grammars whereas the SRN model is only able to learn specific paths, as the discussion beginning on page 403 indicates. For example, Figure B.13 shows a bidirectional link in an FSA grammar that the model is shown to be able to learn.

flow diagrams

As a specific winning node is selected for each SOM surface on each pass through the temporal learning structure, the sequence of winning nodes can indicate the actual temporal sequence that have been learned. This allows for the construction of *flow diagrams* as discussed in Section B.2.3.4 on page 415.

counting

The model is shown to be able to ‘count’ in that it can learn the continuous sequence 7 a a a a a a, predicting a 7 character after six a characters. The model is able to differentiate the a/a transitions on to separate winning nodes based on the previous characters in the sequence. This is discussed on page 422 in Section B.2.5.

explanation & verification

Potential use of neural networks in industrial applications has been hampered by their so-called *black-box* nature; that is, while they may behave correctly within certain bounds there is no way to explain their actual (learned) behaviour or to prove (verify) that their behaviour is appropriate. The generalisation on the hidden layers of the ABC temporal model, in forcing a choice of winning node, may provide a means of explanation and verification as we discuss on page B.2.6.4.

semantic localisation & context

In a discussion on the learning of English embedded sentences starting on page 439 we show how the model forces identical words to be associated with the same winning nodes, thus ensuring localisation of semantic nodes. Even though a sentence may be extended, the individual words excite local nodes on each SOM, with appropriate recurrent linkages between words.

In association with *context*, which we discuss on page 442 (as well as in the current chapter), we are able to provide a mechanism for the generation of sequences and the correct application of case (number agreement, tense and so on) within extended sequences (sentences).

multiple linked maps

Beginning on page 462 in Section B.3, we show how multiple temporal learning structures may be combined into a hierarchy and trained to learn extended language tasks. These tasks include learning a context-sensitive sequence ($a^n b^n c^n$) and learning to parse a c-like computer language.

2.4.2 Learning the Alphabet

In order to test the robustness of the sequence learning, a task of learning the alphabet was given to the network. This example is also of interest in that it shows a simplified method of obtaining a vector representation from a visual object. This point is taken up in a number of later sections in a discussion of the mechanisms of the visual system. In fact the method of obtaining a vector representation for the alphabetic characters in this example can be regarded as equivalent to the method suggested in the full system. In effect, the actual mechanism of obtaining a vector to be used as a representation for a visual object (or part of an object) is not important, so long as the vector so obtained is able to discriminate (either by itself or in a sequence) the object or part.

Figure 2.15 shows a typical ‘pixel’ representation of the alphabetic letters. Each letter is given by a combination of black or white pixels in a box five pixels wide by seven pixels in depth. By simply linearising the array of pixels for each character—concatenating each of the seven rows end-to-end—we obtain a vector 35 bits long.

This is certainly not how the eye works, but the principle is equivalent for our purposes here. The vectors for the letters A, B, and C are illustrated in Figure 2.16.

The task was to teach the network to reproduce the alphabetic sequence by training it to predict the next letter in the sequence, given a particular letter; for example, given D predict E, given P predict Q, and so on.

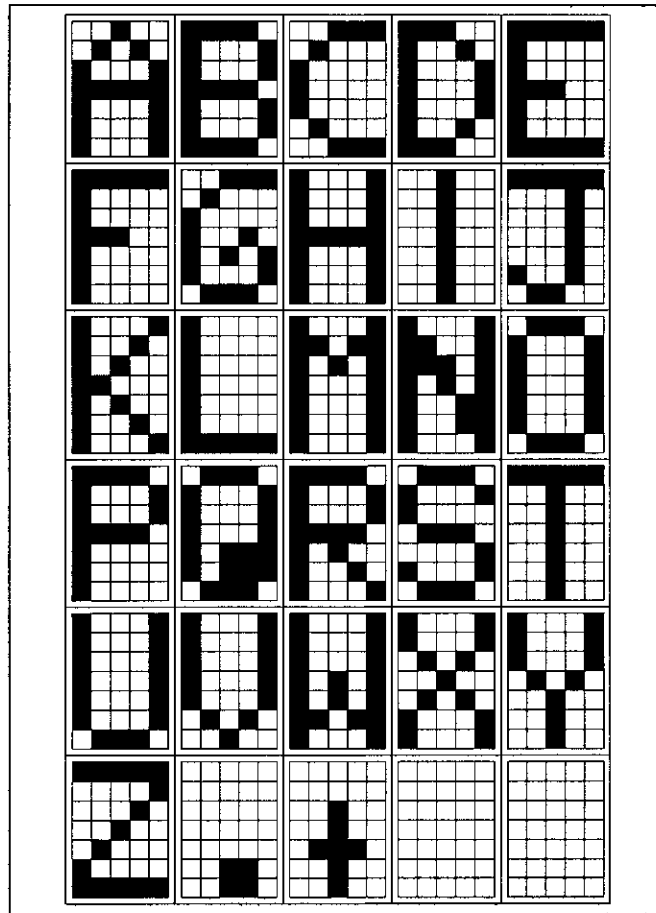


Figure 2.15: "Visual" Representation of Characters.

The training data consisted of a single string of characters

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

which was repeated end-on-end so that A followed Z to start another cycle. The transitions are thus $A \rightarrow B$, $B \rightarrow C$, \dots $Z \rightarrow A$, $A \rightarrow B$, and so on. The input

00100010101000111111100011000110001	A
11110100011000111110100011000111110	B
00111010001000010000100000100000111	C

Figure 2.16: Vector Representation of Characters.

vectors were as described above, and the system output was a 26 bit vector in which the appropriate positional bit was reinforced to indicate the next letter in the sequence. This is a *local* representation.

The network was able to learn the alphabet to 100% accuracy in the training phase.⁹

It is instructive to examine the locations of the winning nodes associated with each letter. On the first SOM surface, shown in Figure 2.17, letters which *look* similar (and hence have similar input vectors) will be located near each other on the mapping surface. For example, P and R, Q and O, E and F (as well as others) are neighbours on the surface. The letters are also well spread out over the surface, indicating that the Kohonen algorithm used is working well and is distributing the input vectors across the map.

Figure 2.18 shows the winning nodes on the second SOM surface for each transition. The first character of each pair is the character from the previous iteration whose outputs are taking part in the mapping via the recurrent loop. The second character (after the slash) is the current character. Thus A/B is the winning node for the transition A \rightarrow B, which is expected to produce a vector corresponding to the letter C as output.

Testing again used a separate program to read in the calculated weights for the two

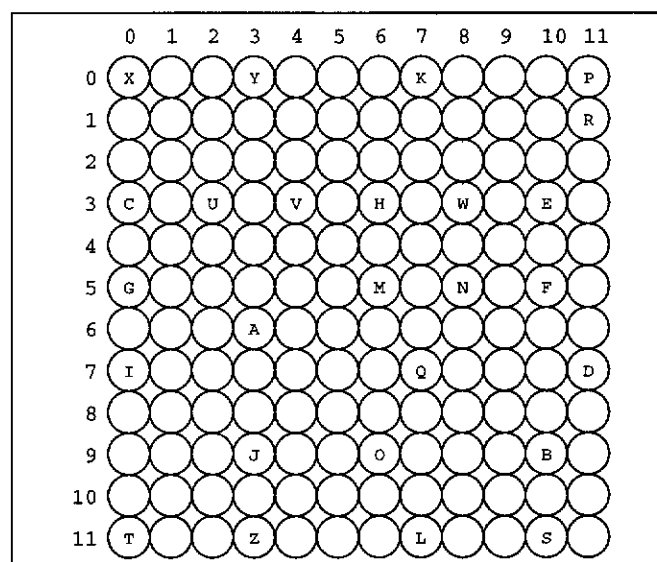


Figure 2.17: Mapping Of Alphabet On First SOM Layer.

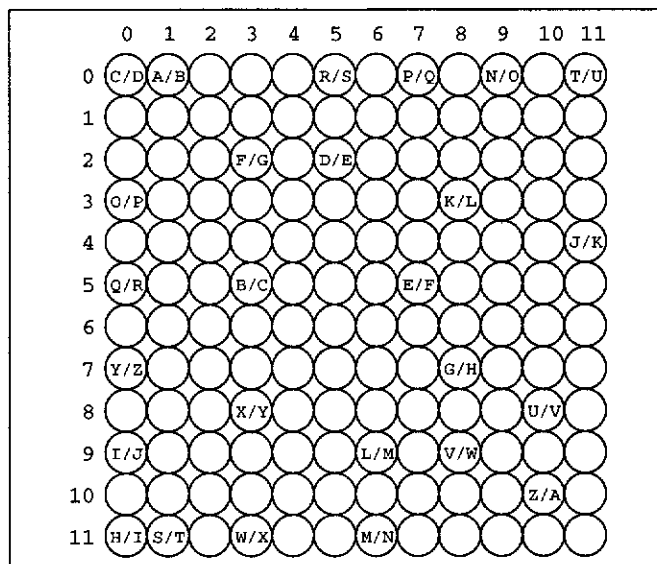


Figure 2.18: Mapping Of Alphabet Transitions On Second SOM Layer.

SOM surfaces and the motor weights. The test program fed in characters as input and checked the resultant output vectors. The output vectors matched the required (next character) vectors in 100% of the test cases.

This process as described above provides a means of learning a series of transitions, and returning the next in the sequence when any element of the sequence is given as input. The method gives us a *single* output for each input, albeit the next in the learned sequence. But for motor control, we need a mechanism whereby a whole *sequence* of outputs can be returned if a 'seed' input is provided.

2.4.3 Motor Control—Sequence Production

The process as described in the previous section allows for the learning of a set of transitions that are part of a temporal sequence. The sequence is, however, only at the input stage, and the process is passive. We have not yet succeeded in learning and *reproducing* a sequence. The network used in the previous section is shown in Figure 2.19.

The network is able to learn a sequence by recording the local context changes—the transitions between successive inputs. It produces an output vector that indicates that

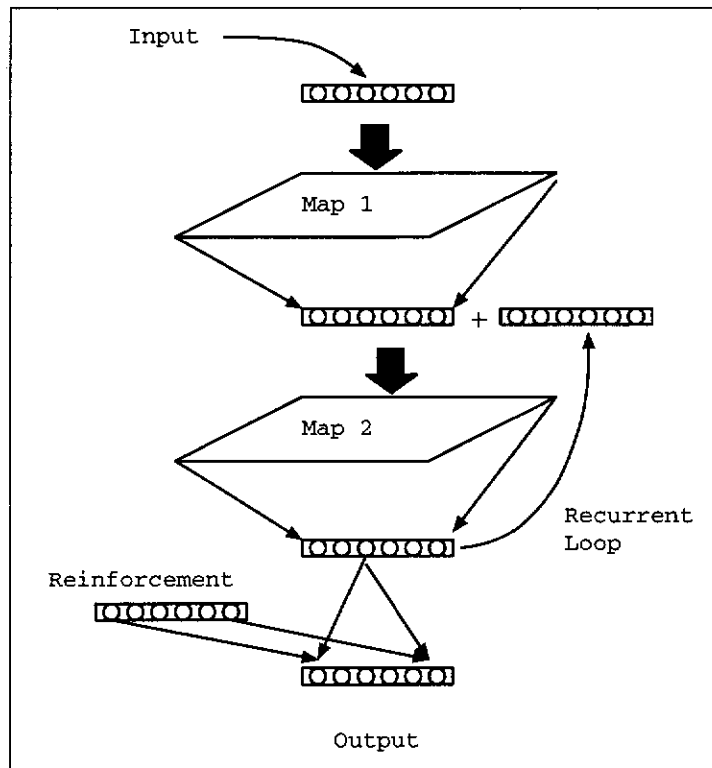


Figure 2.19: Learning a Temporal Sequence.

the next in the sequence has been learned.

But what if we were to train the network to output the *actual* vector that would have been required as input to get the *next* letter; that is, (for this alphabet example), if we feed in an vector corresponding to the letter A, the output required is a vector corresponding to the letter B, if we input a B vector, the output required is a C vector, and so on.

For example, if the input vector is 00100010101000111111100011000110001 (the letter A), then the required output for that input would be the vector representation of letter B, namely 11110100011000111110100011000111110.

We can then train the system to learn the transitions as per the previous section, and freeze the weights. Now that the transitions have been learned, we can link up the output and the input vectors for the testing phase, as is indicated in Figure 2.20.

This second recurrent link is now capable of continuing a sequence that is initiated

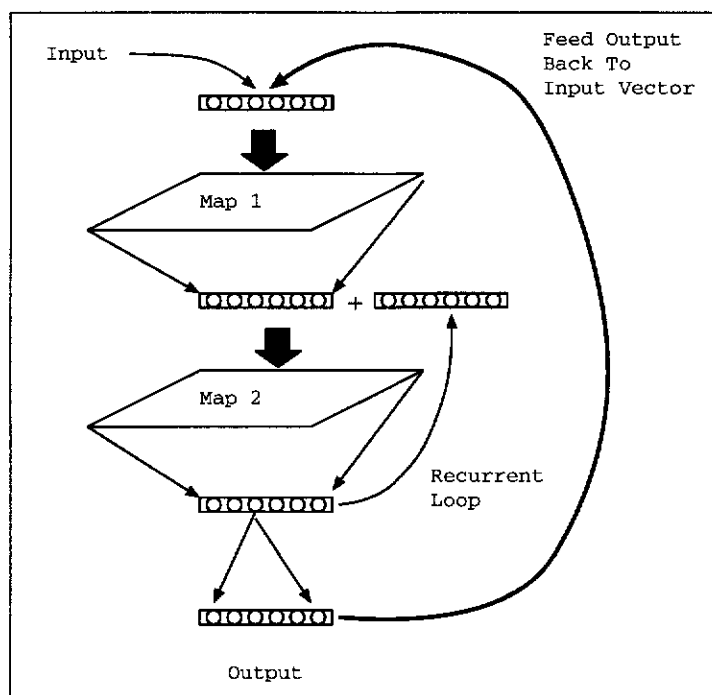


Figure 2.20: Generating a Temporal Sequence.

by a seed character; for example, an initial input of a vector corresponding to an A character will, on the first forward pass, output a B. This output will then become the next input and on the next forward pass will generate a C, the C as input will output a D, and so on.

Given the seed character, the network could ‘feed itself’ by successively feeding its output vector back as inputs. Given the vector ‘bits’ for the letter A, for example, the network would produce a vector for the next character in sequence—the vector corresponding to the letter B.

In order to equate the input and output vectors, the output vectors were turned into ‘binary’ form via a threshold—if an output vector element was less than a threshold value (usually 0.3) the element was equated to zero, whereas if the value was greater than the threshold the value was equated to one.

When this vector was fed back as the input vector, it should in turn produce the vector for the next character (C) as output, this in turn fed back to the input, and so on for a number of iterations determined by a program parameter.

The results of this self-feeding mechanism were impressive. Once the sequence was learned it proved to be very robust and enduring.

The learning was again tested with a separate program that read in the saved weights, and given a seed word, cycled through successive predictions of the network. Calculations for the progress of the vectors through the network were done as per the training phase, but no learning was performed during this testing stage. The test program was given a seed character and asked to reproduce the next n characters in sequence.

The correct alphabetic sequence could be generated using any character as seed, the sequence continuing on from the seed character. The sequence was continued for up to 500 characters from each seed without problem. The initial seeds correctly generated the required learned sequences.

Various tests of the alphabet sequence generation are described in Appendix Section B.4.1.

Sequence generation is seen as the key to understanding motor sequences, which, as we discuss later, includes language and ‘thinking’ (or ‘self-talk’). As well as learning the return vector to be used on the next cycle, the mechanism can also produce actual muscle output (as described in Figure B.9), and so perform a motor action sequence.

The mechanism shows how a neural mechanism may store a temporal sequence, not by storing the whole sequence as a memory trace in a contiguous order, but by storing a series of context transitions and linking these transitions via recurrent connections.

2.4.4 LAPS—Learning And Production of Sequences

In this section we describe a self-contained, self-organising and self-supervising mechanism for temporal learning. The overall network, which will be referenced by the acronym LAPS (Learning and Production of Sequences), is illustrated in Figure 2.21.

Input vectors $Input(t)$ are supplied to the network. These vectors are processed and flow through the LAPS system, the results of which arrive at the output of the second map at time $t + \Delta t$. By this time, the input vector has moved on to $Input(t + \Delta t)$ and

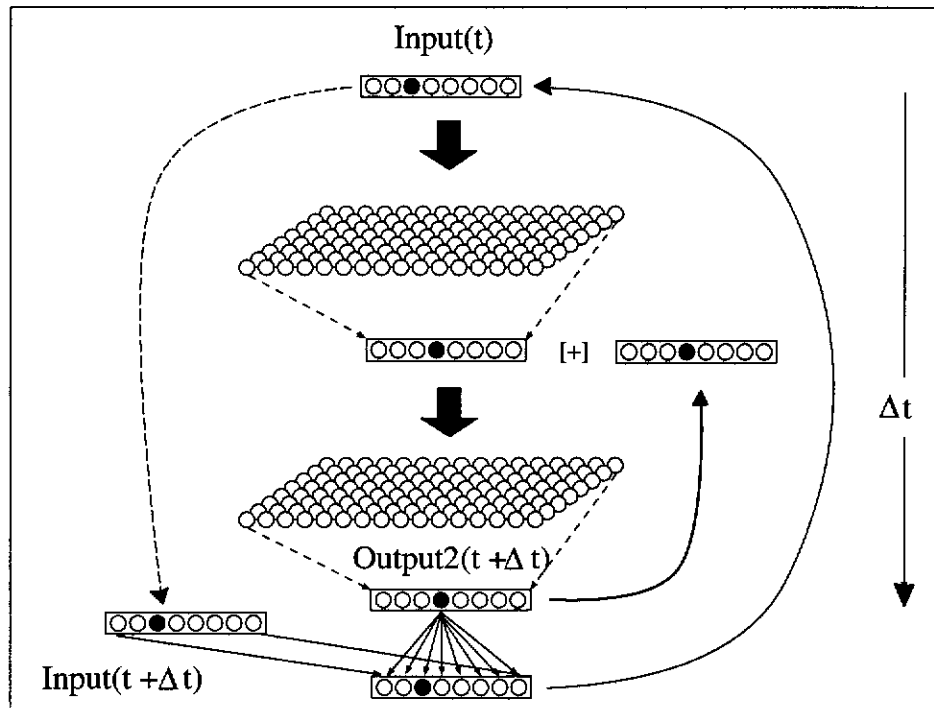


Figure 2.21: LAPS—Learning and Production of Sequences.

it is this later signal at time $t + \Delta t$ that is to be learned as the *next* of the sequence to be associated with the original input $Input(t)$.

That is, the effect of $Input(t)$ passes through the network and arrives at the outputs at the same time that $Input(t + \Delta t)$ arrives at the inputs. The new input vector may then be used to train the weights as the reinforcement signal as described previously.

To illustrate the principle, consider the case of learning the alphabet again. The vector associated with the letter A arrives at the inputs and is processed by the network to produce a set of outputs on the second map at time $t + \Delta t$ ($Output2(t + \Delta t)$). The next character in sequence, B, is then used as the teacher/reinforcement signal to update the motor weights. Thus B is learned as the successor to A.

Of course the flow of the A vector through the network to arrive in time to be associated with the B vector is artificial in this case. However, in the general case, the vectors that are associated are naturally selected by the time of transfer (Δt).

Thus the LAPS model is self-teaching and unsupervised—it just learns the temporal

sequences experienced at its inputs. All interactions except for the input sequences are internal, and it requires no external interaction or supervision—it could be said to be *self-supervising*.^{† ‡}

2.4.5 Semantic Priming

So far we have just dealt with ‘syntactic’ and statistical features of motor control. One of the proposals of this thesis is that language learning and production is just a special case of motor learning and control. * The same mechanisms of learning and production apply to both motor actions (say the learning and reproduction of muscle movements to serve a tennis ball), speech (learning and reproducing sound structures via voice muscles), and as will be discussed later, even thinking.

The idea that language can stand on syntax alone has been promoted by some linguists, but is disputed here. This issue is symptomatic of a general problem with the current approach to linguistics which treats speech and language as a serial procession of strings and does not take into account the parallelism of cognition.

The view taken in this thesis is that language and processes associated with language, (words, concepts, grammar) are learned and not innate. A full discussion of these views on language is to be found in an analysis of the implications of the ABC model in Section 5.1.

Language is learned *in context*—the symbols (labels) used to denote objects and concepts are directly associated with perceptual categories. This ‘meaning’ is directly

[†]One other connection to the outside world is relevant—the ‘value’ or emotional context. If an event is high in emotional context (such as fear, anger, sexual awareness) then the rate at which the weights are modified is increased—a higher gain. For events without high emotional context the rate would be lower—a low gain. Emotional or value context may be physiological (innate fight or flee responses) or social. Verbal criticism or praise may also increase the gain and thus provide a faster learning or recall of the sequence. This component is not included in the LAPS model at this stage.

[‡]It is also interesting to speculate on the connection between this proposed mechanism and our perception of time. More will be said on this topic when we re-examine some implications of the outer recurrent loop of the LAPS architecture and the prediction or anticipation of future events.

*This is not a new idea, and various researchers have made similar proposals. The ABC model, however, provides a mechanism.

linked to the language symbols. Language syntax does not operate in isolation from semantics, and the semantic context obtained from various sources is important in language. In fact, the two cannot be separated.

What is meant here by *context* is not just the surrounding words as is the case in many linguistic discussions. For example, an often-used pair of sentences is

- (a) The countess threw the ball.
- (b) The pitcher threw the ball.

Here the surrounding word-context *may* be important in deciding which ‘ball’ is appropriate. But in cognition, context includes *all* associated information pertaining at the time, whether it results from previous words, from visual senses, or from internal memory (general knowledge) ongoing during the hearing or seeing of the words in the sentence.

So context must not be restricted to just the previous words but is a much broader concept. Perhaps it is useful to consider the difference between *serial* context and *parallel* context. The serial context (previous words) may provide part of the parallel context. For example, if a sentence prior to sentence (a) above was

The countess was asked to throw the first ball of the world series.

then this serial context would help determine the interpretation of the subsequent sentence (a). Conversely, watching a cartoon about an animated glass with a penchant for elegant dining and dancing would give a totally different meaning to sentence (b).

In this experiment, a LAPS network is required to learn a number of very similar sentences describing simple geometrical figures. The network was trained with sentences such as THE SMALL RED TRIANGLE where the size component was one of SMALL, MEDIUM, LARGE, the colour component one of RED, BLUE, GREEN, and the shape component one of TRIANGLE, SQUARE, CIRCLE. All sentences started with the word THE. †

†The purist grammarian would call these *noun phrases*, but this detail is not relevant to the discussion.

A context/semantic component was added to the LAPS network. This semantic component consisted of an additional six 'bits' of information that was concatenated to the input vector mapped to the first SOM surface. For the moment we will simply say that this context information is provided from elsewhere in the brain. Of the six bits, two each were used to indicate the size, colour and shape of a 'perceived' object. The arrangement is shown in Figure 2.22.

Note that the semantic component is not part of the LAPS learning module—it is simply there to provide a semantic means of disambiguating the syntactic sentences. The full vector (including the semantic component) is mapped to the first SOM surface, but the semantic component is not used to train the motor weights for the outer recurrent loop of the LAPS module.

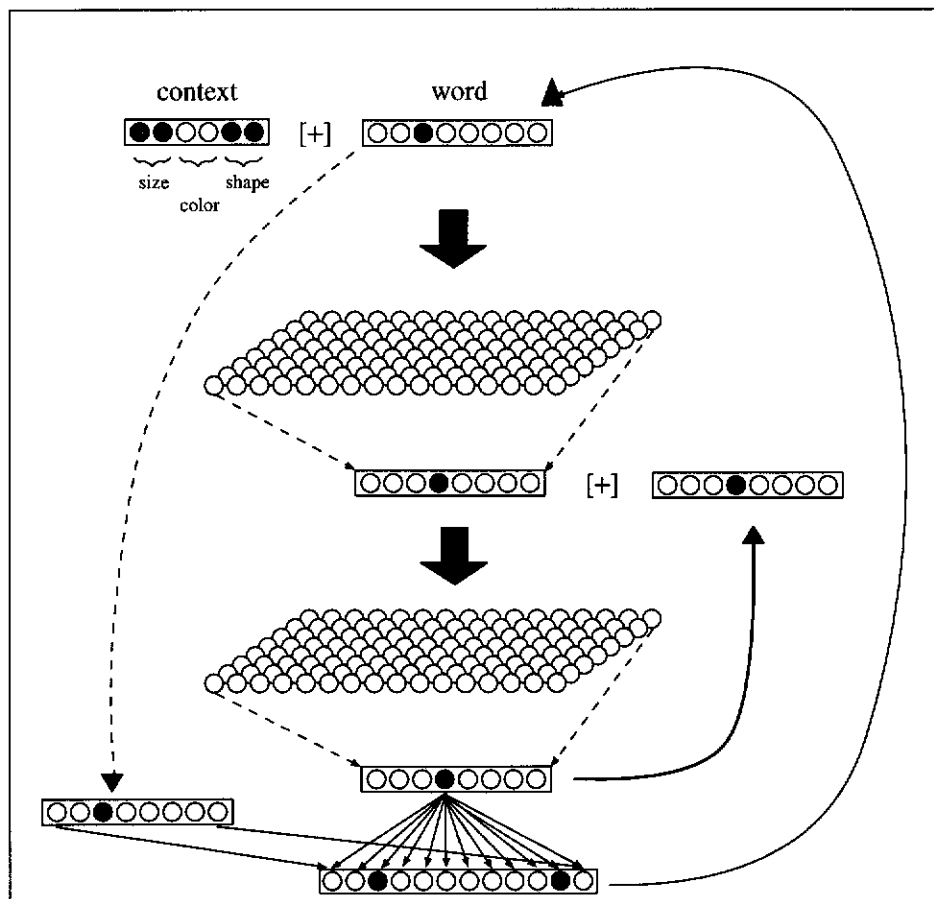


Figure 2.22: Semantic Priming.

Of the 27 possible sentence combinations, 3 were left out of the training data. Thus training consisted of 24 sentences, while the sentences left out were:

the small red triangle
the medium blue square
the large green circle

Following training, the resulting set of weights were tested by seeding with an appropriate context vector (e.g. 011110 for small blue square) and the word THE. The test program was required to generate the next three words following the seed. As all of the learned sentences began with THE, the subsequent generation of words must be determined by the semantic content.

All 27 semantic vectors were used in the testing, so the network was asked to generate the three sentences it had not seen before (when given the appropriate context vector).

The results obtained were extremely good. A typical run was: 23 of the 27 sentences were produced exactly as required to match the semantic component, with 4 incorrect. The incorrect sentences initially included the three unseen sentences, but the reasons for the sentences being marked as incorrect were interesting. Of the three words required to be generated following the seed THE, (such as SMALL BLUE CIRCLE), a typical run would produce results that were *close* to being correct—for example, (2/3, 2/3, 1/3, 1/3) where the fractions indicate the number of words correct both in content *and* position. If the required sentence was SMALL BLUE CIRCLE and the actual result was SMALL RED CIRCLE, then this would be marked as 2/3. A response CIRCLE SMALL BLUE, however, would be marked 0/3. Even though the words are correct, their positions are not.

We can refer to those sentences which have a correct positional structure as *syntactically* correct, whereas those sentences which have the correct content words as *semantically* correct.

Virtually all of the sentences generated were syntactically correct—the order of generated words was ‘size’, ‘colour’, and ‘shape’.

In ‘grammatical’ terms, the sentences all followed the structure **adjective adjective**

noun. In addition, the order of the adjectives was as required by grammatical ‘rules’.

Adjectives in English appear to show a strong ordering preference. For example, one would always use the ordering `the large red triangle` and not `the red large triangle` (and certainly not the ordering `the triangle large red`). This left-right ordering is generally true of all noun phrases. †

Consider the noun phrase example given by Winograd (1983, page 515): *a beautiful new red wooden wagon*. This ordering is considered correct—other orderings, such as **a red wooden new beautiful wagon*, are considered to be incorrect. An explanation given by linguists might be based on perceived semantic classes and an ordering between these classes—to quote Winograd:

“The exact rules for selecting and ordering these elements are very subtle; they have to do with a kind of scale of ‘inherentness’ on a semantic level, and nobody has really worked out a good formalism for them.”

But the current example suggests an alternative. This aspect of English is a natural result of the ordering of the words as perpetuated by speakers within society. If a child learning English hears the words in a certain order, the linkages are learned in that order, and they will reproduce the words in that order. No ‘semantic classes’ or ‘rules’ are required.

In only three cases out of 162 generated sentences (i.e. 6 runs of 27 cases) were the generated words syntactically incorrect, and these cases occurred in early tests in which ‘cross-talk’ between the winning nodes was present. To understand this cross-talk, we need to examine the winning nodes of the first SOM surface. Figure 2.23 illustrates

†The ‘rules’ as sought by linguists appear to be very complicated (Winograd 1983), and include *pre-determiners, determiners, ordinals, cardinals, modifiers, heads, and qualifiers*—in that order. Each of these categories are divided into further sub-divisions; for example, *modifiers* is further divided into *describers*, which can be used predicatively (such as “The baby is *asleep*” but not **“The asleep baby”*) and *classifiers*, which can be used attributively (such as “The *former* mayor” but not **“The mayor is former”*). Some adjectives can be both attributive and predicative, such as “The *red* car” and “The car is *red*”. Describers precede classifiers in the left-right ordering. For a full discussion see (Winograd 1983, page 512).

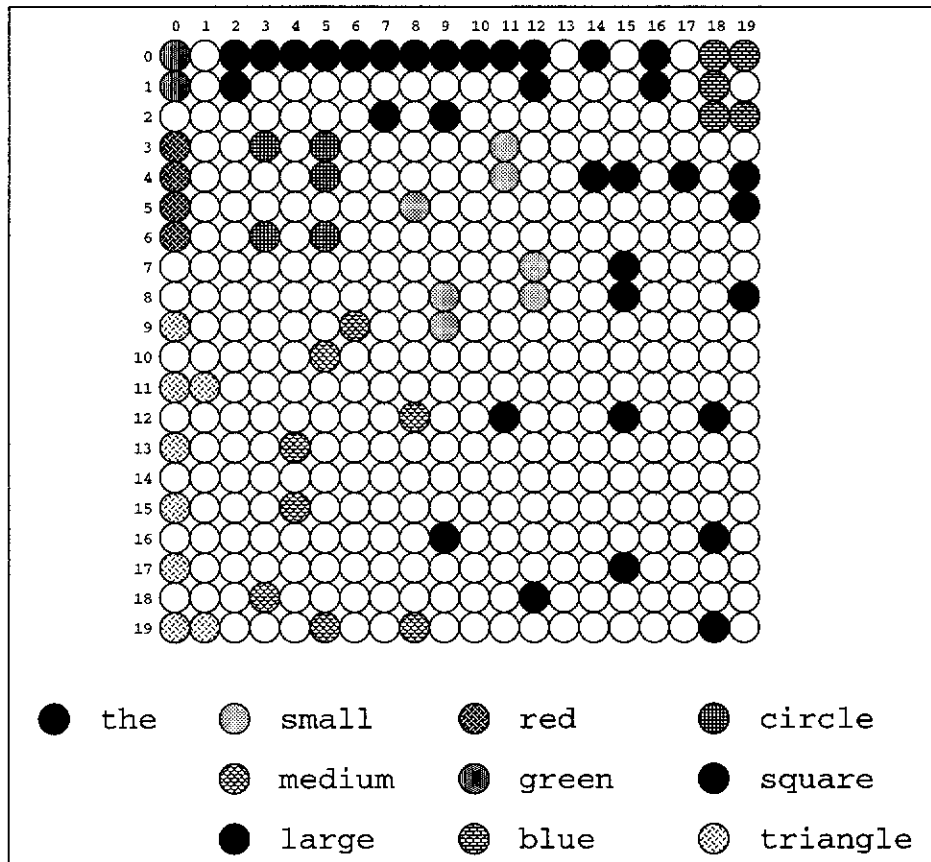


Figure 2.23: Badly Separated Semantic Map of Winning Nodes—SOM1.

the winning nodes associated with each input word. You will notice that **green** is only associated with 2 winning nodes (top left) despite the fact that it occurs in 8 sentences. The word **large** on the other hand is associated with a full compliment of 8 winning nodes (lower right).[†]

Despite the fact that these early attempts did not adequately separate the winning nodes, the results were none-the-less quite good.

To overcome this 'over-generalisation' problem, the 'exclusion' modification was used to ensure that no node on each map could be the winner for two input vectors (per epoch). This 'exclusion' method is discussed in Section 2.3.6. This procedure allowed

[†]The reason for the uneven distribution is not known at this stage, although there appears to be a slight bias in the learning algorithm to favour nodes to the top left (near 0,0) of the surface as shown. This will be investigated at a later date.

for better separation of the winning nodes, as is shown in Figure 2.24.

Once this modification was made, of the 27 possible semantic vectors, at least 24 were correctly generated in the production of the next three words following the seed THE. Further, the unseen examples, although often not completely semantically correct were very close, with in general two out of three terms being correct. All were syntactically correct.

It is also interesting to examine the second SOM surface which records the transitions between words. This is shown in Figure 2.25. The areas marked off by the solid lines indicate semantic areas where the winning nodes for a particular word are clustered. In most cases the semantic map is also ordered within each primary region; for example, for the TRIANGLE primary area, the nodes for GREEN TRIANGLE are grouped, as are those for BLUE TRIANGLE and RED TRIANGLE. Notice however that the GREEN semantic

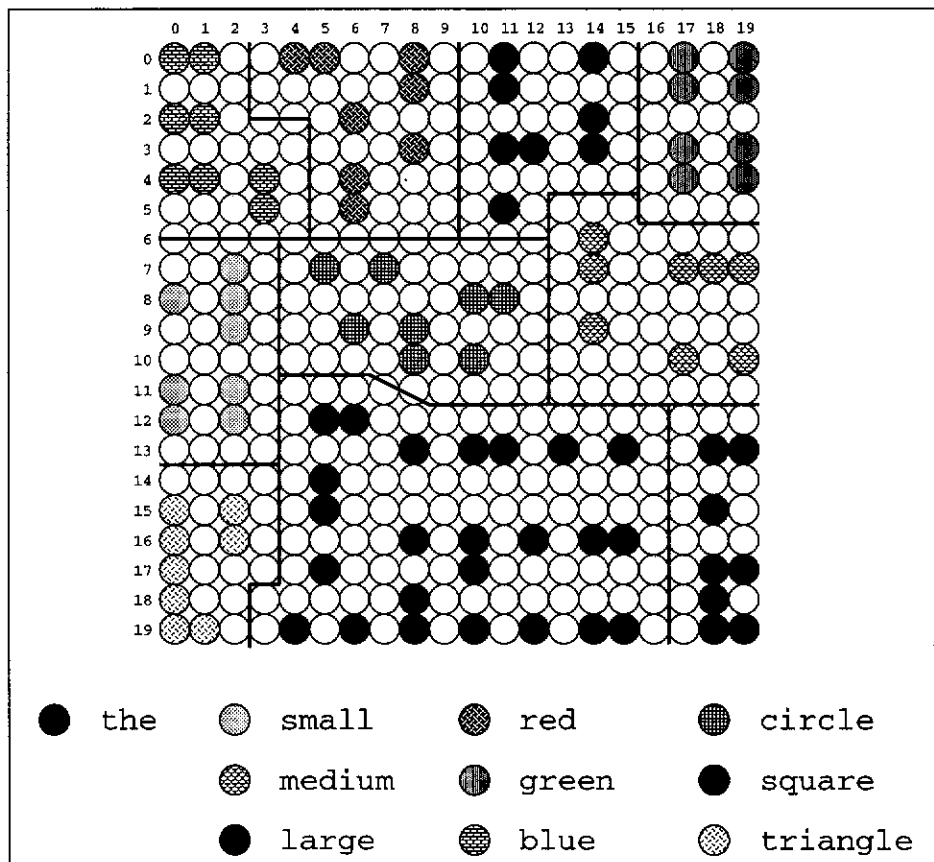


Figure 2.24: Semantic Map of Winning Nodes—SOM1.

region has not completely separated, with the nodes for LARGE GREEN straddling those of SMALL GREEN. This lack of separation may be a cause of the generation of incorrect sentences; in this particular case, the request to generate a sentence with the semantic vector indicating LARGE GREEN CIRCLE actually generated SMALL GREEN TRIANGLE.

The following table gives a sample of typical runs of the semantic priming example. Note that the asterisk (*) means that this sentence was not in the original training set.

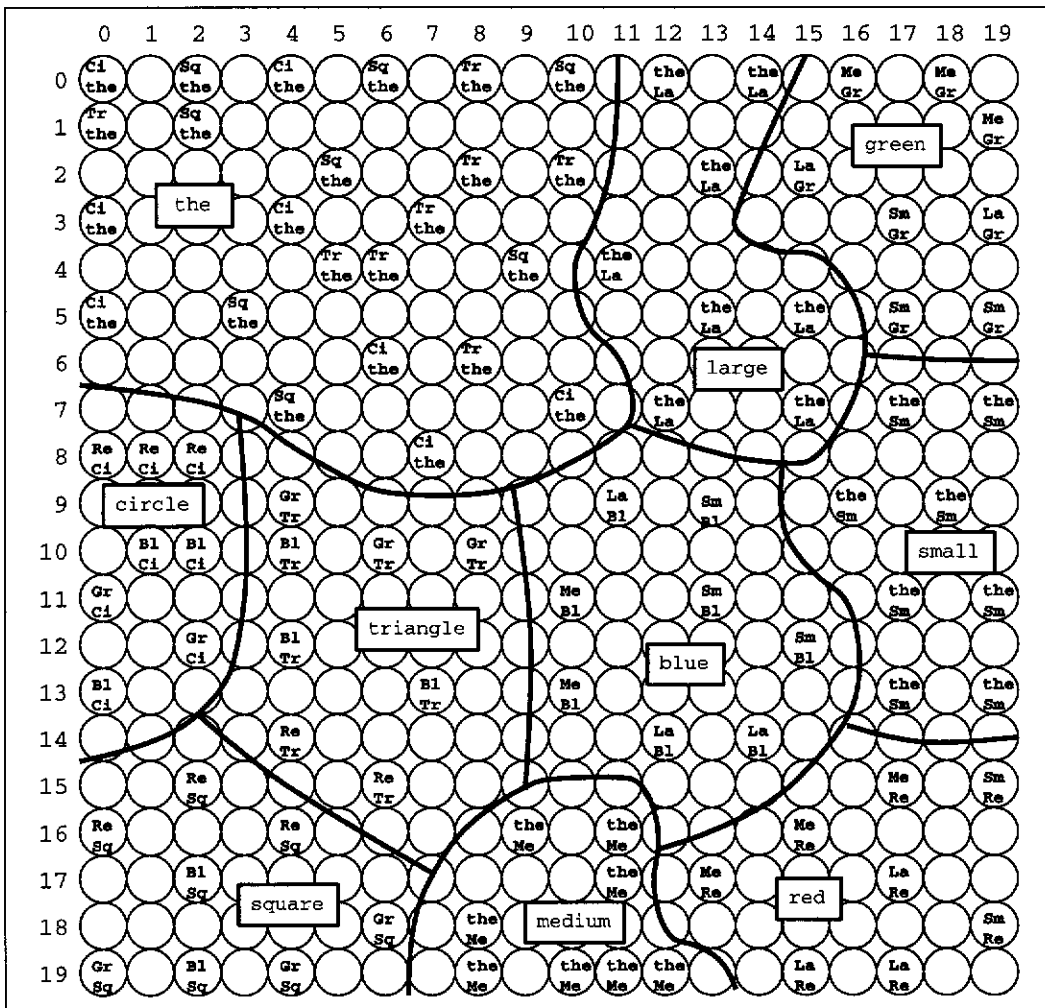


Figure 2.25: Semantic Map of Winning Nodes—SOM2.

Run No.	Number Correct	Number Incorrect	Fraction Correct	Notes
1	22	5	$\frac{1}{3}, \frac{1^*}{3}, \frac{2}{3}, \frac{1}{3}, \frac{1^*}{3}$	One sentence not in the training set was generated correctly
2	23	4	$\frac{2}{3}, \frac{2^*}{3}, \frac{1^*}{3}, \frac{1^*}{3}$	
3	24	3	$\frac{1^*}{3}, \frac{1^*}{3}, \frac{2^*}{3}$	Exclusion code used

In 2 of the 6 runs performed, the system *was* able to generate all three elements of one of the unseen examples correctly—perhaps these are the first examples of a ‘naturally’ generated semantic sentence.

Despite the simplicity of this example, it is clear that context can be an extremely important component in parsing and generating sentences. It also shows how novel sentences may be generated, provide the semantic components (the words) have already been used in other contexts.

The question as to the source of the semantic information used in the example is discussed in a later section.

Note that in this example, each ‘concept’ is represented by more than one node. This is also discussed in a later section.¹⁰

This exercise gives us some insight into a reply to Chomsky’s claim of a separation between syntax and semantics. Consider Chomsky’s (in)famous sentence

(1) Colourless green ideas sleep furiously.

The claim is that this sentence seems to be acceptable in a sense that the sentence

(2) *Furiously sleep ideas green colourless

is not.[†] The first sentence is claimed to be *syntactically* well formed (and hence syntactically correct) yet is *semantic* nonsense (ideas don’t sleep, much less sleep furiously, and they certainly aren’t green and colourless at the same time, etc, etc).

[†]The standard linguistic convention is used—any sentence that is considered “ungrammatical” is preceded with an asterisk. A sentence that is considered questionable is preceded by a question mark.

But because the ‘component parts’ of sentence (1) seem to be in the ‘correct’ order determined by the syntax of English, we can parse the sentence and understand it. We can ask questions about the content of the sentence such as “What are the ideas doing?” and the answer is clearly “sleeping”. The same does not appear to be the case for sentence (2).

Chomsky and his followers claim that this selective sensitivity to the syntactic structure of language is strong evidence for an separate, innate linguistic module (this point is discussed more fully in Section 5.1).

The examples in this section, however, indicate that syntax is no more than learned temporal sequencing. The fact that speakers of English use a certain sequence of words means that children exposed to these sequences will learn the words in the same order—the ‘grammar’ is perpetuated by the usage of words within society.

One certainly *does* need the ‘meaning’ of the words in sentence (1) to make any sense of it, and the positions of the various words which relate to objects (nouns), attributes (adjectives and adverbs), or actions (verbs) are the relational linkages which connect the words. Attribute descriptors (adjectives) usually precede the object (noun), and they tend to occur in a certain order. English is predominantly a SVO language[†] and so the action (verb) usually refers to the preceding object. The ordering is important as a cue to which words relate to which other words. The particular ordering used in English is an historical accident.

There is certainly no need to postulate a separate innate language component on the basis of sentence (1). Should the whole English speaking peoples of the world wake up tomorrow and decide to speak in the manner of sentence (2), then after a period of adjustment, their *children*, I suggest, would simply learn the new sequence and maintain that sentence (1) was somewhat strange.

[†]The major ordering is subject-verb-object.

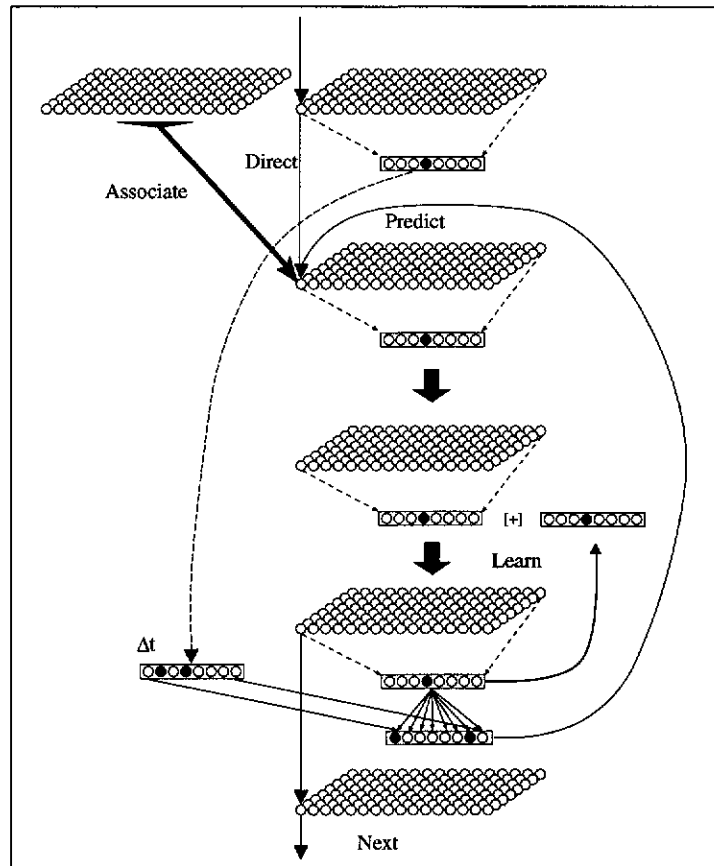


Figure 2.26: Dynamic LAPS Network.

2.5 Dynamic LAPS Structure—Predictive Knowledge

In this section, we present a brief discussion of the role of prediction (or anticipation) in cognition. We also look at the way in which the LAPS structure can provide a form of predictive knowledge of the world.

Figure 2.26 shows the full dynamic LAPS network. This shows inputs into the LAPS module from both direct and association sources.[†] Also shown is the recurrent linkage—here labelled ‘predict’. Figure 2.27 indicated the possible neural connections corresponding to this model.

Many researchers have shown that language users are *primed* by syntactic and semantic expectations. Subjects will recognise words faster if the target words are preceded

[†]The association sources are described in Chapter 3.

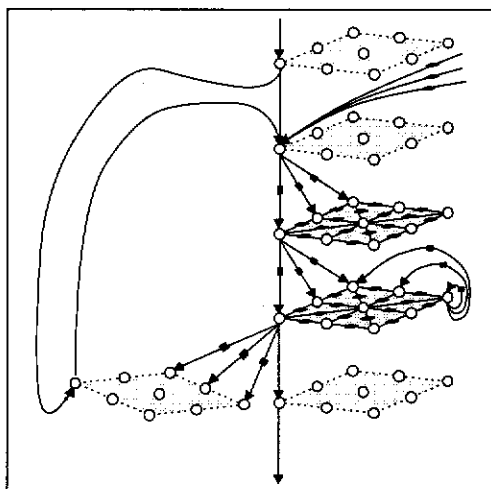


Figure 2.27: Dynamic LAPS Neuron Layout.

by consistent and associated words (Erlich & Rayner 1981, Fischler & Bloom 1979, Stanovich & West 1981, Stanovich & West 1983). See Barsalou (1992) for a review.

For example, the recognition of 'spoon' is much faster in the first of the following sentences than in the others.

Andrew ate the ice cream with a ———.

Oliver touched the slimy moss with a ———.

Anna ate the spinach salad with a ———.

The semantic association of spoon and ice cream is stronger than that between spoon and the object of the other sentences. Further, subjects are slower at recognising words that syntax and semantics predict should not occur; thus people recognise spoon slower in the third sentence than in the second.

Marslen-Wilson & Tyler (1980) conducted experiments in which subjects were given cue words, and asked to find certain target words in sentences. The sentences included normal sentences, syntactically correct but nonsense sentences, and scrambled sentences which were neither semantically or syntactically correct. They found that recognition of the target word occurred after 200 msec in normal sentences, which was even before the presentation of the 369 msec word was complete (Marslen-Wilson & Tyler 1980, page 28). In other words, the subject was able to respond before he had heard the

whole word.

Even if just a *category* is used to prime the target word (for example, if the target word is *lead*, the subject might be primed with *a type of metal*), recognition, while slower, was never-the-less achieved before the target word presentation was completed. However, when the target word was embedded in a scrambled sentence, reaction times increased by some 90-150 msec.

The first 200 msec of a word corresponds roughly to the first two phonemes (Marslen-Wilson & Tyler 1980, page 28). An analysis of the possible words in the English language which could be compatible with the initial two phonemes of the words actually used in the experiment, led Marslen-Wilson & Tyler to conclude that a selection process operating solely on the acoustic-phonetic input would have only a small chance of selecting the single correct word-candidate. Marslen-Wilson & Tyler suggested that parallel lexical, syntactic, and even interpretive ‘searches’ were being made.

The cocktail-party effect is the ability to pay attention to some incoming sound stimuli while ignoring others—selective listening (Bourne, Dominowski & Loftus 1979, page 39). In one experiment, listeners were presented with two spoken messages at once, and asked to ignore one and to repeat the other (shadowing). Subjects are able to perform shadowing very well, with some subjects being able to repeat back the input with delays of only 250 msec (Marslen-Wilson & Tyler 1980, page 25).

Shadowing subjects showed very little memory for the second (ignored) message, not even noticing that the voice changed from male to female, or from one language to another. Further, subjects will tend to anticipate and fill in more likely words if errors are deliberately inserted into the original message—for example, inserting ‘penknife’ for ‘petrol’ into some context more appropriate to a fuel (Marslen-Wilson 1973, Marslen-Wilson 1975, Marslen-Wilson & Welsh 1978, Marslen-Wilson, Tyler & Seidenberg 1978).

These results, along with other related findings, suggest not only that word interpretation is immediate, but also that word recognition is somehow *anticipated*.[†]

[†]Priming also occurs in visual processing (Barsalou 1992, page 42).

The traditional view is that as each word in a sentence is encountered, “you update syntactic hypotheses about the sentence’s constituents and semantic hypotheses about its meaning. In turn, these hypotheses make predictions about words yet to be encountered” (Barsalou 1992, page 245). The mechanism suggested in this view is that ‘top-down’ processes (so-called higher-level knowledge or procedures) in some way interact with the ‘lower-level’ perceptual processes.

The model presented in this thesis suggests a different view. The recurrent loops in the LAPS network return a vector containing the ‘most-likely’ input to follow. This is shown in Figure 2.26. This outer loop was described in Section 2.4.3, where it was shown how the outer LAPS recurrent loop could re-generate a learned sequence of alphabetic characters.

By playing back a ‘copy’ of past recordings of the world, the production loop of the LAPS architecture can support and predict (to some extent) the next input expected. This anticipatory cuing, in the case of the more regular sentences described above, should allow for the faster reaction times. However, we were not able to fully explore this aspect of the proposed architecture for this thesis, but will do so in future research.

2.6 SOMA and LAPS Networks as Modules

Figure 2.26 shows a number of sources for the input to the LAPS network. These sources are *associate*, *direct*, and *predict*.

The *direct* link is the input emanating from the appropriate direct-line source. For example, for the visual system, the direct input might be a vector resulting from a spatial-frequency filter.

The *predict* link is the recurrent link described previously in this chapter, and discussed in Section 2.5 in relation to its contribution to predicting the most likely next input.

The *associate* link is input from other associated sources. For example, in the visual system, a number of separate filters are found, providing separate vectors for spatial frequencies, colour, and motion. Each of these three sub-modalities is mapped sepa-

rately to a SOMA or LAPS network. However, in addition, the sub-modality vectors are cross-associated so that each can support the others. The association of input vectors is discussed in the next chapter.

It is appropriate to separate out the LAPS and SOMA networks so that we can describe them as a separate module. The LAPS and SOMA components are self-contained and so may be separately designated and described.

The shaded area in Figure 2.28 represents a LAPS module. In the simulation of the ABC model discussed in Appendix C we join various LAPS modules together. To simplify matters, we may need to replace the actual neural structure of a LAPS or SOMA network with an abstract diagrammatic representation.

The diagrammatic technique we will use is that of a *data-flow* diagram, as is shown in Figure 2.29. The shaded area of Figure 2.28 is replaced by the box of Figure 2.29. The clear areas represent input/output connections that enable a data-flow diagram to join up with other modules. The top connection sites are *input* sites, while those along the bottom of the data-flow diagram are *output* sites. Input sites may be connected to output sites from other modules further up in the processing path.

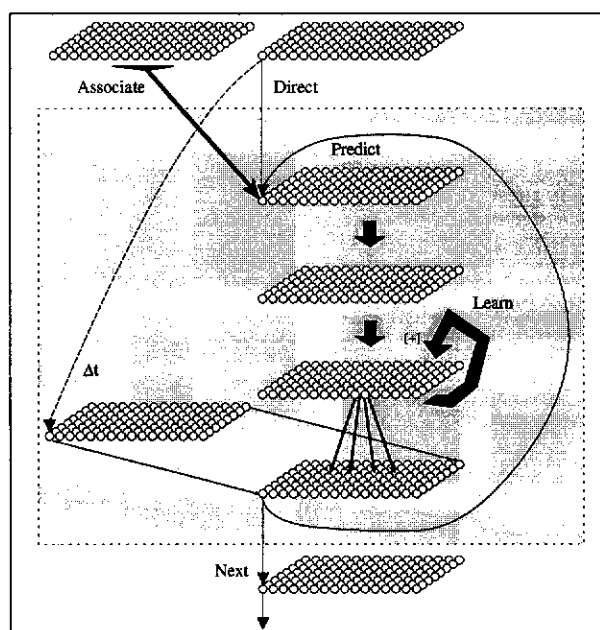


Figure 2.28: Dynamic LAPS as Module.

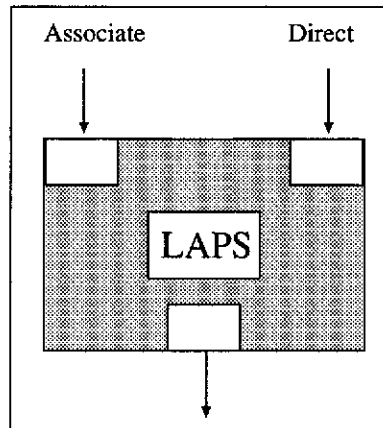


Figure 2.29: LAPS Unit Data-flow Diagram (Two Input Stream).

Note that each LAPS module may have a number of input sites, as shown in Figure 2.30.

We can represent multiple input (and multiple output) data-flow diagrams by incorporating a number of input (or output) sites. For example, Figure 2.31 indicates a LAPS module with three input sites and one output site.

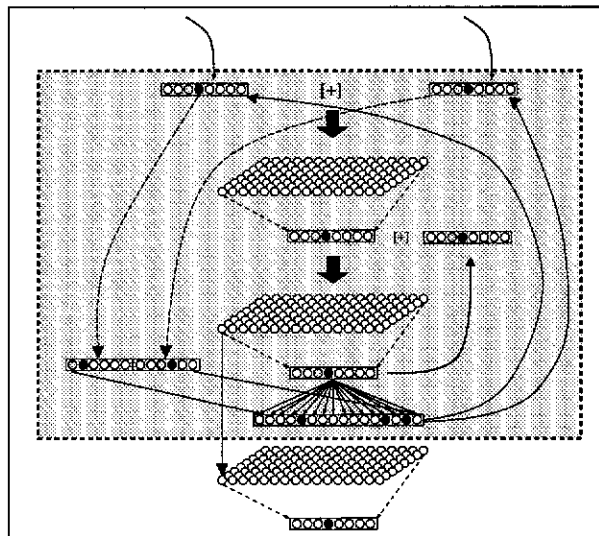


Figure 2.30: Double Direct Input LAPS Unit.

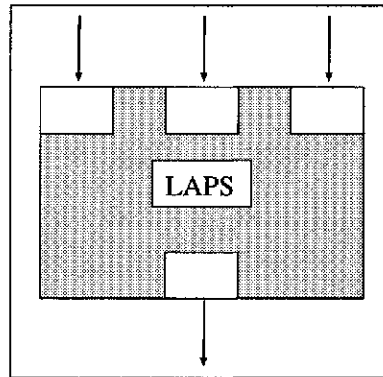


Figure 2.31: LAPS Unit Data-flow Diagram Symbol.

2.7 Discussion of Temporal Learning Results

The preceding sections (and Appendix B) show that the recurrent self-organising map architecture is an extremely powerful temporal learning mechanism. In many cases, the model was able to achieve results superior to the currently popular SRN model (Elman 1990).

Further, the model was able to obtain much more ‘symbolic’-like effects than the FFNN or recurrent SRN models. This resulted from the strong tendency to generalise on the hidden layer (second SOM surface), and resulted in the possibility of finding flow diagrams to describe the implicitly learned ‘rules’. The topological ordering and winning nodes on the SOM surfaces allows for more *categorical* results.

There are objections to the SRN approach of Elman as a cognitive model. These objections are mainly those issues also confronting FFNNs, and include the use of a supervised learning mechanism (backpropagation) which does not appear to be biologically motivated, a lack of any explanation as to the source of input vectors (which are often chosen rather arbitrarily to suit the problem at hand), and the artificial construction of the context nodes (which must equal the size of the hidden units).

The SOMA network does not succumb to these objections. The learning method is based on the biologically-feasible Hebbian learning of synaptic weights, and on the updating of weights on the SOM surface using the Kohonen algorithm. This algorithm has been given biological feasibility and a physiological interpretation in Kohonen (1993).

Further, all learning is unsupervised, or *self-supervised*.

A discussion on the justification and the source of the input vectors to the SOMA model is undertaken in the next chapter.

The method, however, is computationally expensive. As such, it is difficult to go beyond the reasonably small-scale examples described in the preceding sections. The problem is the use of software *simulations* serial computers.

Many researchers have not appreciated the fact that biological neural networks are *hardware* devices, although current *artificial* neural networks are software solutions. While software simulations of neural networks are useful, they are problematic and slow, and only allow small cognitive problems to be examined. By concentrating on software, we neglect the much more powerful applications that will only become available with hardware solutions.

What is required are massively-parallel hardware devices, (perhaps based on the LAPS model), which will enable a scaling to a very powerful methodology for temporal learning. At present, we are really just scratching the surface, and much work remains to be done before temporal learning phenomena, such as language and motor control, can be said to be well understood.

Chapter 3

Adaptive Behavioral Cognition (ABC)—The Model

3.1 Introduction

Before describing the full ABC model, it is appropriate to make a disclaimer—it is not claimed that the model presented in this thesis completely describes the structure and working of the brain. The brain of humans and the higher vertebrates has evolved over many millions of years. Various components have been laid down at different periods of evolution, and it may even be valid to view the brain as several interacting “sub-brains” rather than as a single unit.

The brain of vertebrates consists of three basic divisions; the cerebral hemispheres, the brain stem and the cerebellum. This thesis considers only the most recent addition phylogenetically—the cerebral cortex. The cerebral cortex is most pronounced in higher vertebrates, and from the study of aphasia and other deficits has been shown to be the major area for cognitive function (Kandel & Schwartz 1983).[†]

The sheer number of neurons and connections in the brain is bewildering, and the significance of the projections between the various cortical and extracortical areas is

[†]Even the cortex of (at least) higher primates is believed to have a dual evolutionary developmental path (Sanides 1969, Sanides 1972).

not well understood. In this thesis, an attempt is made to examine general principles that may be used in forming an overall reverse-engineering “design” for the brain, and by so doing, to try to form an understanding of how humans and animals are able to conceptualise, to learn, and in the case of humans in particular, to use language—given the neurobiological, behavioural and other data that we currently have available.

A number of important features of the brain of higher animals need to be incorporated into any model:

- the ubiquity of topographical maps in the cortex and other brain areas,
- the prevalence and systematic character of feedback loops signifying a dynamic, time-dependent system,
- the numerous projections to lower brain centres, including motor-relevant areas such as the striatum (basal ganglia), superior colliculus and cerebellum,
- the cross-linking of the various modality sensory projections in the association areas.

These, and other known properties of the brain, are incorporated into the ABC model in the following sections.

3.2 Building the ABC Model

In this section we tie together the results and the network structures from the temporal learning experiments discussed in the previous chapter, in order to build an extended cognitive model.

The ABC system is based on ‘matrix’ or ‘array’ transformations—essentially transforming one SOM surface into the next. However, it is sometimes easier to conceptualise the model if, at times, we think of these arrays as vectors. This enables us to better fit the ideas expressed about the ABC model into the current literature on artificial neural networks (ANN).

3.2.1 Sense Data and World Filters

The external world provides various forms of energy that may be sampled by an organism. Specialised neural structures (sensory receptors) transform these natural stimuli that impinge upon our bodies into neural signals. In addition, the brain receives inputs from other internal sources such as from the movement and tensions of skeletal muscles, and from the viscera. †

These different sensory systems may be subdivided into three categories: the *exteroceptive* systems which are sensitive to external stimuli, the *proprioceptive* systems which provide information about the relative positions of body segments and the position of the body in space, and the *interoceptive* systems which provide internal bodily measures such as blood pressure and blood glucose levels. We are usually not conscious of the interoceptive signals, whereas we appear to be generally aware of the exteroceptive and proprioceptive stimuli.

In the following discussion we concern ourselves mainly with the exteroceptive systems, although there is no reason to doubt that the same principles could apply to the other systems as well. The exteroceptive systems of concern are the visual system, the auditory system and the somatosensory system. ‡ Each of these modalities has several sub-modalities as outlined in Table 3.1.

Each sub-modality ‘filter’, via its sensory receptors, provides a number of neural responses *at each instant of time*. It is not appropriate to examine these sensory systems in detail at this stage. We examine this point briefly in the discussion of temporal learning in the visual system (Section 4.6), and more detail is provided in the discussion of the simulation of the full system in Appendix C. All we need to consider at this stage is that the sensory systems, both external and internal, initially provide a ‘vector’—that is, a reading $v_s^m(t)$, of the sensory receptors at time t , where m is some modality, and s one of its sub-modalities.

The ABC model used in simulations assumes that an appropriate filter vector has been

†Internal organs such as the heart, stomach, lungs, tear ducts, glands, blood vessels and so on.

‡We do not concern ourselves with the olfactory or gustatory systems as these do not belong to the neocortex but are older systems.

Modality	Sub-modalities
Vision	Spatial Frequencies & Angles Colour Motion
Auditory	Frequency (Tone) Bandwidth (Spectral profile) Symmetry and Rate of Change (Prosody) Intensity
Somatosensory	Touch-pressure Position Sense—static limb position and limb movement (kinesthesia) Thermal Sensations—separate for both hot and cold Pain Sensation

Table 3.1: Exteroceptive Modalities and Sub-modalities.

obtained. In many of the temporal simulations, for example, the vector value assigned to the inputs was assigned at random. In the full simulation of the model, each fixation provides appropriate vectors for the visual modality component from:

- a spatial-frequency and angle filter (based on the Wilson Modified Line-element theory (Wilson & Gelb 1984)),
- a motion filter (based on a Reichardt motion detector),
- a simple RGB colour filter.

Each other modality provides appropriate vectors based on the recording of the external world made by the particular filters. At each time instant, every vector is mapped to its own particular SOM. That is, the spatial-frequency and angles vector is mapped to a SOM, the motion vector mapped to another SOM, and the colour vector mapped to a third SOM.

Each of these maps will self-organise to the actual distribution of mapped vectors, thus providing a representation of the external world, and with an appropriate allocation of the surface area being given to the frequent incoming vectors within the creature's experience.

3.2.2 Self-Organising Maps

The sub-modality vectors are mapped to self-organising layers of the primary and secondary sensory areas. The primary and secondary sensory areas in primates are well known to contain topologically organised mappings of the incoming sense data. In the case of the visual system, these are areas V1 and V2, with corresponding areas A1 and A2 for primary and secondary auditory areas, and S1 and S2 the somatosensory primary and secondary areas. These areas are shown in Figure A.5 (a) for the rhesus monkey. The process of topological mapping to various sensory areas is consistent with the mapping to the initial SOM surface as described in the previous chapter.

The maps thus form a statistical record of the incoming data. If a change is experienced in the statistics of a particular region of the incoming data, then the self-organising map will accommodate by adapting and allocating a larger (or smaller) portion of the map neurons to this region (for example, see Florence & Kaas 1995).

A more detailed discussion of the sensory systems of the brain may be found in Kandel & Schwartz (1983) or Sekuler & Blake (1990). Other collected articles are found in Arbib (1995b) or Gazzaniga (1995b).

3.2.3 Recurrent Loops

As discussed previously, the cortex of primates is richly connected by back-projections to previous layers, consistent with the recurrent loop structure of the ABC model. Recurrent projections for the monkey visual system are shown in Figure 4.17. Appendix A also cites strong neuroanatomical support for the model. Figures A.5 (d), and A.6 (a) and (b) show projections from each of the association areas to the prefrontal and premotor areas, while Figures A.6(c) and (d) show the back-projections from the prefrontal and premotor cortical regions back to the perceptual association areas. Neuroanatomical and neuropsychological findings for the monkey and humans are discussed more fully in Appendix A.

The next stage of the model is to combine the multiple sources of information that are available within each modality.

3.2.4 Intra-Modal and Inter-Modal Associative Layers

Association layers surround the primary sensory regions (Pandya & Yeterian 1985, Fuster 1985). These are of two types; *intra-modal* association areas which link various regions within each modality, and *inter-modal* association areas which link downstream areas of separate modalities (for example, visual with auditory, visual with somatosensory, and so on). The intra-modal association areas for the rhesus monkey are shown in Figure A.5 (b), while the inter-modal association areas are shown in Figure A.5 (c). A full discussion of the association areas of the monkey is given in Appendix Section A.5.

Associative linkage between all of the maps within one modality is achieved in the ABC model in a manner shown in Figure 3.1. A second layer of neurons, termed the *association layer*, accepts direct inputs from the corresponding SOM layer for each sub-modality, as well as inputs from other SOM surfaces within the modality. The link between each SOM layer and its corresponding association layer is direct (i.e., unweighted), and each neuron in the SOM layer projects only to the corresponding neuron in the association layer. The linkage is 1 : 1, and we call this weight a *direct weight*.

In contrast, the association link from other SOM surfaces to a particular association layer may be reduced to some fraction of the original. Each neuron in a SOM layer has a weighted link to *every* other neuron in other association layers—the linkage is thus $n : m$. We call these weights *association weights*.

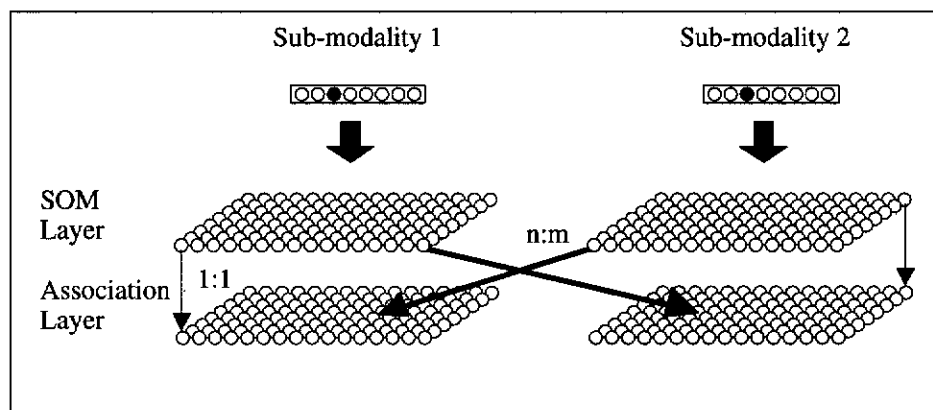


Figure 3.1: ABC Associative Layer.

The association weights are learned by Hebbian reinforcement (Hebb 1949). If both the pre-synaptic neuron (on a SOM layer) and the post-synaptic neuron (on an association layer, excited by the direct weight and possibly other association weights) are firing, then the weight is increased. The weight is decreased if there is no correlation between the pre- and post-synaptic neurons.

The linked connection between sub-modalities allows for associations to be formed between vector inputs—they can mutually support each other or inhibit each other. An input that is perhaps too small to be detected and discriminated on a single map may, because of the mutual support of the multiple maps, obtain sufficient excitation to enable further processing.

An obvious advantage of this linkage scheme is that it is non-metric. The cross-modal association of topographic maps, as well as the use of vector concatenation, eliminates the problem of metrics between multiple modalities (for example, the problem of how to join visual and auditory input features). Instead of absolute metrics, the maps allow for relative (within map) metrics or measurements. Thus the model does not suffer the problem of feature-space representations which attempt to directly link metric features in a single space.

The input vector to the next stage is the output of the association layer, rather than the vectors from the external world filters, as is shown in Figure 3.2. As previously the [+] symbol denotes vector concatenation

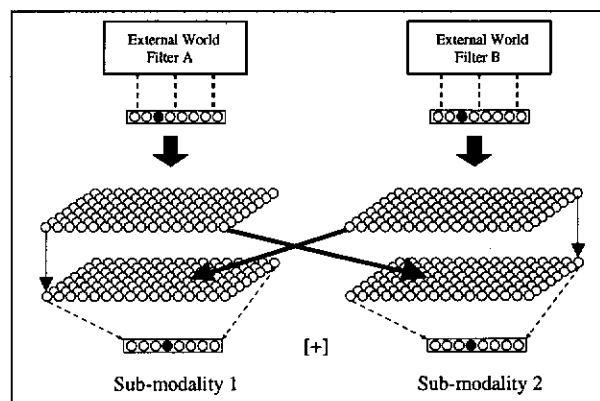


Figure 3.2: Associative Layer and External World Filters.

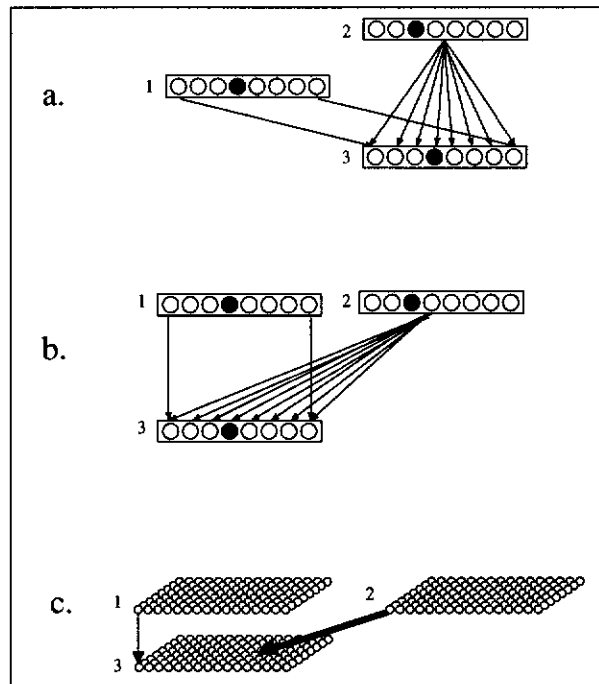


Figure 3.3: Equivalence of Diagrammatic Forms.

3.2.5 Equivalence of Representations

We should also note at this point the equivalence of the diagrammatic representations shown in Figure 3.3. In our discussion of temporal learning we used the diagrammatic form shown in Figure 3.3 (a). This form was used to emphasise that the motor weights between vectors 2 and 3 were to be updated via Hebbian learning in order to learn to reproduce vector 1 . The connections between vectors 2 and 3 are weighted, whereas the connections between vectors 1 and 3 are direct—a weight of unity. Each node in vector 2 is connected to *all* nodes of vector 3 , whereas the connection between nodes of vectors 1 and 3 is only to the single positionally-matched node—that is, only one connection per node.

Rearranging the vectors as in Figure 3.3 (b), and noting that the distinction between vectors and arrays is both artificial and was only introduced for convenience of understanding, we can see that the diagrammatic form of Figure 3.3 (a) is indeed equivalent to that of Figure 3.3 (c).

3.2.6 Model of Visual Cognition

Figure 3.4 shows one possible model for the visual system. While initially appearing to be complicated, the figure can be readily explained. Figure 3.5 dissects the figure into a number of labelled components which are described in the following:

- external world filters e.g., spatial frequency and orientation, colour, motion—only two filters are shown here but there may be more,
- vector produced by world filter,
- SOM mapping of sub-modality vector,
- association layers between sub-modalities,
- vectorisation of association layer output,

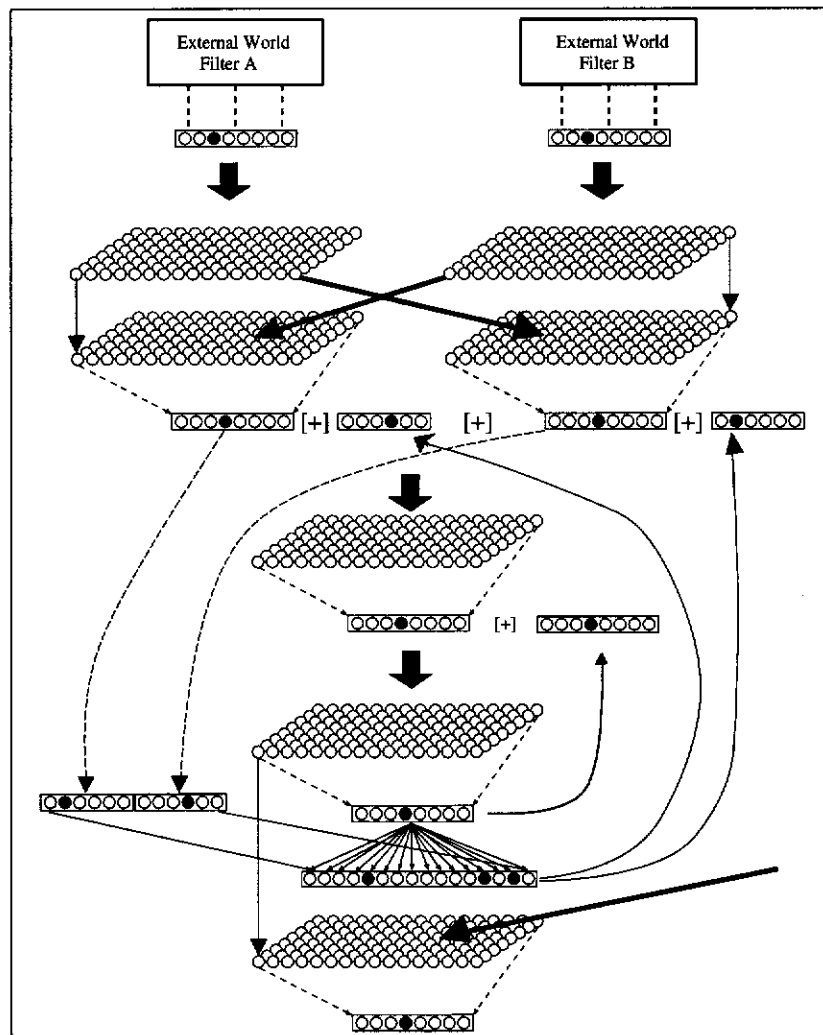


Figure 3.4: ABC Vision Model.

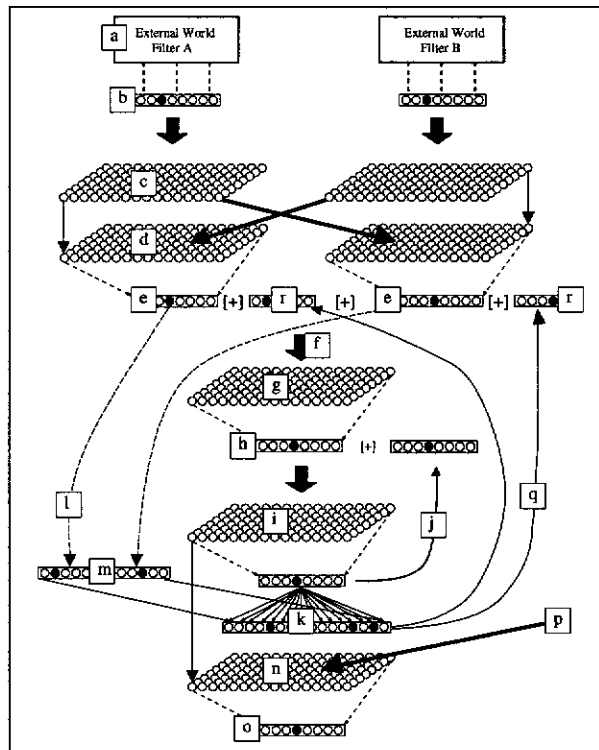


Figure 3.5: Vision System Dissected.

- f. concatenation of sub-modality output vectors for mapping to vision modality SOM,
- g. vision modality SOM,
- h. vectorisation of vision modality SOM,
- i. vision transitions recurrent loop SOM,
- j. recurrent temporal loop,
- k. vision ‘motor’ vector,
- l. ‘copy’ of input vectors to record temporal differences,
- m. time displaced ‘copy’ of input vectors,
- n. association layer to join other modalities,
- o. association vector to join other modalities,
- p. weighted associations from other modalities—auditory, somatosensory, proprioceptive,
- q. sequence generation loop—“imaging”,
- r. recurrent vector.

Incoming sensory inputs are sampled by each filter type (a) to produce a vector (b) which is topographically mapped to the appropriate SOM surface (c). The sensory SOM surfaces are cross-associated on to association layers (d), the output of which provides a vector (e) to be temporally learned by later processes.

The various vectors are concatenated (f), and topologically mapped to another SOM surface (g) which records the overall inputs from all of the sub-modalities of this modality, as well as the temporal sequence information (q).

The neural structures then follow that of a LAPS network as described in Section 2.4.4 (e-m, q, and r). The output from the modality SOM surface (h) is concatenated with the output from the subsequent temporal-transition SOM from the previous cycle (j), for mapping to the temporal-transition SOM (i). Delayed time-displaced copies of the incoming vectors (m) are ‘copied’ (l) and used to train (k) the next-cycle of recurrent ‘expectation’ vectors (r) which are ‘copied’ (q) to be concatenated with subsequent inputs.

The temporal-transition SOM output is copied to another association layer (n) for association with other modalities (p). The output from this association layer (o) is then available for subsequent processing.

The mechanism allows for the recording of temporal sequences within the visual modality, as well as having these filter inputs associated temporally via the use of cross linkages and concatenated vectors.

Other similar structures will record temporal sequences for the auditory and somatosensory modalities, each based upon appropriate filters. For example, the auditory modality could utilise filters such as tonal frequency, tonal intensity, and prosody.

3.2.7 Multi-modal Associations and Motor Control

As well as joining together the components of each sub-modality, the various modalities (somatosensory, auditory, and visual) must be combined and associated with each other—the inter-modality associations. This is achieved in the same manner as the intra-modality associations, and is shown in Figure 3.6. At this level, information from *all* modalities is mapped and learned as sequences.

This component of the model is intended to correspond to the frontal lobe in human and animal brains, and as we report in our discussion on the neuroanatomy of the brain in Appendix A, there may in fact be several levels of recurrence and attachment

to muscles.

Having multiple levels of connected and recurrent temporal learning components means that complex and intricate behavioural sequences are able to be learned. We touch upon this in Appendix Section B.3.1.2 when we see how multiple levels of recurrent SOMs are able to be trained to learn a ‘c-like’ language, and we provide further discussion in Chapter 5 when we examine the language implications of the model.

The motor action component of the ABC model—the back-end—is also shown in Figure 3.6. The earlier processing described so far allows for the storage of temporal sequences and the formation of ‘concepts’. The motor action component allows us to perform some sort of behaviour based on these learned concepts—the activation of muscles or muscle-groups. Of course, this structure will be duplicated many times for different muscle groups (e.g., speech muscles, arm muscles, leg muscles—in fact, all of the voluntary muscles).

As well, there are at least two additional components that we have not yet shown—a proprioceptive recurrent loop that will link the activation and position of muscles with the sensory inputs, and an interoceptive loop to incorporate the internal state of the creature. We discuss the possible role of the interoceptive system briefly in Section 3.2.7 where we discuss the limbic system.

These additional components could be added in a number of ways. For the proprioceptive system, perhaps the simplest is to regard it as the equivalent of an external sense modality, and thus incorporate it in the same manner as previously described.

The interoceptive system could also be added in this manner as an additional vector component to be mapped and vector concatenated. This would give a additional component to the system that would provide a *value-based* context. For example, if the creature is hungry, then the corresponding vector elements from the limbic system will be enhanced, and this will tend to increase the tendency for the subsequent behaviour to proceed toward hunger-relief actions which were previously learned in that context.

One of the striking and important features of the neocortex is that, (with the exception of some olfactory information), almost all of the information it receives, either from

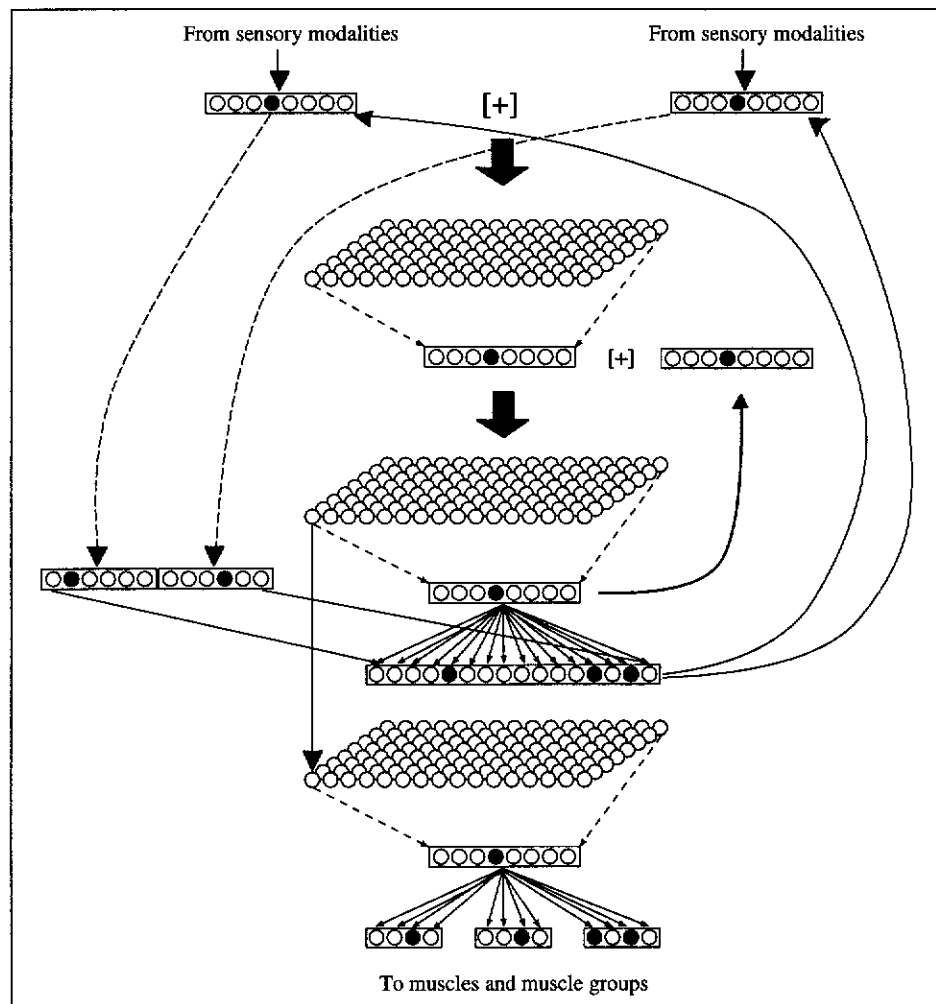


Figure 3.6: Multi-Modal Associations and Motor Control.

external senses or from other subcortical areas, passes through the thalamus (Crick & Asanuma 1986, page 346). The lateral geniculate nucleus, for example, is part of the thalamus and obtains visual-related impulses from the retina. The thalamus may thus be seen as a form of 'gateway' to the neocortex—a place where the incoming 'vectors' are collected together.

Other components of the thalamus provide projections to the primary auditory area, the primary somatosensory area, the association areas, and the various areas of the frontal lobe—the prefrontal, premotor, and motor areas. For example, the ventral anterior nucleus projects to the premotor area, while the ventral lateral nucleus sends many of its fibres to the motor area. These nuclei ensure that the cortical motor mechanisms

are connected to inputs from the cerebellum and basal ganglia. The cerebellum is a brain region thought to be responsible for sensory-motor coordination, while the basal ganglia also has some role in motor behaviour. Both are phylogenetically older than the neocortex (Kolb & Whishaw 1990).

The interoceptive system could also then be included in a very different manner—by providing a means of varying the learning rate of the network based on the emotional content at the time. The limbic system receives projections from all incoming sensory areas, while its output, either directly or indirectly, affects all endocrine, visceral motor, and somatic motor effectors. †

Hebbian learning allows for a learning rate. One of the simplest mathematical formulations of the Hebbian rule is:

$$\Delta w_i = -\mu(x_i - x_0)(y - y_0)$$

where the x_i are the activities of the presynaptic neurons, y the activity of the postsynaptic neuron, Δw_i the increment in the weight between the two nodes, x_0 and y_0 constant base firing rates, and μ a learning rate which is also generally assumed to be constant (Shouval & Perrone 1995).

However, a variable learning rate based on ‘emotional’ context would allow for a differential learning of associations. Inputs from the affective components of the interoceptive system (pain, hunger, and so on) could be used to vary the learning rate.

Variable learning rates would allow for an explanation of ‘single-episode’ learning. For example, if a child puts their hand into a flame, the subsequent pain will produce a high learning rate (gain), and so will increase the resulting weight change to such an extent that they may learn, in a single episode, to avoid such actions in the future. Other repetitive connections with a lower ‘emotional gain’ will be learned at a slower rate based on accumulated statistics.

†Hormone producing glands such as the pituitary, thyroid and adrenal systems, the involuntary muscles, and the voluntary muscles respectively.

3.2.8 Overall Model

We are now in a position to combine these components to get an overall structure for the cortex. The first pass at a full system is shown in Figure 3.7.

This shows the model as being composed of multiple layers of a self-similar component—an extension of the LAPS module introduced in Chapter 2. Figure 3.8 shows this more

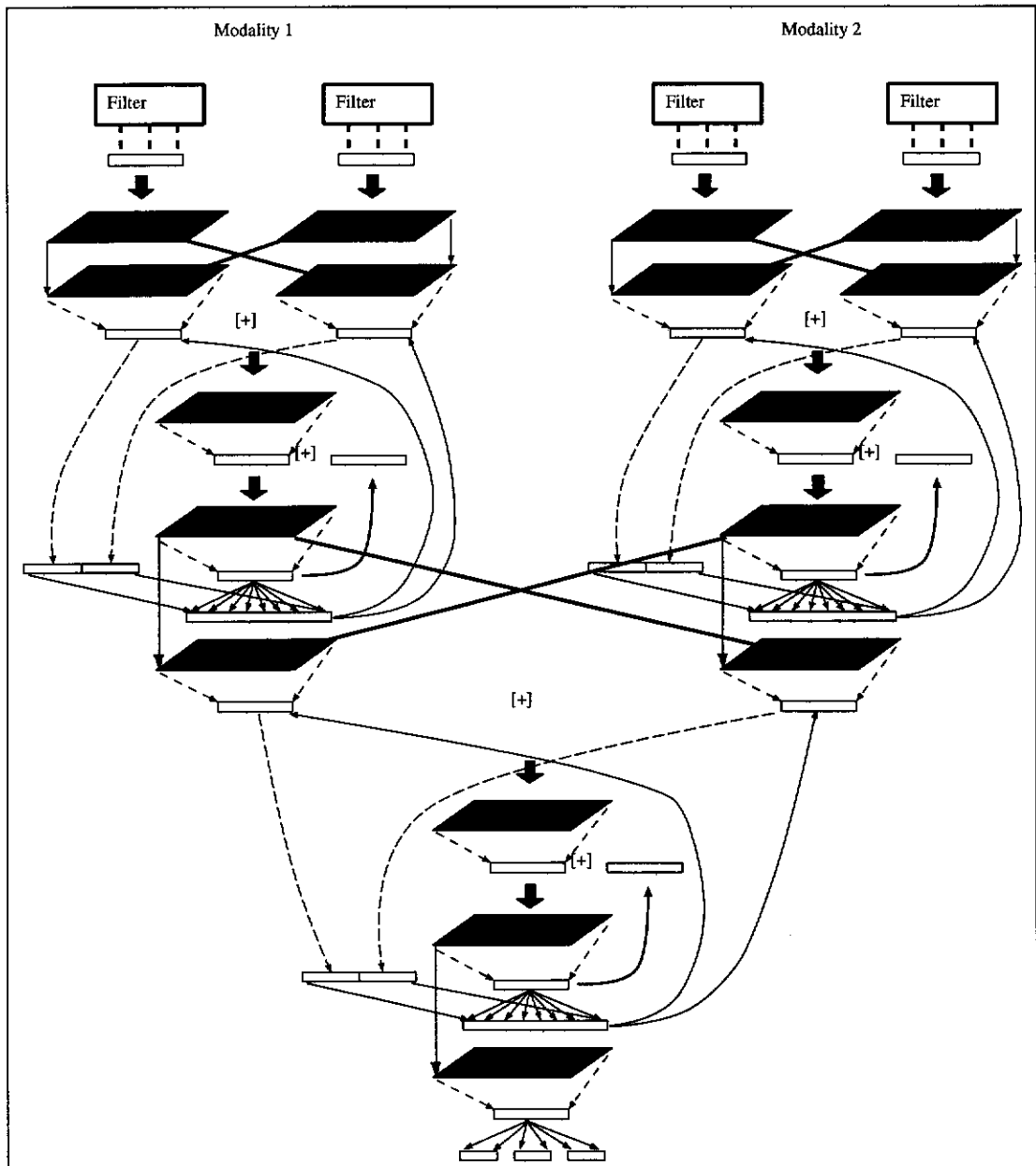


Figure 3.7: Overall Cognition System.

clearly—three LAPS modules are connected between the input sensory filters and the output motor actions.

Each modality may have a number of separate filters which extract information from the surrounding world. These are combined and associated into temporal sequences within each sensory LAPS structure. In Figure 3.8 we show only the visual and auditory modalities—the somatosensory modality is left out for clarity, but the extension is obvious.

Each of the sensory modalities is then combined into a global LAPS module. This is postulated to be the component which is equivalent to the frontal cortex in human and animal brains. The module associates and combines the outputs from the sensory modalities and forms further temporal sequences based on the overall temporal connections between the modalities.

As we saw in Chapter 2, multiple interconnected levels of recurrent loops enables the system to learn arbitrary and complex sequences. Further, there is evidence, as we discuss in Appendix A, that the frontal lobe is itself composed of a number of levels which allow for both more direct and indirect connections to the output motor control

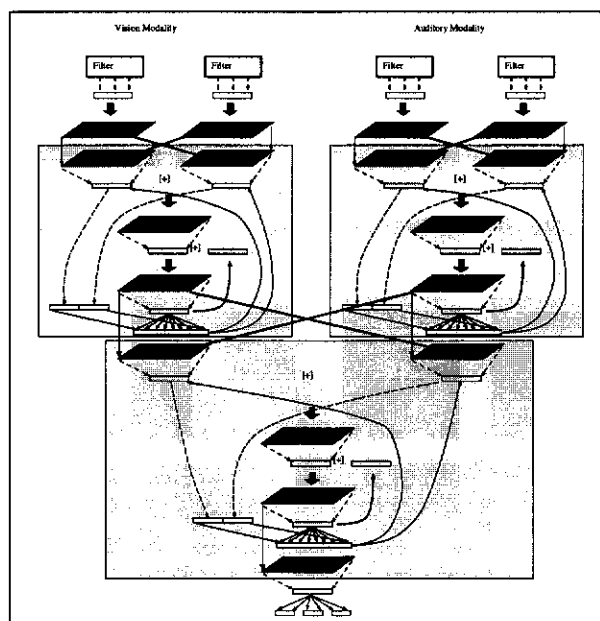


Figure 3.8: Overall Cognition System as Extended LAPS Modules.

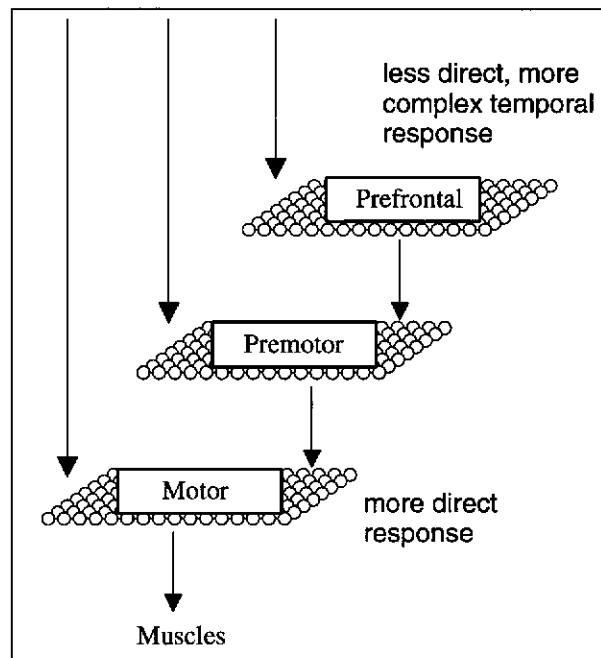


Figure 3.9: Connectivity of Frontal Cortex.

mechanisms. This may be implemented with multiple hierarchically-connected LAPS modules. The evidence suggests that the frontal cortex may be connected as shown in Figure 3.9.

This structure would appear to allow for a variety of behavioural responses to inputs from the sensory and association levels in the brain. The connectivity to the motor cortex would allow for more direct motor responses, while the connectivity to the prefrontal cortex would allow for a less direct, more complex temporal response.

The output of the overall model is a SOM (or a number of SOMs) which is then connected to the various sets of muscles.[†] Winning nodes on this surface (in an extension to the model we allow for multiple winning nodes) activate the muscles, and the temporal sequences provide for coordinated muscle sequences to be generated. The elements of each output vector are connected to separate muscles, allowing for coordinated muscle behaviour.¹¹

[†]We restrict ourselves to the somatic muscles.

3.3 Discussion of the ABC Model

We do not claim that the model as presented is necessarily a complete or precise structure of the cortex, but we do suggest that structures of this kind are what is needed to understand many facets of cognition. The requirement at this stage is to find some initial principles and overall structures for the working of the cortex. These models may then be used for ‘cognitive simulations’ in order to verify that the behavioural outcomes and other properties of the model are consistent with human and animal cognition.

We claim that the model:

- is biologically and neuroanatomically consistent and reasonable,
- has properties more closely associated with the actual brain than either the computational cognitivist approach, or the simplistic FFNN,
- is internally consistent and self-similar,
- is consistent across the whole structure of the brain i.e., the model is an overall general structure of the cortex which incorporates perception, learning, conceptualisation, motor action, language, and even consciousness,
- allows for the many attractive properties associated with the connectionist approach, such as generalisation, soft-degradation, learning and association, and so on,
- provides for massive parallelism in the processing between layers, yet is serial in its use of temporal sequences,
- is a single, integrated architecture for both lower-level perceptual processes and higher-level cognitive processes, and provides a bridge from neural processes to behaviour,
- is consistent with the view taken in this thesis that the processes of the brain are to learn associated and temporally connected sequences, rather than ‘facts’

or ‘representations’, and that the learned behaviours resulting from the associated temporal sequences are the means of cognition, rather than computational processes on representations.

What does the model give us? In this section we examine some of the support for the model—that it provides a better explanation for many aspects of cognition, and how it overcomes many of the problem issues of previous models. In Appendix A we also look at the neuroanatomical, neurophysiological, and neuropsychological evidence in support of the model.

The first thing to note is that it is a unifying theory—it incorporates and brings together features of several other views, including Action Theory, Behaviourism, Situated Cognition, Neural Darwinism, and Connectionism.

3.3.1 Cross-Modal Linkages and Associationism

Associationism has a long history within philosophy and the psychology of the brain. Bishop Berkeley (1709) was one of the first to claim that elementary sense impressions were welded together by association with images of past impressions to form meaningful perceptions. His view of visual perception held that vision gained its meaning from touch. He also concluded that any apparent equivalence of sensory information is the result of sensory integration. Other theorists held that sensations gained meaning through association with actions, which in turn provided feedback stimulation.[†]

Associationism flourished as a philosophy of mind during the eighteenth and nineteenth centuries, and included such figures as John Locke, David Hume, Alexander Bain, and John Stuart Mill. Their contention was that differences in ideas seem tied to differences in sense-experience, thus rendering the theory of innate ideas implausible. They held that ideas are related and that cognition could best be explained “by principles relating to how sensations, ideas of sensations, and ideas themselves are associated one with another” (Honderich 1995, page 62). Bain, for example, insisted that “. . . the primary form of association is the mere contiguity of ideas of sensation in experience.”

[†]The ABC model allows for both of these types of association.

We are of the view that cross-modal linkage of perceptual inputs, based on value/emotion feedback from actions in the environment, is one of the keys to perceptual categorisation and cognitive function. It allows for the formation of temporally correlated attractors and sequences, which in turn permits the recording and actioning of appropriate behaviours given particular happenings in both the external environment and internal makeup of the creature.

3.3.2 Neural-Based Not Symbol-Based

The point to be made here is that the ABC model is based solely upon assemblies of neurons, and is a physical ‘hardware-only’ system. We simulate the model on a computer as discussed in Appendix C, and this simulation model is certainly symbol-based, but the object that the model is simulating (the ABC model of the cortex) is in no way based upon symbols. Linguistic representations and symbols take no part in the underlying processes of the model.

This is an important point, and so we discuss the issue of language, as well as that of representations and computationalism, in much more detail in Chapter 5.

3.3.3 Recurrent Feedback

The ABC model allows for recurrent feedback at several levels. As we show in Appendix Section B.3, multiple hierarchical recurrent networks may be trained to learn so-called *context-sensitive* sequences such as $a^n b^n c^n$ and the c-like language. As we discuss further in Chapter 5 while discussing language, we contend that recurrent feedback at several levels gives the overall system the ability to learn arbitrary sequences.

This is an extension of the work of Doya (1995), who has shown that recurrent neural networks can model arbitrary dynamical systems, which in turn follows the results of Hornik (1991) who proved that multi-layer neural networks can approximate arbitrary mappings. Further mathematical analysis needs to be performed on the ABC structures to fully appreciate the implications of the temporal structures that are able to be learned.

3.3.4 Sense Data and Filters—Sampling of the World

The first thing to note is that one of the major functions of the brain is to interact with the outside world in some way so as to give some survival advantage to the creature. This requirement is obvious, but is often overlooked.

Perception of the external world (as performed by humans and other living beings) is really a sampling of the world through various modalities (sight, sound, taste, somatosensory, echo-location, magnetism, etc.). Even within a single modality, there may be several ways of extracting information. For example, with electro-magnetic radiation, information may be obtained as to the colour of the light (Caelli & Reye 1993), the wavelength of extracted lines (spatial frequency) or the angle of extracted lines (spatial angle) (MacLennan 1991, Zetzsche & Caelli 1989), or motion (Marshall 1990). Other sub-modalities include prosody, intensity, and frequency for the auditory system.

There are many possible modes of information available from the world—electro-magnetic radiation in the form of light and heat, vibrations in the air and the ground, chemicals in the air and as components of other objects, the earth's magnetic field, electricity, and numerous others. All of these forms of information are available to any living entity that can evolve a receptivity to that mode, and all have been used by various animals.

The experience of each of these modalities forms a 'map' of the world as experienced by the animal through that modality. The brains of (at least all higher order) animals are known to include topographic maps that reflect various modalities. These modalities can include virtually *any* source of external differential information about the environment that may aid the creature in ongoing survival.

In evolutionary terms, an improved world view may be obtained by refining existing 'filters', and/or adding new sensory sampling 'filters'. The more maps used by an animal, the more 'evidence' (or associations between modality maps) to carve up the world. Provided the information can be synchronised, several maps would allow for a better quality 'perception' of the world.

If a creature could evolve a suitable 'filter' capable of receiving some external information, and that information could be cross-linked with other incoming information so

that some form of synchronisation between information sources can be achieved, then the creature would have a means of altering its behaviour based on external factors. And having evolved a receptivity to a particular form of 'sense-data', the creature may be able to somehow use that information to behave appropriately in various circumstances, thus achieving a survival advantage over having no information.

Multiple different filters have evolved in animals. For example, sound (including ranges beyond human recognition), light (including infrared and ultra-violet), chemical (smell, taste), electricity (the platypus uses the electrical discharges in the muscles of its prey), magnetism, pressure (touch), and polarisation of light, are just some of the external phenomena used by various animals. The use of various external sensory "filters" appears to be consistent across the phylogenetic spectrum.

The simplest way to use this information would be to form a direct connection between the detection of a particular 'signal' from the mode, and some behaviour. This is a simple reflex action such as light avoidance in some insects, or the direct connection (via the spinal chord) between a tap to the patellar tendon of the knee of a human patient and the knee jerk reflex (Kandel & Schwartz 1983, page 18).

Reflexive links are common to lower creatures, but this form of direct connection is not very adaptive, and any change in external conditions could reduce the effectiveness of the reflex. However, reflexes are simple, direct and very fast.

An improvement might be to have some adaptive method of handling the input modality information. There is much evidence for topological maps in the brain, and as we show elsewhere, this may be brought about by a self-organising process. With self-organisation, the statistics of the external world (as recorded by the creature over time) form an adaptive linkage between senses and resulting behaviour.

One issue that has proven a stumbling block for the acceptance of self-organisation and association is the need for a narrow band of 'data'. This presents no problem for audition because we can readily separate out the various tones and intensities—for example, a loud shrill of a certain frequency may be associated with danger. But with the visual modality, the field of view of the retina appears to be too large. Associations with such a large field of view would not be possible, with too much extraneous detail

being recorded and confusing the required associations. What is required is a narrow field of view. But a creature could also utilise a large field of view to ensure that predators may be detected, or that prey can be ‘indicated’ in peripheral vision. This problem may be overcome by developing a greatly magnified central portion of the visual field—a fovea—as well as a less densely sampled peripheral component.

The peripheral field may then be used to direct saccades to regions of possible interest (perhaps determined by novelty detectors such as motion). By saccading to various parts of an object, connected input vectors will be sampled, enabling ‘recognition’ of the object. This forms the basis of the ABC model of vision as described in Chapter 4.

3.3.5 Perceptual Conceptualisation, Labelling, Language and Self-Talk

Most current models of cognition are predicated on linguistic terms. The cognitivist model, for example, is based in the syntactic manipulation of symbols, and certain localist versions of ANNs use ‘symbols’ (concepts) as input features.

The ABC model takes a completely different approach to cognition, especially as regards the role played by language. The model suggests that there are essentially four stages in the development of cognition in humans.

Perceptual Conceptualisation

The initial stage of cognition in humans (and higher animals) is the formation of perceptual concepts, [†] self-organised neural attractors which link sensory inputs with behavioural outputs.

[†]We need to be precise as to what we mean by a ‘perceptual concept’. A dictionary definition of *concept* is “something formed in the mind; thought; general idea” (McLeod 1988). We intend *perceptual concept* to mean some consistent, self-organised attractor within the brain that leads to a particular behaviour, given the appropriate sensory inputs. A perceptual concept may or may not have a symbol/label associated with it. An example of an unlabelled perceptual concept is *catching a ball*—given particular sensory inputs, one can behave to catch the ball. An example of a labelled concept is *dog*—again, given the particular sensory inputs of a dog, one can respond appropriately to the dog concept by perhaps patting it, or by running away, or even by speaking the word ‘dog’.

In agreement with Locke (1690), the model assumes that the brain of the neonate is a conceptual *tabula rasa* (or ‘white paper’ to use Locke’s term). Of course, there are numerous innate reflexive behaviours and autonomic systems which are pre-determined, and other behaviours such as walking are no-doubt pre-wired to a large extent. As well, the overall structures of the brain (as indicated by the ABC model components) are in general pre-wired, although further neural growth and decay can occur. However, all of our concepts (ideas) and even our perception of the external world are derived from subsequent experience through interaction with the environment and other internal bodily sources.

The initial learning mechanism employed by the infant is *differentiation* (and not *induction*). At the start of its interactions with a surrounding world, the infant will have no conceptualisation at all, not even to the extent of being able to discern separable ‘objects’. The model implies that the infant’s only experience will be a form of ‘white noise’. The first separable ‘concept’ will most likely be related to the principle carer, and over time other ‘concepts’ will form sensory attractors of their own and become separable. †

Objects in the near environment of the infant will provide substance and feedback to allow separate attractors to form for their recognition. Pets, toys and the like will become recognisable concepts within the infants world. Their associated sensory inputs, as well as feedback via action and reaction, will be self-organised into differentiable concepts. Initially random movements of the infant will at the same time become more coordinated as the motor attractors self-organise.

Gibson (1969) points out that the young infant at first singles out individual areas when looking at objects. The focus is on high contrast edges, vertices, spots and moving parts. This initial attention to single features then develops into the use of what Gibson (1969, page 345) calls “bundles of features”:

Distinctive features develop later out of these properties as contrasts are discovered . . . the specificity of discrimination is thereby increased. *Gestalten* or higher order structured units develop still later, as bundles of features

†It is inappropriate to denote the perceptual attractors as corresponding to objects at this early stage of development. The separation of concepts into objects occurs later in the development cycle.

are processed with greater simultaneity and relations between features are registered as units of structure.

The stages through which an infant passes in coming to recognise human faces is an illustration of this progression within perceptual development. At the earliest stage, the infant responds with a smile to dots or lines, whereas later the same response will require the dots to appear within some contour, with the eyes (dots) in the top half of the figure (face). After this, a mouth must be present, preferably in movement, and this must be in the correct relationship with eyes and facial contour. By the age of 5 months, the infant can distinguish between realistic heads and unrealistic ones. As stated by Gibson (1969, page 347): “development seems to proceed from simple contours to differentiated features to structured relations or patterns to unique patterns of individual faces.” †

The formation of attractors in neural networks is discussed in Jordan (1986), Venkatesh & Pancha (1990), Kelso, DeGuzman & Holroyd (1991), and Nicolis (1989). ‡ Attractor basins extend over multiple levels and multiple maps in the ABC system, and are regions of equilibrium in a complex phase space. Nevertheless, we can use a first approximation and consider them as those collective neurons on a single SOM sheet which, if selected as the winning node, will all lead to the same output result—either the same motor behaviour or the same linkage to a following level. This is illustrated in Figure 3.10. An input to the SOM as shown by the arrow will lead to *Action B*, as would an input selecting any other node of the same shading. Of course, the attractors within the model are found over several layers of the SOM surfaces, and may include recurrent linkages.

It should be stressed that this process is a dynamic one, and the differentiation of concepts into multiple new concepts is constantly proceeding. As well, there can

†Note that the infant is concurrently learning from *actual* human faces at the same time as these experiments are being performed.

‡It should be noted that while the output of a FFNN is a continuous function of the weights if each unit has a smooth output function, (such as a sigmoid function), the output of a simple recurrent network (SRN) can change drastically with an infinitesimal change in the network weights when it passes through a bifurcation point (Pearlmutter 1989, Pineda 1988, Guckenheimer & Holmes 1983). Thus the behaviour of recurrent nets is potentially much richer and more complex than that of FFNNs, and may include irreversible components.

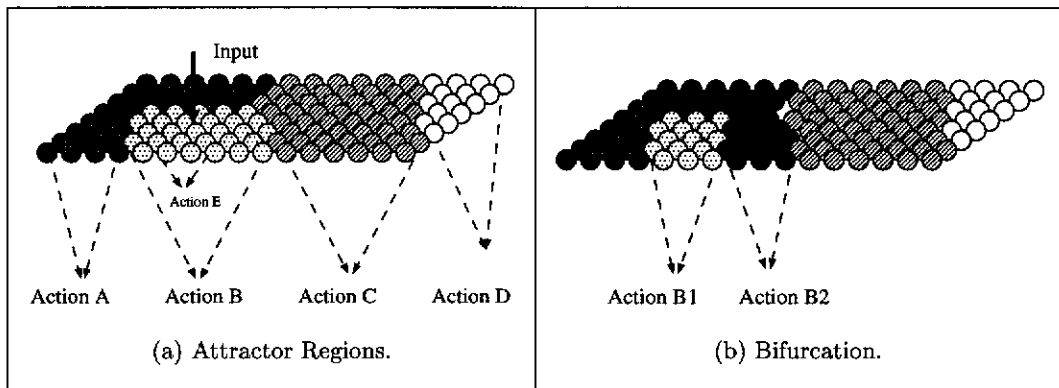


Figure 3.10: Attractor Basins and Bifurcation.

be the opposite process of two previously separated concepts (which once produced different behaviours), subsequently aggregating into a single behaviour and hence a single concept.

Perceptual conceptualisation links sensory inputs (both externally derived and internally generated) with behaviours. As we stated previously, conceptualisation and behaviour are inseparable—the concept is formed by the internal neural weights and linkages, while the behaviour is the external manifestation of that concept. Thus concepts are behaviour based—if there is no ‘need’ for a concept then none will be formed. For example, a bird may not form a concept of ‘tree-bark’ or ‘tree-roots’ as these differentiations are not needed in the bird’s environment.

The separability, discreteness and overall regularity of objects, properties and events in the external world allows for the separability and discreteness of perceptual concepts. The actuality of the world determines the extent and scope of conceptualisation. For example, Braunwald (1978) reports the case of a child failing to initially unify the visual inputs of a dog with the sounds of a dog barking. Subsequent experiences of the two concepts in coincidence, as well as common labelling (as described below) will fuse and relate the two components to the *dog* concept.

The process is one of discrimination. As stated by Harnad (1982), “two distinct inputs are discriminable in virtue of a receiver’s capacity to respond differentially to them; and two inputs that are responded to differentially are, by that token, discriminable.” The discrimination may result directly from the properties of the inputs, and may require a number of iterations of recurrent linkages to resolve. For example, a red

object may be discriminated from blue objects solely on the property of wavelength. Additionally, the discrimination may be brought about by some *internal* context—perhaps some memory of prior encounters with the objects, or some expectation, or even the current affective state of the creature.

To know something by intuition is to “receive knowledge by direct perception”.[†] This is the ‘feeling’ that something is a particular way without being able to *put it into words*. The outward expression of that ‘feeling’ is some behaviour—perhaps avoidance behaviour in the case of a feeling of ‘danger’. Another example of perceptual conceptualisation is the ‘reading’ of body-language.

The building of *expert systems* has been one of the very few success stories of Artificial Intelligence, albeit a limited success. Here an expert is queried so that the ‘rules’ that he or she uses in their expertise may be transferred and implemented on a computer. However, one of the major difficulties in this process is trying to get the experts to describe how they go about making decisions.

Having worked on expert system development with experts, it is clear that they are able to look at a problem and make a determination very quickly, but when asked to put their decision process into words they find it very difficult or are unable to do so. It is often only after a long and sometime tortuous analysis that a ‘knowledge engineering’ team is able to cobble together an ‘expert system’. The experts *do not* have the concepts in ‘linguistic form’, nor do they necessarily have associations to other related labelled concepts.

Perceptual conceptualisation also explains our ability to catch a ball, to play a forehand drive in tennis, and many other behaviours which do not require a linguistic form. Other examples are art appreciation, gestalt phenomena, the ability of a chess grand master to ‘see’ the next move, the reading of an x-ray image by an expert and many other behaviours that are based on sensory inputs and a ‘feel for it’. The outward manifestation of the perceptual conceptualisation is *skilled* behaviours.

There are no rules and no linguistic component. Practice and repetition plays a major role in learning and honing a perceptual skill. For example, we need to hit a tennis ball many hundreds of times, make slight adjustments to our forehand shot (possibly

[†]The Australian Pocket Oxford Dictionary.

through linguistic instruction from a teacher or by self-talk as will be discussed later) until the attractors linking the sensory inputs (including proprioception) and the output motor behaviours for the forehand shot are finely tuned. The process requires dynamic temporal sequence learning involving feedback and 'expectations' as we discussed in Chapter 2.

It is instructive to note the stages in child development observed by Piaget (see, for example, Cowan 1978) during this initial period, which is termed the *sensorimotor* stage by Piaget, and is broken down into six sub-stages:

- Sub-stage 1—*the emergence of directed behaviour (birth to 1½ months)*.

Beginning at birth, reflex reactions (such as grasping) are initiated and motor actions start to become more coordinated. For the first month or so, each sense modality is independent of every other, but soon the infant is able to sustain attention to a stimulus as long as it remains in the field of vision or hearing or touch. Generalised bodily movements are replaced by more active exploration.

- Sub-stage 2—*coordination and early goal direction (1½ to 4 months)*.

There is a beginning of coordination in sensorimotor behaviours, and seemingly chance activities are repeated endlessly. The infant begins to imitate an adult, provided the adult first copies an action just made by the child. Perceptions and understandings of objects are still tied to the infant's actions and are not preserved when the object disappears from view.

There is, as yet, no cognitive separation of self from physical or social objects, and the infant's preoccupation is with his body. Infants react responsively to people as interesting objects, but there is as yet no truly *social* interaction.

- Sub-stage 3—*intention; beginning of independence of actions and thought (4 to 6 months)*.

The infant may actively search for, and select, certain movements which have given rise to interesting effects, indicating the beginning of *intentional* goal-directed activity. Behaviours are modified and generalised to new situations, and the focus shifts from the child's own actions to consequences in the world. Piaget suggest that behaviour directed by conscious goals does not usually emerge until this sub-stage.

•Sub-stage 4—*the first permanent object (6 to 12 months)*.

The first real differentiation between assimilation and accommodation is observed through longer behavioural sequences. Objects are perceived as permanent for the first time, although this permanence is initially inconsistent. Also at this stage, the first real imitations are observed, and both gestures and sounds new to the child are copied.

Rather than simple repetitive functions, some form of preparation, practice and coping is observed. The fascination with peek-a-boo games occurs at this time.

•Sub-stage 5—*knowledge acquisition through trial and error (12 to 18 months)*.

The child shifts to an investigation of causes of events. Random trial-and-error groping is coming under the control of already-established behaviours as if knowledge of the physical world has become a spontaneous, conscious goal.

Coordination of behaviours is still limited to sensorimotor rather than mental combinations, although language starts to emerge. Play begins to be combined with imitation in the form of rituals—repetitions of actions which the child has performed at some other time.

•Sub-stage 6—*combinations of symbolic behaviours (15 months to 2 years)*.

This stage represents a culmination of the sensorimotor development and a transitional period to mental behaviour. Although symbols are starting to be internalised, it is in a sensorimotor form that still lacks the flexibility and reversibility of the conceptual symbols which come later.

Problem solving is achieved through cognitive trial and error, but it is clear that solutions are not insightful but achieved over long periods. Object permanence is now fully developed, and the child is able to follow extended and coordinated behavioural sequences to achieve a goal.

Symbolic play and deferred imitation extend the child's cognitive trial and error investigations to the emotional and social arenas.

However, the sensorimotor behaviours and memory, even in their internal form, are still tied directly to concrete events which have been experienced by the child. They are as yet unable to think of hypothetical events.

For Piaget, the entire sensorimotor period represents a rich foundation for all conceptual and symbolic thought, with crucial developments in object conceptions occurring before the development of the spoken word.

We contend that humans share this initial stage of cognition with higher animals. Animals are also able to form perceptual concepts based on sensory data and linked to behaviour; for example, a trained dog doing a trick on verbal command. Note that there is a continuity between perceptual categorisation and innate behaviour. While one is hard-wired in the neonate, the other is a soft-wired reflex learned through self-organisation during the life of the creature.

We can say that behaviour which relies solely on perceptual conceptualisation is 'non-conscious' behaviour in the sense that it is not able to be verbalised. † It is not linked to the linguistic component of cognition, and as we discuss later, linkage to the linguistic component of cognition via self-talk is in large part the requirement for 'consciousness'. As these perceptual categories and behaviours are not linked with linguistic labels or sequences, we are unable to talk or self-talk (think) about them, and so they do not form 'conscious' behaviours. The subject will *not be conscious* of why they undertook the particular behaviour associated with the perceptual conceptualisation, even though they may be 'consciously aware' of their actual behaviour. For example, the expert may not be able to put into words just why they made a particular decision, but they are well aware of the decision.

Perceptual conceptualisation attached to behaviours continues throughout the person's life, (usually referred to as *skills*), and may include behaviours such as tennis, gymnastics and driving a car.

Perceptual conceptualisation has of course been discussed previously by a number of others in various fields. These include the idea of nameless concepts (Nelson 1974), and cognitive meaning in advance of language (Bloom 1973, Clark 1973b, Clark 1973a, Greenfield & Smith 1976, Huttenlocher 1975, MacNamara 1972, Nelson 1973, Nelson 1974, Piaget [1946] 1962, Sinclair-deZwart 1973, Slobin 1973). The excellent monograph by Eleanor J. Gibson (1969) discusses the principles of perceptual learning and development. There is also a whole literature on *implicit learning* which extends from

†Note that we do not imply that it is *impossible* to verbalise perceptual concepts, but that there is no linguistic linkage available to the subject unless and until a labelling is put in place.

Helmholtz's ([1910] 1962) concept of "unconscious inference", through to recent studies by Schacter (1987a, 1995), and especially the learning of skills by Hayes & Broadbent (1988), Lewicki, Czyzewska & Hoffman (1987) and Willingham, Nissen & Bullemer (1989).

Labelling

The second stage in the development of cognition is labelling—the association of essentially auditory † sequences with existing perceptual categories. This model turns around the usual linguistics/cognitivist discussion in which words have a separate existence in the mind, and are somehow later linked to some meaning in the world—the 'symbol grounding problem'.

An important issue concerns the mechanism of association of sounds (labels) with other sensory inputs experienced by the child. Adults monitor very closely their infant's focus of attention and adjust their own gaze to maintain shared experience (Butterworth & Grover 1988, page 9). A mother will monitor the focus of her infant's attention very closely, and will vocalise at suitable moments in the interaction, thus creating a tutorial environment that Bruner (1983) called the *language acquisition support system*. Schaffer (1984) has shown that the majority of episodes of joint activity arise as a result of the mother monitoring the infant's line of gaze. Not only does the mother make use of the child's line of gaze, but the child also redirects his gaze to share the focus of the mother's attention (Scaife & Bruner 1975, Churcher & Scaife 1982). Further, Butterworth & Grover (1988) provide evidence that in the first 18 months of life, the infant uses a number of mechanisms to enable it to look where others are looking. Joint visual attention also allows for deictic gestures, so that for example, attention may be drawn to a particular object by referential pointing.

In Chapter 4 we describe a model of vision consistent with the ABC model. This model rejects the current *spatial-fusion* model in which retinal images from successive saccades are subsequently fused together to form an internal 'pictorial' representation of the external world. Instead, we propose that there is no fusion, and the appearance of full peripheral vision is an artifact. The mechanism proposed in Chapter 4 suggests

† Another perceptual means of labelling a perceptual concept is through the use of visual signs made with the fingers and hands—the initial stage in learning a sign language such as American Sign Language (ASL).

that vectors obtained from foveal images are learned as temporal sequences. The point to be made here is that the *foveal* view of the child is able to be temporally associated with a concurrent auditory input (the mother's naming of the object or event).

Children's early word usage does not correspond exactly to the same meaning as an adult's use of same word (Bjorklund 1989, page 143). One frequently observed phenomenon is *overextension*, in which the child supposedly extends the meaning of a word to cover a broader range of referents than the adult meaning actually covers. For example, a child may call any four-legged animal a 'doggy', or 'horsy'—whatever their initial label for a 'large-separable-hairy-friendly-thing' was. The perceptual concept for such an object must already exist within their brain prior to the linkage of the label.

In fact, the overextension may be much greater than a simple inclusion of all 'similar' animals. For example, Braunwald (1978) discusses the acquisition of the first 50 words by her daughter, Laura, from age 8 months to 2 years. † She emphasises the acquisition of two words in particular—*ba*, which to Laura extends to meaning "ball, round object, milk and liquid served in a cup", and *bow-wow* which was used by Laura to refer to "dogs and dog-like toys, miscellaneous animals and cars as well as the sound of barking, car engines, airplanes, birds' chirping and noises from the outside which are audible in the house".

Within the ABC model, early labelling may refer to extended perceptual concepts. As further cues are discerned, associated and self-organised, the initial perceptual concepts may be bifurcated (split) on differential sensory inputs. As well, existing labels may be refined and new labels added. For example, the initial 'doggie' may become 'dog', 'horse', 'cow' and 'Fido'—a particular individual dog. Whereas all animal-like-things were 'doggie', the particular brown one with the white stripe and particular high-pitched yap is separated out as that animal with the label 'Fido'.

The traditional view of this process is that the child *is* able to discern all of the animals as being of different *types*, but as they do not yet have the language terms for these other concepts, they generalise the word meaning to describe objects which are perceptually similar to the original referent (Clark 1973b), or have a similar function (Nelson 1973). The meaning of the word is described as having a number of semantic

† See Bowerman (1978) for another similar study.

features [†] and it is assumed that the child gradually acquires a complete set of features for a particular concept, thereby restricting the number of referents to which the word can apply.

We suggest that this account is incorrect on at least two points. Firstly, we contend that the child is not yet capable of discerning the animals as being of different types. Initially, the child will not even be able to separate figure and ground, so that the very concept of ‘object’ is not appropriate at this early stage in the child’s development. The differentiation of sensory inputs into recognisable objects and events is incomplete, and the child’s perceptual attractors (which enable the differentiation) will accept all of the perceptual inputs from each of the animals as if they were the same. The viewing of each of the animals by the child will elicit exactly the same response—the behaviour of speaking the word ‘doggy’.

This point needs emphasising. It is not that the brain of the child will not later self-organise so that it can then discriminate the various extended meanings—a statement about capability—nor that the perceptual means used by the child (for example, their eyes and ears) cannot differentiate the input signals—a statement about equipment—but rather that their brains are initially organised such that the broad perceptual attractors associated with the sensory inputs of the extended meanings are currently all linked to a single output behaviour. The self-organisation of their visual (and auditory) system is still incomplete, and is proceeding at the same time as the labelling of current perceptual attractors.

The second point is that we reject the linguistic basis for initial perceptual conceptualisation. As we discuss further in Chapter 5, the view that learning is induction over linguistic terms must be abandoned. The premise of linguistic terms being based on other (more primitive) linguistic terms (features) as necessary and sufficient conditions has a long history in western philosophy. This requires a set of absolute primitives (simples) from which all other word meanings can be obtained. We reject this *representationalist* view (following the seminal work of Wittgenstein (1963) and Heidegger (1978)), and instead propose that conceptualisation is based on *sensory primitives*, as exemplified by spatial frequencies and angles, motion and colour for vision, frequency

[†]For example, the concept *bird* may include the features *animal*, *flies*, *has-feathers* and *has-beak*, as well as some specification of typical sizes and shapes.

and intensity for audition, and so on—that is, *vector* ‘primitives’ rather than semantic primitives.

The ontology of this process needs emphasising. The traditional view has it that the infant needs to *figure out the referent* of a word when the labelling event is ambiguous. Her development task is to discover the match between a word which she hears and its real-world referent (MacNamara 1972). But this account is much too simplistic. The child does not yet have command of *words*, nor the ability to separate out specific objects or events in the world. It must also be emphasised that the process of retrieving a label by speaking is a behavioural response—the child is learning a particular behaviour to accompany certain sensory and motivational inputs.

Vygotsky ([1962] 1986) stressed that word meanings grow and change within the child, and he regarded this as central to an understanding of the development of thought and language. For example, Braunwald (1978) emphasised that the listener of the infants uttering must use the situational context to understand the communication request of the infant. The actual words † used by the infant may initially convey a variety of ‘meanings’. Some example word/meanings from the study are shown in Table 3.2.

The words in the child Laura’s lexicon are idiosyncratic and context dependent. Most of the initial 50 words used by her depend upon the social interaction between the child and her interlocutor within the framework of specific situational contexts. A knowledge of the social context is required in order for a word to function as meaningful communication. Only eight of the initial fifty words used by the Braunwald child are used with essentially the same range of reference, and hence meaning, as the standard adult word. Near half are partial phonetically-incomplete imitations of adult words, with the use of onomatopoeia accounting for 3 out of 4 self-invented words.

The words relate to nominals (*Father, Mother, cookie, cheese*), actions (the word *oh-oh* is used to express notice of perceptually unexpected events), modifiers (*hot, bye* meaning ‘all gone’) and personal social references (*me*—a whining, volitional word intended to make the listener aware that she is unhappy about something).

Nelson (1973) reported that among the children studied by her there were marked

†Braunwald suggests that the use of the term *word* in relation to infants at this stage of development refers to any recognisable union of a phonetic form and a meaning.

Child's Word	Adult's Word	Range of use in order of occurrence	Range of use at 16 months
Dada/daddy [0;9(8)]	Father	Father; doll; baby in a photograph; Laura herself in a mirror; Mother in photograph; people in general; sister; to call attention to herself; occasionally to any man	Father; sometimes to designate any man
Bow-wow [0;11(7)]	Dog	The sound of (a) barking, (b) an airplane, (c) car engine, (d) birds, (e) any outside noise audible in the house; a toy dog; the sight of a car; the sight of a dog.	Dogs and barking only.
Ba [1;0(9)]	Ball	Ball; round objects including a grapefruit, an orange, a seedpod and the (round) doorbell buzzer; to request first and second servings of liquid in a cup.	The entire range of use remained productive.
Da [1;0(23)]	Down	To request locational or positional change. Used where the words up, out and in would be situationally appropriate.	Specifically used for request to get down from her high chair or in other situations where down is the appropriate request for positional change.
Ooo/Hoo-oo [1;2(18)]	Hot	Hot food; hot car seat; hot pavements; cold car seat; ice cubes; to the oven.	To things which are hot; to the oven

Table 3.2: The Form and Meaning of Infant Words. (Adapted from Table I Braunwald (1978)).

individual differences in the 'types' of words used—some used many object names while others used more words expressing feelings and needs. Further, the use of words does not correspond exactly to adult word types, making it difficult to classify words according to this taxonomy. Even common nouns, when used by an infant, appear to function not just as a label but rather as a substitute for some more 'complex utterance' (Benedict 1979).

Braunwald concludes that Laura *extends* the meaning of some of her existing words as a strategy for increasing her vocabulary. We contend that this is not the case, but

that the extension is already in place, brought about by insufficient differentiation. As stated previously, the process of associating labels with referents is one of differentiation from overgeneralisation rather than individual allocations of labels to referents.

All overextensions appear to be based on some similarity which the infant perceives between the (adult extended) referents. The similarity could be a subjective, affective state experienced by the child as well as a perceptually objective or functional feature of the referent. As an example, the word *cookie*, while initially referring to cookies and crackers, was extended on the basis of perceptual and functional features to include other unusual foods, and by other subjective or affective dimensions to include other forms of pleasure such as (a) to request music on the hi-fi or car radio, (b) to request someone to rock her, and to identify the rocking chair, and (c) to talk about ice cream.

Braunwald (1978) performs an extensive analysis of the evolution of various words as used by the child. One such word is *ba*, which was initially learned in association with a ball but was subsequently used in reference to all other objects and actions with some 'roundness' component, including drinking milk from a cup. Her conclusion is that "the differentiation of *ba* in the drinking event can best be described as the progressive recoding of comparable repeating situations into increasingly more explicit and informative terms."

A progression in restructuring an overextended word is described by Clark (1973b, page 85), and is illustrated in Table 3.3. The original use of the word *bow-wow* was in reference to dogs. The subsequent overextension is gradually refined and restructured to discriminate the different animal groups. In line with the generally accepted explanation of this phenomenon that the narrowing of meaning is brought about by the addition of certain 'features', Clark provides a number of what he believes may be possible features in column 3. Other examples of overextension and subsequent restructuring are also given in the article, as well as examples of the differentiation of closely related words.

The ABC model suggests an alternate explanation. The narrowing of the referent field meaning is brought about by a bifurcation of pre-existing perceptual attractors, and a subsequent linking of new terms providing differential output behaviour. The bifurcation is indeed based on 'features', but within the ABC model the features are low-level features such as spatial frequencies and angles, motion, colour and so on.

Word	Semantic Domain	Possible Criterial Feature(s)
bow-wow	dog(s)	shape
bow-wow	dogs, cows, horses, sheep, cats	shape
(a) bow-wow (b) moo	dogs, cats, horses, sheep cows	sound, (horns?)
(a) bow-wow (b) moo (c) gee-gee	dogs, cats, sheep cows horses	sound size, (tail/mane?)
(a) bow-wow (b) moo (c) gee-gee (d) baa	dogs, cats cows horses sheep	sound
(a) bow-wow (b) moo (c) gee-gee (d) baa (e) kitty	dogs cows horses sheep cats	shape, sound

Table 3.3: Overextension and Restructuring of Words. (Adapted from Table 8 Clark (1973b)).

As well as overextensions, some underextension may also occur. For example, the word *goggie* (referring to doggie) was used by Laura in relation to a dog on a toy block. Presumably, some component of the block differentiated this ‘dog’ from the more general term *bow-wow*. As well, various components of an adult concept may initially be learned by the infant as separate, unrelated perceptual concepts. Braunwald describes how her child initially matched the label *bow-wow* to separate visual and auditory concepts. On the visual side, *bow-wow* referred to dogs and cars, while on the auditory side it referred to the sounds of barking, car engines, birds’ chirping, overhead airplanes and other outside audible noises. Subsequent reinforcement by other speakers, and exposure to coincident occurrences of the concepts (i.e., viewing a dog while it barks) will re-shape the perceptual concepts to conform to the regularities in the world.

At this stage in the child’s development, the function of the word is to designate or isolate one aspect of a situation from others. Thus the word has an indicative but not a signifying function—Vygotsky suggested the term “oral indicative gesture”. For

the child, the word is for some time more a *characteristic* of the thing than a *symbol* for it. Vygotsky maintained that the child does not attain conscious awareness of the symbolic significance of speech for a long time. It is also clear that at this time the word does not have a constant or stable objective meaning.

The conclusion of Braunwald (1978) is that “cognitive, social and linguistic knowledge must somehow dovetail together to permit [the child] to discover the linguistic and social conventions used by her speech community for encoding informatively a given intention in a given setting.”

Rather than it being an active process of the child trying to figure out a jig-saw puzzle of fitting words to referents, we contend that the process is one of simultaneous self-organisation of many components, including visual and auditory maps, as well as dynamic feedback from the environment and other members of the social group.

The interpretation of language by a child is strongly embedded in the context of the occurrence. For example, a reference to ‘milk’ is usually accompanied by the actuality of milk being offered to the child. MacNamara (1972) argues that it is only to the extent that children understand the *situation* that they are able to work out the meaning of words.

Although perceptual conceptualisation does not require linguistic terms, the labelling of concepts allows for much richer and complex concepts to be formed. Learned labels may also be used for further differentiation within concepts. Consider a simplistic example—to many, all wines taste roughly the same, with only broad differentiation on say, the level of sweetness. Their ‘concept’ for wine taste is broad. An expert taster, however, will have learned to differentiate many subtle tastes and aromas, as well as different viscosities and colours. Through training and practice, and by associating new labels to the subtle variations of taste (such as ‘old leather’, ‘woody’ or ‘sulphur’) and other sensory inputs, the expert is able to form finely discriminating wine concepts. The labels become part of the discriminating process by providing differentiated behaviours for the fine differences in perceptual inputs. The labels themselves are then available for subsequent associations and temporal sequence learning.

Humans share the capability of labelling perceptual concepts with higher animals. For example, some monkeys have been trained to associate auditory or visual labels with certain objects and events, and the chimpanzee Washoe was trained to communicate

in American Sign Language (ASL), eventually responding to more than sixty signs (Gardner & Gardner 1969, Gardner & Gardner 1975). See also Terrace (1985) for other animal studies.

Any model of cognition must be able to account for the labelling of objects and concepts. This inability to account for the origin of labels, and the lack of any means whereby new labels may be generated, is a serious problem for other cognitive models, including cognitivism. The ABC model provides a mechanism for the origin of labels—they are simply attractors (or combination of attractors and temporal sequences) formed in the auditory modality (in the case of spoken labels), or other visual attractors (or combination of attractors and temporal sequences) in the case of written labels.

In summary, the process of labelling of perceptual concepts is not one of 'learning the meaning of words' in the usual sense, but rather one of attaching an arbitrary word (label) to an already established (albeit only partially established) meaning unit—a perceptual attractor. The child may perceive either a similar form or common function (or both) in learning an overextended meaning for a word; that is, which goes beyond the limits of the accepted adult norm. Non-linguistic contexts are essential for the early stages of the growth of word-meaning.

Language

A child's use of single-word utterances gradually shifts to a wider range of communicative acts. Initially words are combined in simple two-word 'sentences' such as 'bow-wow gone' or 'that cookie'. The classes of the words still do not correspond to adult forms. Studies by Slobin (1972) and others have found that language development lags behind cognitive development—the child must understand what he or she wants to say before he can say it.

The acquisition of linguistic competence is a subject of much debate. Some suggest innate modules specifically for grammar (Chomsky 1965, Chomsky 1975b), while others such as Piaget ([1923] 1960) suggest that grammatical expression depends upon the sensorimotor development of the child—for example, the child must act upon his environment before he is able to develop the notions of *actor*, *action* and *object* and to use these as components of grammatical constructs. Piaget's theory suggests that cog-

ognitive and language development is a progressive reorganisation of cognitive processes as a result of maturation and experience.

According to Piaget, children construct an understanding of the world around them through experience. He introduced the concepts of assimilation and accommodation to explain the continual process of understanding the world through experiencing discrepancies between what they already ‘know’ and what they discover in their environment. When a child discovers something that is reasonably similar to what they already know, it is *assimilated* into their existing knowledge. However, should they encounter something that is different from what they know, they either ignore it or modify their way of thinking to *accommodate* this new knowledge.

Piaget suggested that there are a number of stages of cognitive development through which each child must progress. The sensorimotor stage (from birth to around age 2) involves essentially trial and error learning. Children tend to rely on reflexes, modifying them to adapt to their world. During the preoperational period (from about ages 2 to 7), children are able to ‘mentally represent’ events and objects (the semiotic function), although their thoughts and communications are typically egocentric, and they are limited to a single focus consideration of problems. The concrete operational period (usually from age 7 to 11) finds children learning the rules of conservation and reversibility. Their thinking is typically more organised and rational, but they are usually not able to think abstractly or hypothetically. Finally, during the formal operation period (beginning at around age 11), the adolescent is able to think in an abstract manner and is capable of higher-order reasoning.

Vygotsky ([1962] 1986) has stressed the importance of social context in the development of language and thought. According to Vygotsky “the true direction of the development of thinking [or language] is not from the individual to the social, but from the social to the individual” (Vygotsky [1962] 1986, page 36).

As was indicated in Chapter 2, language is seen within the ABC model as learned temporal sequences of meaning units that are based on inputs provided by other members of a society. These meaning units (labelled concepts) are usually initiated as sound sequences coming from the speech of others, but may be from other sensory sources such as hand movements. The use of words and sentences by members of the society, both directed towards the infant and to others, and in the specific contexts in which the

words are appropriate, will, over time, allow the child's brain to associate, self-organise and learn temporal sequences so that the child comes to learn and duplicate the words and word sequences of the society.

The ABC model does not require the presence of 'grammars' in the brain, and we reject strongly the 'string-processing' model of languages. † Grammars are seen as external representations created by linguists (in the same way that a house plan is an external representation). The apparent uniformity of language is an artifact imposed by the consistent language usage within a speech community. The consistent language usage provides consistent sensory cues for temporal sequence learning.

The learning and production of linguistic sequences continues during development, and becomes a major component of cognition in its own right. We are able to learn and use abstract words that are only loosely attached to real objects and events. Moreover, as we discuss in Chapter 5, labelling can be used to further finely bifurcate perceptual concepts, enabling a rich description of objects and events in the world.

Much work needs to be done to link together the empirical studies of Piaget, Vygotsky and others, with the self-organisational processes of the ABC model. However, based on this (admittedly cursory) examination, the outlook is promising.

We defer most of our discussion on language (and self-talk) until Chapter 5. However, we need to make one point here; that language is another form of skilled behaviour in the same way that hitting a tennis ball or playing a piano is learned behaviour. All are concerned with the learning and reproduction of temporal sequences. The ABC model treats language in exactly the same manner as other learned skills.

Self-Talk

'Thinking' for most people is usually thought about in terms of language (Bjorklund 1989, page 140). Language is not the only component of thinking, but most of our thinking appears to be involved with language usage, only with ourselves as listener—self-talk.

Language is such a large component of thinking that most current theories of cognition have it as the *mechanism* of cognition. The cognitivist model places language at the centre of its model, with cognition thought to be an essentially linguistic pro-

†See Chapter 5 for a definition and rebuttal of the 'string-processing' model.

cess. The ABC model rejects this positioning of language as the central mechanism of cognition without reducing the importance of language to thinking.

Both Piaget (Piaget [1923] 1960) and Vygotsky (Vygotsky [1962] 1986) describe two distinct categories of childhood speech—*egocentric* speech and *socialised* speech. The two are in general agreement on socialised speech. For Vygotsky, “the primary function of speech, in both children and adults, is communication, social contact.” (Vygotsky [1962] 1986, page 34).

Egocentric speech is language used by a typical 3 to 5 year old child that is directed to *himself* rather than as a form of communication with someone else. According to Piaget ([1923] 1960, page 33), “the child talks to himself as though he were thinking aloud. He does not address anyone.” This *private* speech is observed both when children are alone and in social settings.

Piaget believed that the egocentric speech of children simply reflects their general egocentric perspective of the world. As the child becomes increasingly able to de-centre their cognition and perception, and see the point of view of another, their private speech decreases. So for Piaget, egocentric speech plays no functional role in cognitive development but is merely symptomatic of ongoing cognitive activity.

In contrast, Vygotsky believed that private speech plays a crucial role in the development of the child’s thought processes. He believed that egocentric speech is a *vocalisation* of the internalised “speech” that occurs within the child. Vygotsky termed this internalised speech “inner speech”, but we prefer the term ‘self-talk’ to emphasise that this language component of thought is a communication with oneself.

In order to benefit from the self-regulatory function of language (see Luria 1961), Vygotsky felt that children must essentially talk to themselves, using their speech to guide their thoughts and behaviour.

Vygotsky indicated that there is a strong link between social experience, speech and learning. The aspects of reality that a child is ready to master lie within what he called the zone of proximal (or potential) development. This term refers to the range of tasks that a child is currently unable to accomplish without guidance from an adult or more skilled peer.

Egocentric speech enables the child to direct their own behaviour, acquire new skills and otherwise work through unfamiliar situations. When a new task is encountered,

the child will initially state out loud those features of the problem that seem puzzling. Later, as their competence grows, this private speech turns into inaudible muttering, until finally, when the cognitive operations necessary to succeed at that task are well practised, the child thinks the words silently.

Vygotsky maintained that “the most significant moment in the course of intellectual development . . . occurs when speech and practical activity, two previously completely independent lines of development, converge.” He maintained that social communication underlies all uniquely human, higher cognitive processes. It is through a process of communication with mature members of society that children learn to master activities and think in ways that have meaning in their culture.

Egocentric speech may be seen as the child ‘thinking’ out loud. Over time, this private speech becomes internalised as silent, inner speech—“ . . . those conscious dialogues we hold with ourselves while thinking and acting” (Berk 1994, page 62). Vygotsky hypothesised that these self-guiding comments help children direct their actions. Children tend to talk to themselves more often when working alone on challenging tasks, and when the child needs to take charge of their own behaviour.

The amount of social discourse determines the onset time of private speech. American middle-class children speak out loud to themselves with increasing frequency between the ages of four and six years. However, within other cultures, the onset time may extend to later ages depending on how frequently parents converse with their children—less social discourse between child and other members of their society usually results in a slower development of private speech (Berk 1994, page 62).

It has also been observed that children who progress more rapidly from audible remarks to inner speech are usually more advanced in their ability to control motor activity and focus attention. The development of private speech and task-related behaviour go hand in hand. Further, as performance in a task improves, private speech diminishes. †

Advice from adults provides children with the means they need to use private speech effectively. Giving assistance with a challenging task provides the child with spoken directions or strategies to help it succeed. The child may incorporate the language from

†We discuss this aspect of cognition in more detail in Chapter 5.

these conversations into private speech and later use this language to guide their own efforts. As expected, private speech predicts future gains better than concurrent task success.

Vygotsky's views on private speech have recently been corroborated by a number of researchers including Berk (1994), Diaz & Berk (1992), Berk & Garvin (1984) and Berk & Landau (1993).

Vygotsky maintained that the young child is bound to the perceptual field. He suggested that it is through the development of the child's play activity that thought and meanings are liberated from their origins in the perceptual field, providing the foundation for the further development of speech and its role in advanced forms of thinking and imagination.

The child's early play concentrates on reproductions of situations or actions that the child has actually experienced. For example, in playing with dolls, the child reproduces the actions of his caretakers. This stage Vygotsky saw as transitional to more abstract forms of thinking, and is initiated when the child uses one object to function as another in a play situation—for example, using a stick in place of a horse. Even then, Vygotsky argued, the child's initial substitution does not represent a manifestation of symbolic capabilities as exhibited by adults, and it is only within the activity of play that the child begins to separate the object's meaning from the object itself by using another object as a pivot. This process then enables the child to develop forms of imagination and abstract thought.

What is clear, according to Vygotsky, is that thinking depends on speech, on the means of thinking, and on the child's socio-cultural experiences.

The ABC model proposes a radical new alternative for the process of inner *linguistic* thinking which is in agreement with the observations of Vygotsky and Piaget. We propose that a linkage exists in the human brain that connects an area adjacent to the vocal motor areas, to another area adjacent to the auditory input region. A vector of 'pseudo' or simulated 'voice output' that uses the same 'codes' as those found in the input auditory areas, would then form the basis of the pseudo-auditory (or linguistic) component of thinking.

The ABC model for ordinary speech is indicated in Figure 3.11. The auditory

world-filter receives sounds from the environment [†] and the resulting vector is processed within the brain. The speech behaviour requires a previously learned speech motor action [‡] to be reproduced, with the output vectors from the subsequent temporal sequence connected to the speech generating muscles.

Figure 3.12 shows this model extended with the inclusion of a further *internal* recurrent linkage between the motor output areas and input auditory areas. The additional linkage provides an additional source of ‘sound’—the silent sound of verbal thinking. Provided that the child is able to learn *voice-equivalent* vector components, he or she now has another source of verbal *instructions* or *communication* that is self-generated.

We propose that an area adjacent to the motor cortex that controls the muscles of the face, tongue, jaw, and throat (one such area is Broca’s area) is available to learn

[†]We make no comment here as to the possibility of multiple auditory filters, one or more of which may be more finely tuned to receiving human-voice sounds.

[‡]To head off the obvious objection that whole sentences cannot be learned, we remind the reader that the temporal experiments in Chapter 2 and Appendix B which relate to *context* indicate that the learning of whole sentences is not required. The reproduction of learned *words* in contextually generated sentences will enable arbitrary and novel sentences to be generated.

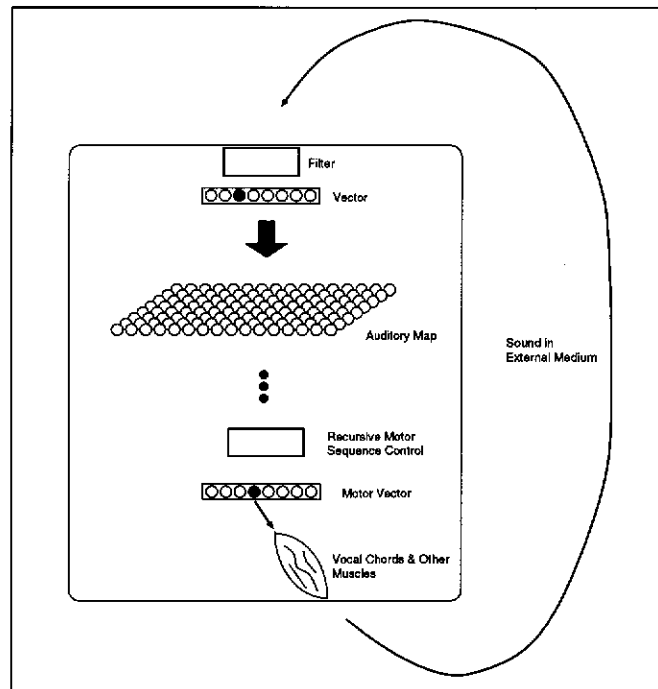


Figure 3.11: Vocal Sounds in External Medium.

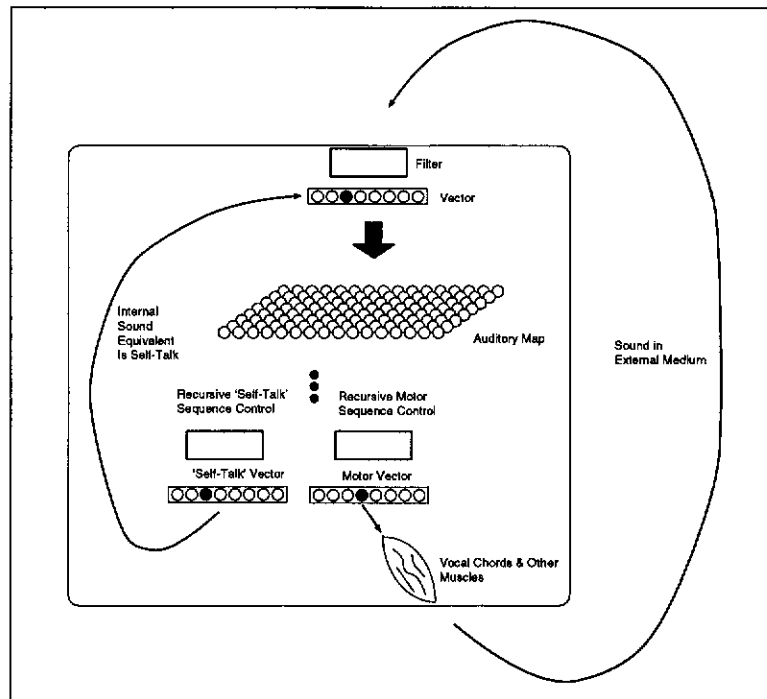


Figure 3.12: Internal Self-Talk as Sound-Equivalent Vectors.

the pseudo-voice vector codes and sequences required for inner speech. Wernicke’s area is an area adjacent to the primary auditory areas, and is connected to Broca’s area by a collection of nerve fibres called the *arcuate fasciculus* (Kandel & Schwartz 1983, Graham 1990). We postulate that this linkage via the arcuate fasciculus may provide

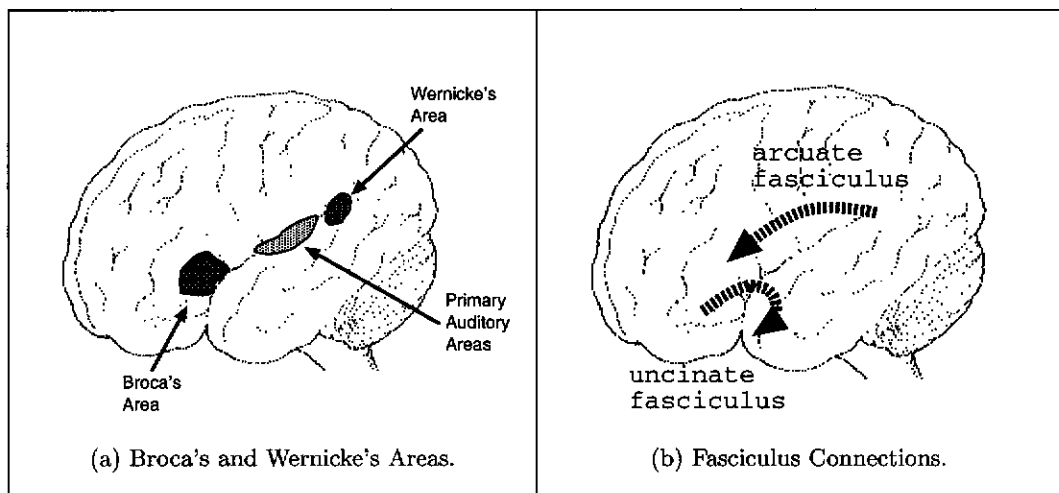


Figure 3.13: Postulated Self-Talk Areas and Connections.

the means of supplying Broca's area with copies of the auditory codes required for self-talk. The recurrent link for the return of the self-talk codes may be provided by the *uncinate fasciculus*, another bundle of long fibres which connect the orbital, middle and inferior gyri of the frontal lobe with the anterior region of the temporal lobe (Diamond, Scheibel & Elson 1985, Carpenter 1991). This region is near to, or forms part of, the tertiary association area of the auditory system. †

The relationship of Broca's and Wernicke's areas is shown in Figure 3.13 (a). This figure also highlights the relationship of Wernicke's area to the primary auditory area. Figure 3.13 (b) indicates the postulated long fibre connections.

Note that the codes and sequences that drive throat and facial muscles to make the sounds of speech are not the vectors required for this internal recurrent loop. Rather, 'copies' of the incoming auditory vectors are required. This is consistent with the structure of the SOMA and LAPS modules which were described in Chapter 2.

The use of egocentric speech by children is then postulated as a training period during which the child is able to learn the internal vector codes, initially in conjunction with external speech. The requirement for external speech falls away as the internal codes, and thus internal speech, are learned.

Thus, in the ABC model, thought (at least the component of thought that is self-talk), is seen as an extension of motor behaviour that has become internalised. We choose to also call self-talk a behaviour, even though there is no actual 'action' other than a flow of neural axon discharges. In this way we are able to show a continuity between language motor behaviour and self-talk.

The processes of speech and self-talk are separate. In most circumstances, they are coordinated, but sometimes we may find ourselves saying and thinking different

†Except for the terminology of 'program', this description is similar to that of Bloom & Lazerson (1988, page 284):

... the underlying structure of an utterance—its form and meaning—arises in Wernicke's area. It then passes through a collection of nerve fibres, called the *arcuate fasciculus* to Broca's area. There, the impulses evoke a detailed and coordinated program for vocalisation—a program for how each lip, tongue and throat muscle must move. The program is then transmitted to the adjacent area of the motor cortex that controls the face, and the appropriate muscles are activated.

things—with ‘two voices’. As well, one can be easily transformed into the other and each develops under the other’s influence.

There are other components to ‘thinking’ such as mental imagery. In a similar manner to which we have postulated an internalisation of auditory sounds to form self-talk, mental imagery may be explained as the result of a recurrent linkage within the visual modality. However, there is an essential difference between the recurrency resulting in self-talk, and that of self-imagery. Language is an active, social process that may be extended and reproduced readily, whereas the visual system is somewhat passive and individual, relying on personal experiences. Mental imagery plays less of a role in thinking than does self-talk. We can include self-talk in the full system diagram, as shown in Figure 3.14.

In summary, the ABC model views language as the learning and subsequent reproduction of temporal sequences. Speech is produced by connected voice motor actions which produce sounds in the environment which are later picked up by a listener. The sounds heard by a listener correspond to previously learned auditory attractors which are associated with the various conceptual components of the sequence.

Thinking is an extension of this motor control behaviour, but with the ‘sound’

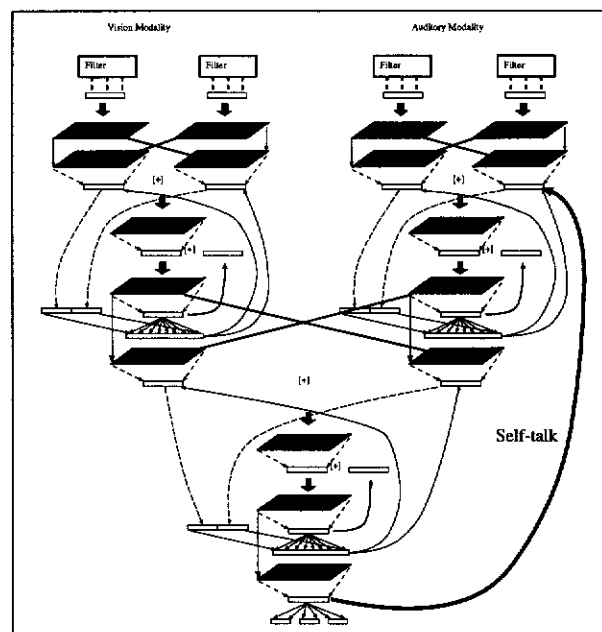


Figure 3.14: Self-talk Model.

vector signals travelling along an internalised recurrent loop rather than the sound being directed to the outside world for transmission.

The need to engage in private speech never disappears. Whenever we encounter unfamiliar or demanding activities in our lives, egocentric speech resurfaces. It is a tool that helps us overcome obstacles and acquire new skills.¹²

3.3.6 Motor Sequences

Piaget ([1946] 1962) posits universal cognitive capacities which are *not* specific to language, but which also structure cognition in other modes. The ABC model also ties together the processes of language, self-talk and skilled motor behaviours. In essence, there is no difference between the processes of language, conceptualisation and skill development. Learning to hit a tennis ball uses the same basic mechanisms as learning to speak a language.

Output motor behaviour is learned temporal sequences. The process is one of sequences of vectorial excitation of muscles and *not* inverse kinematics. Each learned output vector element contains an excitation strength for a particular muscle or muscle group. In combination with the other vector excitations, and following on from previous vector excitations, the muscles are then able to perform smooth coordinated actions.

In learning muscle coordination, the initial random muscle movements observed in infants gradually becomes directed and coordinated when visual, somatosensory, proprioceptive and other sensory inputs are temporally associated and self-organised. For example, the child will learn to focus on near objects and at the same time obtain a corresponding proprioceptive ‘reading’ for their arm when they touch something, such as a toy. The initial reaching may be at random or it may result from some inbuilt mechanism for exploration, but the effect will be to associate and learn coordinated proprioceptive and eye stereo/focus ‘settings’ for the reaching task.

Muscles may be driven by perceptual concepts (for example, playing a tennis shot or reaching for a toy), or by ‘voluntary’ control through a ‘linguistic’ path. The means of obtaining this ‘conscious’ control of muscles is essentially the same as the control of

muscles by perceptual concepts, only here the addition of a speech (and subsequently self-talk) component provides an additional context. The motor behaviour is initiated by language, either external or internal.

Luria (1961) proposed three stages in the ability of language to regulate behaviour (see also page 140, Bjorklund 1989):

- Between the ages of 18 months and 3 years, the language of others can initiate motor action in the child, but language does not function to inhibit an ongoing activity.
- From about 3 to $4\frac{1}{2}$ years of age, speech has an *impulsive* role in regulating behaviour. Children respond not to the conceptual message of a language command but to its physical characteristics—they respond more to the physical energy of the spoken word.
- From $4\frac{1}{2}$ to $5\frac{1}{2}$ years of age, children are able to use the meaning of speech to regulate their behaviour.

Luria (1961, page 92) suggests that “the regulatory function is steadily transferred from the impulsive side of speech to the analytic system of elective significative connections which are made by speech.”

The means of obtaining voluntary (conscious) control of a muscle is illustrated in Figure 3.15. † The connection from a learned perceptual attractor to the muscle is shown on the right of the figure. The muscle may be activated whenever the appropriate winning node associated with the perceptual attractor is selected.

Once the perceptual attractor has been associated with the (non-conscious) control of the muscle, and while (at some later time) it is actually activating the muscle, a second attractor resulting from an auditory or self-talk attractor (speaking or thinking of the

† We prefer the term *voluntary* rather than *conscious* as the latter word covers a number of associated phenomena as we discuss in Chapter 5. By this we mean that the person can use speech, or more usually, self-talk, to directly control muscle action, as opposed to the more automatic (non-linguistic or non-conscious) control of muscles that comes about through the execution of some motor skill.

label) is simultaneously active, then Hebbian learning can link the auditory attractor to control the muscle. The reason for this is that the post-synaptic node resulting from the non-conscious muscle control is firing simultaneously with the pre-synaptic node associated with the auditory or self-talk attractor. The post-synaptic node fires as a result of the muscle being activated, while the pre-synaptic node fires as a result of the auditory attractor firing a particular node on the SOM surface. As a result of Hebbian learning, the weight between the nodes (indicated as a solid black rectangle in the figure) is increased. Over time, if the weight increases to a sufficient strength, the auditory node will be sufficient to initiate the muscle action on its own, thus enabling the child to obtain voluntary control of the muscle.

Thus, although the initial non-conscious control was achieved through random associations, the voluntary control is as a result of the simultaneous use of the muscle and voice (or self-talk).

The connection between conscious and non-conscious control of muscles is a complex one. We discuss this aspect in some detail in Chapter 5. For example, a typist can ‘voluntarily’ move their fingers to strike the keys either under the control of language (self-talk)—a conscious action—or, in the case of an expert typist, more as a direct association between the sight of the words on the page and the required muscle control—a non-conscious behaviour.

This interchange between the linguistic control of behaviour for novices of a particular

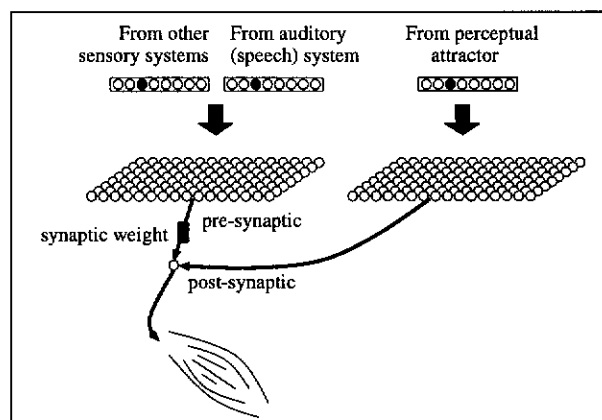


Figure 3.15: Voluntary Control of Muscles.

skill, and the direct perceptual attractor control of behaviours used by experts in that skill, is discussed in Chapter 5.

3.3.7 Parallel versus Serial

It is clear that the ABC model contains elements of massive parallelism yet retains seriality. The massive parallelism is achieved through the use of only local interactions between neurons within and between SOM surfaces. At each instant of time, *all* of the neurons in a single SOM layer act in concert, each taking part in local inhibition and updating of synaptic weights. If there are millions of neurons within the SOM layer, then each is not processed sequentially (as is required on a von Neuman computer, and hence performed in the *simulation* of the model reported in Chapter 2 and Appendix C), but rather they proceed in parallel. Likewise, the associative linkage between maps is done in parallel. †

The serial aspect of the model results from the fact that the input to all of the neurons of a SOM surface requires some time to pass to the output axons. There is a ‘minimum’ length of time that would be required for an input signal (say some sensory input) to pass through the system and induce some output behaviour. The use of recurrent loops in the architecture means that recognition times are not fixed, but may require some temporal sequence processing. The processing is serial because we appear to live in a world in which *time* is required for behaviours and actions.

There is some debate in computer circles about the pros and cons of parallel computer architectures—for example, SIMD (Single Instruction, Multiple Data) and MIMD (Multiple Instruction, Multiple Data) architectures. It is the author’s opinion that (symbolic) parallelism, as a computer architecture, has failed to live up to the early promises of researchers. As well, although most modern serial computers are capable of some parallel operations (such as separating memory fetches from operations of the main CPU), these devices are inherently serial, with a bottleneck occurring at the CPU. Parallel models may have multiple CPUs, but then the bottleneck is simply transferred to the means of coordinating the multiple data streams. We contend that all of these

†See Section 3.3.9 following for a more detailed explanation of what we mean by *parallel* processing.

devices are better thought of as *serial* devices, with the SIMD and MIMD architectures better described as *multiple-serial* architectures.

True parallelism requires only local update rules, such as is used in the ABC model. Millions, even billions of operations may occur, in a coordinated manner, in a single time step. When global coordination or decisions are required, massive parallelism is reduced to serial processing.

It should be noted that the whole ABC process occurs dynamically; that is, at each instant there is always information flow over the maps and around the recurrent loops. These processes cannot really be looked at in isolation, and the flow is continuous (albeit punctuated into separate vectors through some process of quantisation and synchronisation, such as the division of time into small increments and stepping through the process at each time interval).

3.3.8 Self-Organisation

Traditional Cognitive Science, and in particular Computer Vision research has assumed that particular *patterns* exist in the world, and that it is up to each creature to learn to recognise these patterns—the process and science of *pattern recognition*.

It is claimed that this is not the case, and that this scheme is much too ‘inflexible’ to be accepted, except in the case of certain innate reflexes. Pre-existing patterns do not exist in the world in the sense that new dangers and opportunities for a creature may come into being at any time, and rather than being able to respond to fixed external stimuli, a creature will be better served if it is able to learn the changing composition of its environment. The self-organisation of the external stimuli and the association of these inputs with appropriate behaviour would appear to be a more appropriate strategy for adaption and survival.

Self-organisation is appropriate for an adaptive, dynamic system—not pattern recognition. The fixed connections of localist FFNN solutions are also not appropriate.

Churchland, Ramachandran & Sejnowski (1994) remind us, in relation to child development, that “correlation-based models for self-organisation of primary visual cortex

during development have shown that some properties of the cortical cells, such as ocularity, orientation, and disparity, can emerge from simple Hebbian mechanisms for synaptic plasticity” (Swindale 1990, Linsker 1986, Miller & Stryker 1990, Berns, Dayan & Sejnowski 1993).

Separate self-organisation of the various modalities and sub-modalities detected by a creature will produce multiple topological maps, and as we show in Appendix A the neocortex contains numerous topological maps. Separate maps allows each sub-modality to be reorganised according to the statistics experienced by the creature in that sensory domain, irrespective of the other stimuli. The *important* statistics within each sensory domain will not be confused or diluted by inputs from other domains. †

In a sense, the actual ‘filter’ providing the input to each map is not important, provided it allows for a range of vectors to be self-organised. There may be no such thing as an ‘optimal filter’ in this sense.

In the ABC model, we make the assumption that each modality filter is fixed, but there is some evidence that even these filters are themselves self-organising. For example, various experiments performed on cats by Blakemore & Cooper (1970) indicate that the spatial frequency and spatial angle filters of higher animals undergo some form of self-organisation during development. Other anthropological data suggest that people raised in environments typical of western society, (in which there are many straight lines and right angles), develop different spatial frequency filters than others who are raised in an environment that has less linearity and a higher frequency of curved and irregular objects in their environment (Segall, Campbell & Herskovits 1966).¹³

Before Darwin, it was not seen how creatures could have the characteristics they do without some form of *design*. Darwin showed that the self-organising principles of evolution can provide the means of generating increasing diversity and adaption within organisms. Similarly, some philosophers and cognitive scientists think the certain components of the brain, such as the so called Universal Grammar (UG) unit of language,

†Separate maps also avoids the “metric” problem of computer vision—the issue of how to form a composite feature space from components with different metrics. For example, combining *colour* and *shape* axes.

must be innate as it is claimed that children receive inadequate language exposure for self-learning (for example, children are generally not exposed to negative examples of grammar, nor are they taught explicit rules of grammar—at least not initially). These researchers suggest that the UG component has somehow evolved and is now innate.

We claim that the likelihood of a language unit of such complexity evolving in the relatively short history of human language is very remote, and not sufficiently adaptive to be accepted. Further, there would appear to be insufficient ‘information’ in the DNA code to uniquely determine the neuron connectivity for such a complex language unit.

We contend that self-organisation of the brain can account for language and cognition without resorting to innate structures. The question revolves on whether language is obtained by humans in phylogenetic or ontogenetic time. We discuss this point further in relation to language in Section 5.1.

3.3.9 Hardware, Representations and Self-Organisation

One major advantage of a self-organising brain is that it does away with the requirement for representations, so that the system may be built solely in hardware.

We perhaps need to be careful of our use of the word *representation*, so some initial clarification is required. Of course, some aspect of the external world must be embodied in the brain for it to behave in accordance with external objects and events. However, it does not need to incorporate representations in the sense of *standing-in-place-of*.

Self-organisational *changes* to the structure of the brain are more akin to modifying a resistance or capacitor on an electrical circuit than to storing a new ‘representation’ in a computer. The change is made to the basic structure of the brain, and is not simply an additional piece of data to be manipulated at a later stage. The change is *actual* and *permanent*.[†]

A trivial example will illustrate the nature and advantages of a hardware-only solution.

[†]Obviously the overall process is *dynamic*, so the permanence is only fleeting, but the point still holds. At a neural level, the synaptic connection is strengthened via some chemical means and any further changes to that particular synapse in the future will be from that weight value.

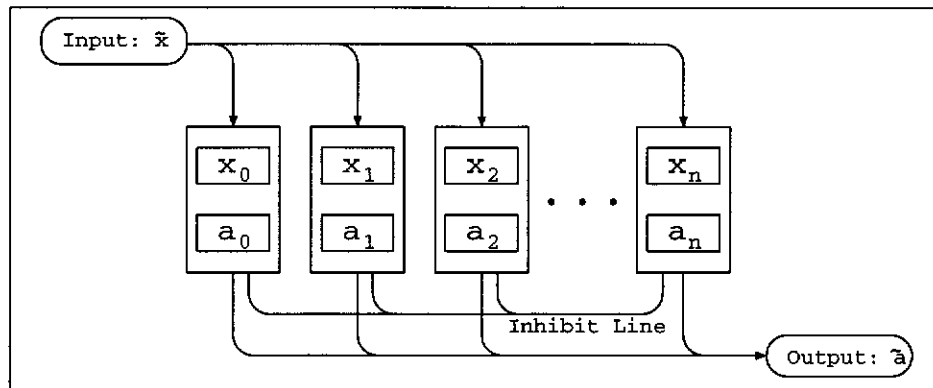


Figure 3.16: Hardware Kohonen Map Model.

Let us develop a *time-delay Kohonen map*—a structure in which a set of units operate in parallel to determine the closest match between an input activation vector and a set of prototype vectors. Such a structure is shown in Figure 3.16. [†]

The input vector, \tilde{x} , is distributed to each of n units. In parallel, these units compute a distance (an error) between their prototype vector x_i (initially set at random) and the input vector \tilde{x} of the generalised form $d_i(\tilde{x}) = f(\|\tilde{x} - x_i\|)$. Each unit then, *independent of the others*, waits a time proportional to the distance measure d_i . When a unit's time is up, it activates the inhibit line which switches every other unit off—a form of lateral inhibition. At the same time, it places its address, a_i onto the output line. If a unit detects that the inhibit line has been activated while it is waiting, it switches off.

The net result of this process is that the address corresponding to the closest matched unit will be the output of the system. This architecture is designed for nearest neighbour pattern matching and can be used to build a content addressable memory—the output of the unit may be used to directly locate 'behaviour' associated with the winning node.

Following the output for each cycle, the appropriate updating of the winning and near-neighbour nodes may be undertaken. Circuitry for this operation is not shown.

The first point to make about this network is that the 'search time' (or 'lookup time') to find the winning node is approximately constant ($O(1)$) no matter what the size of the SOM array. The search time is actually $O(\|\tilde{x} - x_a\|)$ which will tend to a constant

[†]Kohonen (1993) also discusses the implementation of "Winner Take All" biological circuits.

value as the map self-organises to match the incoming vectors.

Of course, the brain utilises *lateral inhibitory* connections to force a winning node, rather than a time difference as in our example. However, the principle of table lookup and constant search time still holds.

We can think of the behaviour of the ABC model in terms of computer metaphors. Whereas the cognitivist model uses a process of *search*, the ABC model uses a process of direct *table lookup*.[†] These two alternatives are illustrated in Figure 3.17.

The self-organisation process will, over time, organise each SOM surface into a dynamic ‘lookup table’ which directly links inputs with appropriate outputs. The ABC model suggests that the brain is a series of recurrently linked self-organising maps, each of which is similar to a huge lookup table (i.e., ‘get-this-input’ means ‘do-that-behaviour’). Software search over representations is an absurd waste of time, with the same ‘calculations’ needing to be performed every time a similar situation is encountered. It is much more efficient to self-organise into a direct table look-up.

This practice is used in part in some computer software design—any algorithm may be recast as a look-up table if sufficient pre-processing is done on the ‘representations’

[†]As mentioned in O’Rourke (1993), the view that intelligent behaviour could be achieved with an extremely large (but finite) static lookup table goes back to at least the work of (Leibnitz [1702] 1969). Note that the ABC model suggests that the brain acts rather like a *dynamic* lookup table, and not a static one.

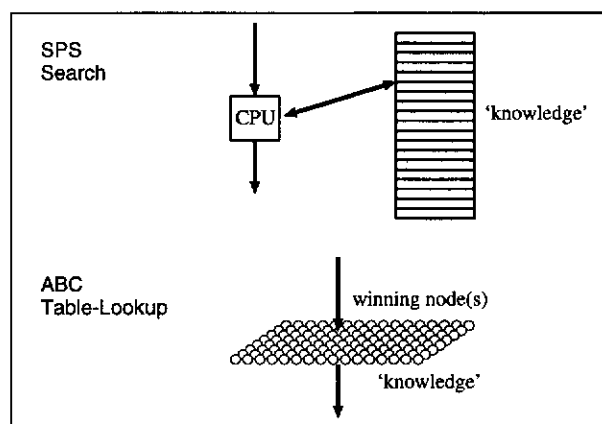


Figure 3.17: SPS/Search versus ABC /Table-lookup

at compile time. In effect, calculations at ‘run time’ are exchanged for calculations at ‘compile time’. A side effect of this is that more (computer) memory is usually required for the look-up method. An example of this technique is the ‘Graph Matching Algorithm’ of Bunke & Messmer (1996).

Of course, biological systems don’t have compile and run-time components—it is all ‘run time’—but all the same, the statistics collected during the life of the creature allow it to directly ‘choose’ an appropriate action in the most efficient manner.

Note also that the ABC model does not employ the form of *static* look-up table that is familiar to computer programmers, but rather a *dynamic* look-up table that is constantly self-organising according to the occurrence of external and internal conditions.

Traditional methods of pattern recognition require extensive searching of a (usually large) feature space, which is composed of n -dimensions—one for each feature. Attempts to cluster within this space are difficult because of the sparseness of this linear space, resulting in poor predictive success. Further, the linearity of the space means that well sampled volumes in the space are assigned on an equal unit hyper-volume basis with sparsely sampled areas. The feature space must be maintained on a metric basis, i.e., with equal divisions along each dimension.

These issues do not arise with (Kohonen) SOMs. Here the non-linear nature of the mapping from an n -dimensional vector to a 2D space ensures that *more* neurons are allotted to frequently occurring input values than are allotted to infrequently occurring input values. The space is ‘warped’ to conform to the input statistics. The ‘feature space’ is not maintained on a metric basis and the SOM self-organises to suit the statistics of the incoming data in a non-linear manner. Further, sparse clustering is not an issue.

Another ‘computational’ advantage of the self-organising hardware model is that large incoming vectors are not an issue. In fact, the larger the number of elements in the incoming vector, the better the subsequent discrimination of the winning node. In contrast, traditional pattern recognition algorithms need to limit the number of input features for computational tractability.

In summary, we must emphasise this point—the evidence is that the functional brain of humans and higher animals self-organises over time so that it may respond almost *immediately* to a situation. It does not have to *search* for the appropriate behaviour, but is able to essentially *look up* the required response. This model would seem to be an emphatically better proposition for the survival of a creature, provided it is able to survive the rigours of the learning (development) period. Such characteristics are reflected in the ABC model.

3.3.10 Categorical Perception

Categorical perception is the simultaneous perception of enhanced *within-category* similarities and *between-category* differences. Human and animal perception has the ability to carve the world into relatively orderly experiences, segmented into ‘chunks’, or segments. This ability is aided by the phenomenon of *categorical perception* (CP) (Harnad, Hanson & Lubin 1991), in which equal-sized physical differences in the physical signals arriving at our sensory receptors are perceived as smaller *within* categories and larger *between* categories. For example, differences in wavelength within the range we call “yellow” are perceived as smaller than equal-sized differences that straddle the boundary between yellow and the range we call “green”. The wavelength continuum is somehow ‘warped’, with some regions being compressed while other regions are stretched out.

Bornstein, Kessen & Weiskopf (1976) describes colour vision and hue categorisation in young human infants (Burnham 1987, page 239):

... both adults and infants categorically perceive the physical continuum of visible wavelengths as separate hues with quite sharp category boundaries.

It has been claimed that colour vision CP, as well as other examples of CP in human and animal perception, is a fundamental property of our sensory systems, modified only minimally (if at all) by experience. Musical pitch categorisation may be an example of CP effects that result primarily from learning. However, CP effects have also been shown to occur purely as a result of learning in experiments with artificial continua.

The generation of CP has yet to be adequately explained, but Harnad et al. (1991) show how CP might arise as a natural side-effect of the means by which certain standard neural net models accomplish learning. Harnad et al. discuss the imposition of categorical perception for neural nets by “warping” inter-stimulus similarity space in a way that resembles human categorical perception.

The mechanism for CP suggested by the ABC model is slightly different. In attaching a label to a perceptual attractor on a SOM surface, such as the association of sound labels (‘red’, ‘yellow’, and so on) to areas on the visual system SOM associated with colour (or subsequent maps), the process of self organisation of the sensory space would result in a deformation of space within the attractor. The self-organisation process will result in a *prototypical* node which will correspond to the colour learned as being statistically the most reinforced example of that named colour. Other similar colours near this prototypical exemplar will also be ‘known’ by the same label—that is, they will produce the same speech behaviour when naming of the colour is requested.

However, the space within the attractor is not linear as is indicated in Figure 3.18 (b). This figure suggests a profile of winning wavelengths associated with each node on (a 1D section of) a SOM. The prototype colour is near to the centre of the attractor, with similar colours maximally exciting nodes in the near neighbourhood. An increment in wavelength which falls across a label boundary will appear to the subject to be larger, due to the non-linear warping, than an equal increment experienced within a category boundary.

3.3.10.1 An Example—Categorical Perception in Speech

An important example of categorical perception is that found in the processes of speech and listening. Burnham discusses CP and speech in both adults and infants, and defines CP as (Burnham 1987, page 239):

Categorical perception is the tendency to perceive a physical continuum discontinuously i.e., for the relevant perceptual system to group stimuli on a continuum such that discrete categories are perceived.

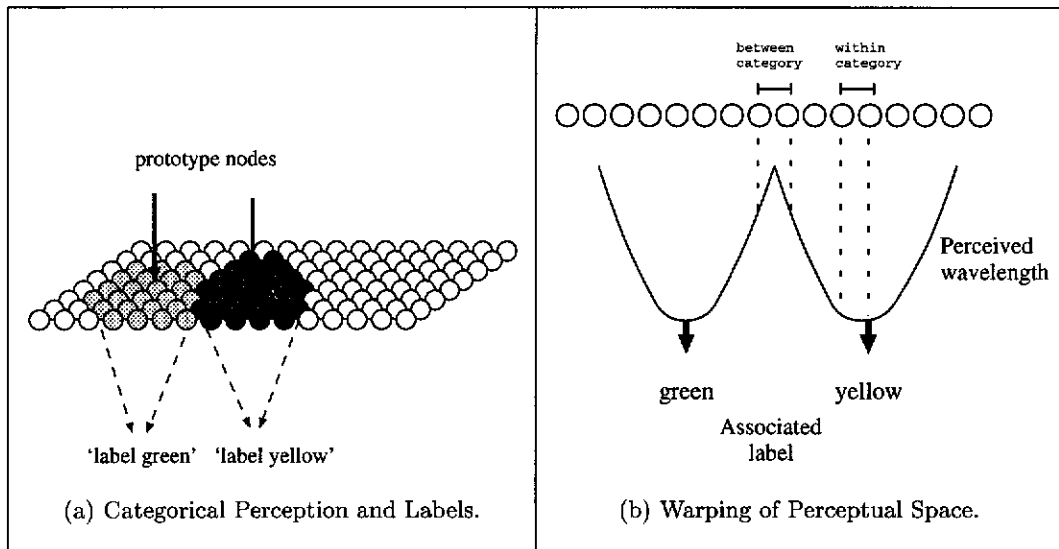


Figure 3.18: Categorical Perception.

Burnham suggests that we must question the claim that humans innately perceive and process speech in a special manner. It is now clear that linguistic experience is involved to some extent. Previous workers had found that infants as young as 1 month old were able to discriminate speech categorically, which strongly suggested that the special manner in which humans process speech was innate. However, the results cited by Burnham now suggest that the categorical discrimination of speech in infancy can no longer be taken to imply a special innate mode of processing for speech by humans.

The evidence suggests that the perception of speech in infancy is *not* truly categorical but becomes increasingly so with experience, especially with the active experience in language acquisition between 2 and 6 years. Linguistic experience modifies categorical speech perception.

Burnham also mentions several references which indicate that categorical perception is not peculiar to speech or peculiar to humans—animals with similar auditory systems to our own also categorise speech sounds.

Kuhl (1978) further suggests that continua with these properties were selected for in the evolution of language systems; i.e., that speech is not perceived categorically because it is speech, but that speech *is* speech because it is perceived categorically.

Eimas (1975) demonstrated that in an English language environment, infants of 2 to 3 months of age can discriminate the liquids /r/ and /l/, and suggested that this ability may be present at birth. Adult English speakers are also capable of making this discrimination. However, Miyawaki, Strange, Verbrugge, Liberman, Jenkins & Fujimura (1975) have demonstrated that Japanese adults, for whom the /r/ versus /l/ contrast is not phonologically relevant, can not discriminate the two sounds. Burnham (1987) also cites other examples of cross-cultural differences in phonetic recognition.

The evidence shows that discrimination can improve with training. For example, six month old infants given extensive training on the /f/ versus /θ/ contrast can actually discriminate the difference. Thus, linguistic experience can facilitate the categorical perception of speech by sharpening categorical boundaries. There are also different boundaries for different languages (e.g., Spanish vs. English).

There is evidence for differential *realignment* of category boundaries over cultures, and for selective *sharpening* of phonemic category boundaries. There is even evidence that the perception of fricative contrasts is *enhanced* or even *induced* by linguistic experience, and that some stop consonants may require specific linguistic experience for their *maintenance*.

Burnham (1987, page 271) suggests that “consideration of theories about the evolution of language systems and ontogenetic development converged on two views: first, that speech perception begins in infancy as a psychoacoustic phenomenon and ends in adulthood as a phonological phenomenon; and second, that different developmental processes may be involved for contrasts which differ on the degree to which they are universal and psychoacoustically based.”

The evidence suggests that an active experience with language directs the language learner’s attention only to those contrasts which are meaningful, and strengthens the view that infants initially identify speech contrasts but do so non-categorically and in a psychoacoustic mode. When language acquisition directs attention to phonemic contrasts, identification of those, and only those contrasts gradually becomes more categorical such that by adulthood speech is perceived in a phonological (and perhaps automatic) mode. Nevertheless, other evidence suggests that pure all-or-none categor-

ical perception is probably a myth even for adults.

The ABC mechanism of self-organisation on SOM surfaces leading to differential outputs suggests exactly these results. Because certain sounds in a particular language will be repeated in conjunction with certain referents, and different referents which require contrastive opposition will lead to differential output behaviour, those sounds will be separated into different attractors.

In contrast, should two similar sounds in a language not lead to separable outputs (for example, /r/ and /l/ in Japanese), then they will be self-organised into the same attractor and hence not be able to be differentiated by an experienced user of that language.

Suitable training, however, will establish the differential outputs thus leading to separability. More work is required on this issue to establish a more rigorous treatment. However, this is left to future research.

3.3.11 “Soft-Wired” Reflex

It is generally held by animal learning researchers that “each animal species possesses a set of specific learning dispositions and is thus species-specifically prepared for combining hard-wired with modifiable neural circuits” (Menzel et al. 1984, page 260).

The linkage between sensory inputs and behavioural outputs in the ABC model is essentially a “soft-wired” reflex (or an “adaptive” reflex).

Reflexive behaviour is essentially hard-wired from birth, and is not altered over the lifetime of the animal. Reflexes are generally involuntary, and extremely rapid as a result of a simple, straight-through, sensory-to-motor connection. A ‘conscious’ decision process † would be simply too slow—perhaps as long as 500 msec (Graham 1990, page 139).

Note that the soft-reflex may require several recurrent ‘loops’ across SOM surfaces

†We discuss ‘consciousness’ in Chapter 5. For the moment, a ‘conscious process’ in the ABC model means that some self-talk (thinking) component is included in the process. That is, the person is aware of the process as it unfolds by thinking of it. This may require multiple ‘loops’ within the recurrent temporal connections across SOM surfaces, and so is a relatively ‘slow’ process.

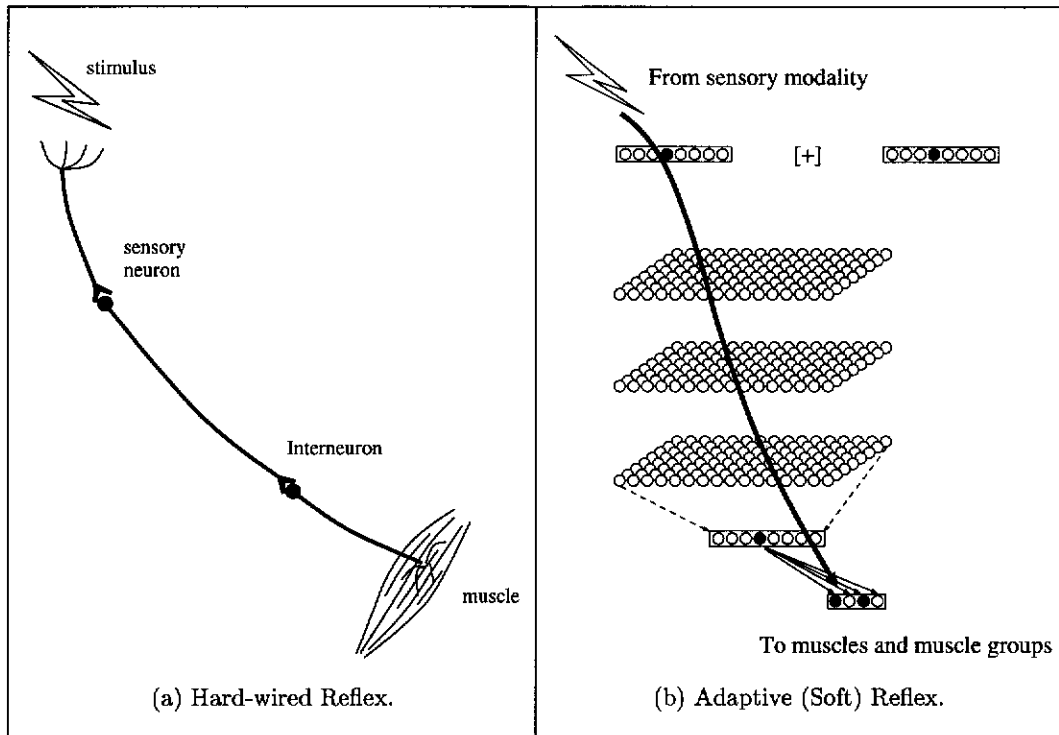


Figure 3.19: “Hard” Reflex versus “Soft” Reflex.

before the appropriate output is achieved. For example, in looking at a face in a crowd, it may take a number of saccades (and hence, as we see in Chapter 4, a number of recurrent loops following a scanpath) before a learned temporal sequence leading to a recognition behaviour is achieved.

If, however, the input has been sampled many times and thus is very ‘familiar’, the self-organised attractors may be sufficient to achieve recognition directly, as shown in Figure 3.19 (b). The mechanism thus allows for extremely fast recognition and action for familiar sensory inputs. Further work is required, but the ABC model appears to provide a mechanism to enable recognition within the approximately 100 to 150 synaptic joins of human recognition (Feldman & Ballard 1982), and possibly the 20 msec view time for recognition of facial expressions (Simpson & Crandall 1972).

The model gives us a continuity between hard-wired reflexes, and the soft-reflexes brought about by self-organisation on SOM surfaces and recurrent loops. There is another extension, that of conscious control of behaviour, and we discuss this aspect in Chapter 5.

3.3.12 Learning

In the ABC model, the 'brain' is constantly self-organising and hence constantly learning. Every experience at every moment causes some small change in some of the synaptic weights and so the process of learning is dynamic.

Learning in the ABC model is a process of change of the whole creature. The creature is 'wired' differently following learning simply as a process of dynamically adapting to its experiences, both from internal sources and with respect to the external world.

Learning is not seen as the addition of a new rule, or a new piece of knowledge, but rather an adaption to incrementally modify behaviour. The synaptic weights within the brain are constantly changing so that each individual is the current sum of their past experiences. Each has a world-view which changes over time—that is, children see the world in a very different light from an adult.

Learning is obviously important to any system. It is a general principle of nature to endow (at least higher) living creatures with an ability to learn so that they are better able to adapt to new situations. If the world was constant, learning would be less important; but it is a dynamic, changing world, and so living things need to respond quickly and appropriately to new situations.

The traditional view has it that learning is adding new knowledge about the world, generally represented in a linguistic form, to a database of knowledge stored somewhere in the brain. Important concepts are *knowledge*, *representation*, *background knowledge*, *search*, *optimisation* and some form of *learning principle*, such as Bayesian probability theory.

The principle method of learning used within the cognitivist paradigm is induction—the forming of some conclusion which is more general (i.e., applies to a wider range of instances) than the premises. Within the discipline of *machine learning*, for example, this usually consists in finding (segmenting) some region of a feature space which encloses example instances. The instances are points within the hyper-volume region, and the region is taken to be the learned 'rule' for the concept. Future (as yet unseen) instances of the concept (or class) are expected to also fall within the region and hence

satisfy the rule. Induction is thus a process of generalisation, usually over linguistic terms (Briscoe & Caelli 1996).[†]

The ABC model questions induction as being the primary method of learning. As with other connectionist models, learning within the ABC model does involve a component of generalisation through self-organisation on SOM surfaces and association links. However, we contend that another major component of learning is a process of *specialisation* resulting from differential behaviours. That is, rather than build up from features and then try to generalise, the ABC model bifurcates neural attractors to differentiate phenomena within the world.

The process of induction (and deduction), we contend, is a learned behaviour and not a fundamental principle of learning. Learning is a much more dynamic and low level process. This is not to deny that some form of induction is used by humans, but that generalisation over linguistic terms is a socially learned process.

In the ABC model, only Hebbian update of the synaptic strength between concurrently firing neurons is used for learning. However, the update is *value-based*, and so is able to account for a number of learning observations. Hebbian learning allows for SOM generalisation, associative links between maps and the learning of temporal sequences.

The bifurcation of attractors, based on behavioural and linguistic differentiation, enables new concepts to be formed. Bifurcation further discriminates the world, allowing more refined actions to take place, and occurs when there is a behavioural (including cultural) reason for it to occur.

A metaphor for bifurcation is the formation of a bubble in a 3D ‘concept’ state-space. An attractor is formed so that any perceptions falling within the bubble are associated with a certain behaviour. New concepts are formed when the bubble is split into smaller bubbles, which more finely carve up the incoming perceptions—the more astute concepts of an expert compared to those of a novice. The metaphor applies to the tennis player who practices so that he or she has small attractor bubbles relating the current speed and direction of the ball with his behavioural output of placing his racquet in its path at a certain angle, with a certain speed, and so on. It applies equally well to, say,

[†]The problems in finding a suitable generalisation are discussed in Section 3.4.

differentiating breeds of dogs based on sensory inputs and linguistic labels.

The value-based component of Hebbian weight update in the ABC model allows for an explanation of the more rapid learning observed during the development period. If the learning *rate* has a high value during childhood, with a subsequent decrease following a developmental period, then learning will tend to be more rapid during the development period. The value-based component also allows for rapid learning in ‘emotion’ charged circumstances by providing a higher learning rate. Value-based learning also allows for *single-example* learning—if the emotional content is sufficiently high then a larger learning rate will allow rapid learning.

The cognitivist paradigm requires learning to be ‘conscious’ (Irwin 1993b, page 367). We discuss this aspect of learning in Chapter 5 when we look at language and consciousness. There we also look at the distinction between explicit (conscious) and implicit (non-conscious) learning.

Language usage greatly speeds up the process of learning by allowing perceptual experiences to be labelled and then subsequently referenced in other situations. It also allows for associations to be formed between objects based upon their linguistic labels rather than on sensory inputs—for example, the many varieties of ‘chair’ often used to explore human conceptualisation in certain texts may have little more in common than their label.

The labels themselves, separately from their real world referents, may take part in learning as separable abstract entities in their own right. They will always, however, have some ‘meaning’ attachment or linkage to perceptual attractors and other labels. In other words, there will always need to be some *grounding* of the terms.

Another aspect of the ABC model is that learning, perception and behaviour may not be treated as separate components of cognition. The model combines ‘recognition’ (or observation, or perceiving) and learning as coincident.

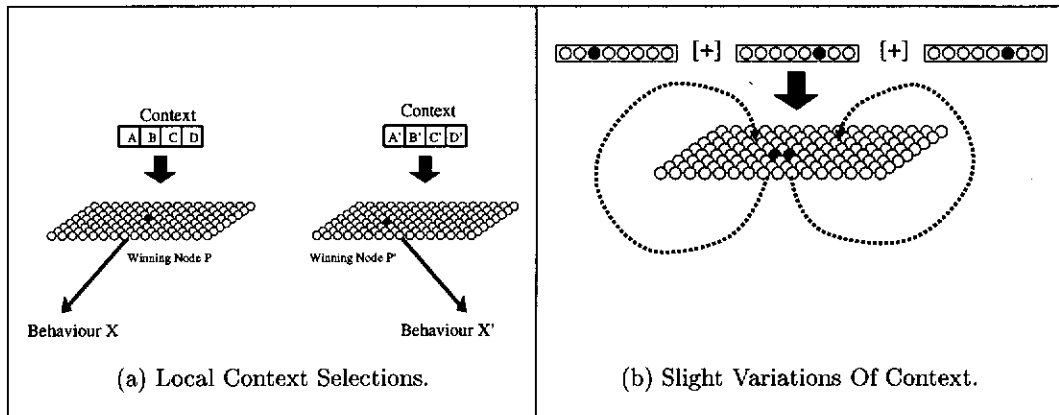


Figure 3.20: Different Context Selects Different Behaviour.

3.3.13 Context

Context is very important component of cognition—the same sensory inputs in slightly different contexts may require very different behaviours.

As was shown in Chapter 2, the process of concatenating vectors enables context to be a major component of the ABC model. Vector concatenation enables information from different parts of body (emanating from the either external sensory sources or other internal sources) to be combined together to form an extended context for subsequent processing.

For example, imagine a vector component from the limbic system being concatenated with other sensory information. If the current situation experienced by the person has a low ‘emotional’ context, then they will tend to go about their ‘usual’ behaviour. But if that same sensory context is punctuated by a scream of *fire*, that auditory input will induce in the limbic system a strong emotional component which will subsequently modify the context, which will in turn produce very different behaviour.

The different (contextual) vector will cause a different path to be taken through the network (and thence around the recurrent loops) to enable different behaviour to be undertaken. This is illustrated in Figure 3.20 (a). Context $ABCD$ will select winning node P leading to behaviour X , while context $A'B'C'D'$ will select winning node P' leading to behaviour X' .

An appropriate context enables a direct 'selection' of the most appropriate behaviour given past experience. Moreover, different local context winning nodes will lead to divergent temporal sequences being traced through the recurrent linkages. This will mean that the different contexts may lead to different subsequent behavioural sequences as illustrated in Figure 3.20 (b). In this figure the recurrent paths are shown for one layer, but the full effect occurs over many layers.

3.3.14 Common Sense

The account of common sense within the ABC model follows on from the discussion of context. Decisions are made because of the prior learning and self-organisation of the person, so that they will choose the most appropriate behaviour given the current context. The person does *not* need to consider any of the multitude of associated information that we know of as *common sense*.

If, however, they are later asked to recall their reasoning for making a 'decision', the linkages between the selected behaviour and other 'knowledge' may be followed to provide some form of rationalisation. The person will be able to reproduce associated and linked learned 'facts' about the context, the situation and the resulting behaviour. However, consideration of the background *common knowledge* is not necessarily part of the decision itself.

This view is supported by studies of neurological disorders. For example, Bisiach and his coworkers (Bisiach, Berti & Vallar 1985, Bisiach 1988) report the case of a patient with a left hemiplegia and a left hemianopia and who was severely anosognostic to both.

Hemianopia is loss of vision in which the patient is unable to see in all or part of a visual field—in this case the left visual field. Hemiplegia is the loss of voluntary movements on one side of the body. Anosognosia is the loss of ability to recognise or to acknowledge an illness or bodily defect. Thus the patient was not only unable to see in his left visual field, or to move the left side of his body, but was not consciously aware that he was unable to do so. When asked to move his left hand and arm, the patient believed he was

actually doing so, and when asked to explain the anomalies in his behaviour resorted to strange rationalisations. For example, when the examiner placed the patient's left hand between his own hands, and asked: "Whose hands are these?"

Patient: Your hands.
Examiner: How many of them?
Patient: Three.
Examiner: Ever seen a man with *three* hands?
Patient: A hand is the extremity of an arm. Since you have three arms it follows that you must have three hands.

Much of the patient's 'common knowledge' relating his concept of 'leftness' and his conscious control of his body is missing, but the rest remains, and so he follows other links in his common sense associations and linkages in order to satisfy the examiner's questions. His explanations border on the bizarre as he searches for a 'rational' reason for the anomalies within the domain of his previous behaviours and knowledge.

Similar common-sense rationalisations are observed in experiments with split-brain patients in which different images are given to the left and right visual fields, or an incongruity is forced between visual and somatosensory inputs (Sperry & Gazzaniga 1967, Gazzaniga 1992).

3.3.15 Errors and Unsupervised Learning

ABC does not require a process of *error minimisation* such as that used in the back-propagation algorithm (Rumelhart et al. 1986) and models of adaptive control (Åström 1995). If a specific function is to be learned, most methods use some form of error minimisation technique such as error gradient descent learning.

The ABC model, on the other hand, does not comprise any component which uses supervised learning. It is the consistency of the external world, and the training from others in society, that forces consistent behaviours in individuals.

Training within society may continue until an individual conforms to some *expected* behaviour, or the individual may persist until they have mastered a skill, but this is

achieved internally within the ABC model solely with Hebbian learning.

3.3.16 Representations

The ABC model is non-representational in the sense that although the world is obviously ‘represented’ in the various weights in some way, these are spread over multiple SOM surfaces and connecting association weights, and are accessed via temporal sequences. For example, an external ‘symbol’ such as cat may be made up of linked weights in the visual system which refer to the perceptual inputs in viewing a cat, linked weights in the auditory system which refer to perceptual inputs in hearing a cat “meow”, weights in the somatosensory system associated with stroking and touching a cat, olfactory weights associated with smelling a cat, as well as culturally obtained visual and auditory weights associated with seeing the characters ‘c’, ‘a’, ‘t’ (i.e., the written symbol ‘cat’) and hearing the spoken word ‘kat’. All of these will further be linked with weights from relevant experiences with cats, such as they scratch, catch mice, and so on. There is no single location containing the concept ‘cat’. There is no representation in the sense of *standing-for*.

Further, some of the weights found in the ‘representation’ of cat may also take part in the ‘representation’ of other concepts, such as ‘dog’.

The linguistic representation ‘cat’ is not a primary component of the cognition process, but is rather a *product* of it. Representations in the sense used in the cognitivist model are externally generated (i.e., outside the brain), culturally-determined phenomena. Within the ABC model, on the other hand, concepts are formed by self-organisation and temporal learning within the cortex, and exhibit themselves through behaviours which reproduce the external, cultural representations; for example, speaking or writing the word ‘cat’.

A qualification to the external status of linguistic representations such as ‘cat’ arises with self-talk. With self-talk, the normally *spoken* word behaviours are executed internally within the cortex (as we discuss further in Chapter 5), and become part of the ‘datum’ of cognition. Since we appear to ‘think’ in terms of linguistic representations,

many have taken this to mean that linguistic representations are much more *fundamental* to the process of cognition, and in fact form its mechanism. This is the view held by the cognitivist model.

The ABC model relegates linguistic representations to their *rightful* place in cognition; not as the fundamental building blocks of cognition, but as culturally determined and learned external adjuncts to the real processes of cognition such as self-organisation, associative linkages, and the learning and reproduction of temporal sequences leading to perceptual categorisation.

The rejection of representations is supported by a number of other authors. Slezak (1993, page 92) makes the point that “the symbolic approach conflates ‘first-person’ representations in our environment (e.g., utterances and drawings) with ‘third-person’ representations (e.g., mappings a neurobiologist finds between sensory surfaces and neural structures)”.

Clancey (1993, Page 87) also provides a view, based on neuropsychological research, which “rejects the hypothesis that neurological structures and processes are similar in kind to the symbols we create and use in our everyday lives”. According to Clancey, the representation storehouse view of memory confuses structures in the brain with physical forms that are created and used in speaking, drawing, writing, and so on. This reduces learning to syntactic modification of the modeller/teacher’s pre-supplied ontology of standard notations, and representing, comprehending, etc., are reified into acts of manipulating representations. The stored-schema models view *meaning* as mapping between given information and stored conceptual primitives, facts and rules; *activity* as executing rules or scripts; and *concepts* as stored descriptions. The ABC model rejects these views.

3.3.17 Continuity Across Phylogenetic Tree

Any model of cognition should be able to base the evolutionary development of phylogenetically later animals on the general brain structures found in recent predecessors; for example, the evolution from, say, reptile to monkey, and on to human. The model

must also be able to accommodate the multitude of sensory mechanisms found in nature, such as light, sound (including echo location), magnetism, electricity, chemicals (smell and taste), gravity (proprioception), pressure and heat (somatosensory), radio frequency (insect antennae), infrared radiation, and so on.

The brain of higher animals is a multi-layered organism with layers corresponding to previous evolutionary developments still in place. Any model must also be able to incorporate each component into an integrated whole.

The ABC model allows for arbitrary ‘world-filters’ to collect a vector to be used in subsequent processing. The vector basis of the model also accounts for phylogenetic continuity via the concatenation of vectors, the older and newer regions then both providing a contextual vector component which is able to be associated and learned in temporal sequences.

3.3.18 Vector Concatenation

The process of vector concatenation to obtain an appropriate context that may in turn be mapped to a SOM surface, is an extremely flexible and powerful mechanism. It allows multiple, interconnected components to influence behaviour either independently or in coordination. For example, appropriate vectors may be taken from certain areas of the brain, concatenated into one composite context, and then, following mapping to a SOM, be used to control an isolated component of cognition—for example, the control of eye movements. The same vector components may be joined in different combinations for other tasks.

The concatenation mechanism also suggests a solution to the problem of combining the phylogenetically disparate parts of the brain. If the older components (for example, the limbic system) provide a vector of their outputs to the newer components (for example, the neocortex), then the newer component can utilise the results of the older component within its processing. The older component vector will be used as another *context* vector to be used in learning behaviours.

This further suggests a mechanism for the interplay of the older phylogenetic areas and

the newer areas during development, and in moments of stress, when more primitive reflex actions take over from the 'reasoned', conscious behaviour of normal circumstances. The ABC model suggests that initially the neocortex is not organised into any form of perceptual conceptualisation, and the weights on the SOM surfaces, association and recurrent links are more or less random. As such, it is unable to provide any coordinated connection between sensory inputs and output motor behaviours. The infant's behaviour is under the control of the older, more reflexive structures of their brain. As the neocortex becomes more organised in connecting sensory inputs and internal perceptual concepts with motor behaviours, as occurs during development, the vector context provided by the cortex makes a more significant and coordinated contribution, in effect taking over from the older components. In this way, the adaptive (and mostly 'conscious') mechanisms of the cortex are able to override the more basic reflex mechanisms. In times of stress and danger, however, an enhanced contribution from the more primitive areas to the concatenated vector will allow the reflexive behaviour mechanisms to resume control.

3.3.19 Topographical Connectionist Approach

The ABC model incorporates an alternate connectionist approach to FFNNs—the formation of topographical maps. The Kohonen Self-Organising Map (SOM) mechanism is used as a representative model in simulations (Kohonen 1984, Kohonen 1990, Chappell & Taylor 1993, Lo, Yu & Bavarian 1993, Kangas et al. 1989, Kangas 1989, Kohonen 1989, Kohonen 1993). While the Kohonen SOM is not necessarily the correct model, it is used because of its computational tractability in the simulations, and because many of its properties are well known.

The SOM method is an unsupervised mechanism for mapping (or organising) the inputs to the map to ensure that similar input values are located close together. As learning with the SOM approach is unsupervised, (and unlabelled), the problem of a priori knowledge is eliminated.

Topologically organised maps are ubiquitous in the brain, and the SOM method has strong biological feasibility.

3.4 Addresses Problems with Alternate Approaches

In this section we briefly discuss some of the problems faced by alternate cognitive models, and the way in which the ABC model overcomes or avoids these issues. We discuss a number of these and other associated issues more fully in Chapter 5.

Real-Time Constraints

While the signal velocities in nervous systems are quite slow in comparison to those in computers, brains are nonetheless much faster than electronic devices in the execution of some complex tasks. For example, human brains are incomparably faster than any existing computer in visual recognition tasks.

Feldman & Ballard (1982, page 206) express the modelling constraints implied by temporal limitations of recognition and action as the *hundred-step rule*:

The critical resource that is most obvious is time. Neurons whose basic computational speed is a few milliseconds must be made to account for complex behaviours which are carried out in a few hundred milliseconds (Posner 1978). This means that *entire complex behaviours are carried out in less than a hundred time steps*. Current AI and cognitive science programs require millions of steps.

The ABC architecture, because of its inherent massive parallelism, is a hardware solution rather than a software-based system. As every neuron on each SOM surface is processed at once (as are the linking association and recurrent connections), the ABC model *is* able to process sensory inputs and produce appropriate behaviour within realistic times. As stated previously, the appropriate metaphor for the model is a *table-lookup* in which the appropriate behaviour may be selected either directly or within a few recurrent loops, rather than the *search* metaphor of the cognitivist model. As such, real-time constraints are not an issue.

Content-Addressable Memory

There are a number of *dissimilarities* between computers and nervous systems. We look at some of these in Chapter 5, but one major difference pointed out by Churchland (1986) is that of storage. Conventional von Neuman computers use memory much like

a library in which each piece of data is located in a particular place in the memory bank. For the central processor to retrieve that data, it must know the address in the memory bank for each datum.

However, human memory appears to be organised along entirely different lines. Human memory is able to ‘reconstruct’ particular memories from a partial or a degraded stimulus. As well, there are associative relationships among stored pieces of information based on considerations of *content* rather than considerations of *location*.

These human capabilities are shared by the ABC model. Generalisation on the SOM layers allow degraded vectors to select a neighbouring winning node, and if this is within the same differential output attractor, will enable the same output as the pristine vector to be selected. Moreover, the SOM surfaces are associative memories, with the contents of the vector selecting the winning node which then leads to some appropriate learned behaviour.

Symbol Grounding Problem

The cognitivist model of cognition stipulates that the processes of cognition may be dis-associated from the external world, and the symbols manipulated separately. The *symbol grounding problem* (Harnad 1990, Harnad 1992) considers the difficulty, if not impossibility, of relating symbols back to meaningful, physical objects in the world using this model.

Harnad has argued that the woes of the cognitivist model relate to the *ungrounded* nature of the atomic symbols. He points out that the initial set of symbols, being arbitrary, is flexible enough to support compositionality (atomic symbols can be combined and molecular representations can be decomposed according to a formal syntax) and systematicity (the atomic and the molecular symbols and the rules of syntax can be systematically assigned a meaning). He asserts however, that the very fact that these symbols are arbitrary, means that they become *ungrounded* (their meaning and interpretation reside in the subject’s ‘head’ and are otherwise ‘rootless’) and give credence to the Chinese room arguments of Searle (1980).

Formal symbol systems, such as a computer program, manipulate symbols on the basis of formal rules or algorithms that apply to the *shapes* of the symbols rather than their meanings; that is, symbol manipulation is syntactic rather than semantic. The

meanings of the symbols are projected onto them by a user/programmer who interprets the symbols and the symbol manipulations; they are not intrinsic to the system itself.

However, humans (using sensory mechanisms), can discriminate and categorise real-world objects and events to which their symbols can be interpreted as referring. The symbols are *grounded*, rather than just being parasitic on the meanings an interpreter projects onto them. Whatever mechanism successfully maps the sensory projections onto the category names is also what grounds the symbol system.

Harnad's solution calls for the grounding of the categorisations in motor and sensory perception. In other words the atomic symbols are grounded in sensorimotor perception and are not arbitrary.

This is exactly the method employed within the ABC model. Perceptual categories, which form a non-linguistic linkage between sensory inputs and behavioural outputs, form the initial mechanism of cognition, and it is to these pre-established meaning units that labels are associated, thus overcoming the symbol grounding problem.

Frame Problem

The *frame problem* relates to the difficulty of maintaining and using symbolic representations in a complicated, dynamic and changing world (McCarthy & Hayes 1969, Dennett 1990). If cognition is based on representations, then how do we choose to perform appropriate actions in given circumstances, and further, how can we “recognise not just the intended implications of [our] acts, but also the implications about their side effects” (Dennett 1990, page 147)? How do we evaluate “the difference between relevant implications and irrelevant implications . . . and ignore the irrelevant ones?” And how can we do this quickly and efficiently in a dynamic world?

Dennett pushes the issue further by seeking an answer to the question of how humans are able to ‘look before they leap’—that is, to think through a situation in a reasoned manner and select a course of action.

This is a fundamental issue for the cognitivist paradigm, and one that has not been satisfactorily answered. The symbol processing model, being based on search (algorithms) and representation, is hard pressed to come up with a solution. A more complicated and detailed knowledge base requires a longer search time, or a more complex representation requires longer update times, invariably resulting in an exponential

increase in complexity and access times. A related problem is just how does the system decide to modify representations in a constantly changing world. Further, the *mid-point phenomenon* suggests that no matter how you try to carve up the world into representations, there is always a tendency to require a mid-point (or combined) representation (or concept). The computational difficulties seem intractable.

By contrast, the ABC model does *not* have the same difficulty with the frame problem. For a start, the ABC model is contextually driven, so that a particular context will ‘select’ the behaviour with which it was previously associated. The creature learned this behaviour in conjunction with a context (which includes external sensory inputs as well as recurrent internal inputs resulting from learned sequences), and will reproduce the same (or closely related) behaviour when placed in the same (or similar) context. The ‘winner-take-all’ nature of the model means that the vast majority of behaviours will simply not be selected, that is, will not be ‘considered’. The behaviour and the context are directly linked.

The direct access of the appropriate behaviour (requiring either no cycles of recurrent loops if the context and behaviour are common and so forming a soft-reflex direct link, or a number of cycles of the previously learned temporal sequences at multiple levels in order to ‘latch on’ to a required behaviour) eliminates the computational intractability issue.

New concepts within the ABC model come about as a result of bifurcation of existing concepts, as well as new linkages and associations between existing concepts and temporal sequences. Learned associations also account for the ‘common knowledge’ issue that is required in a model of cognition. All of our behaviours and concepts were learned within a context and a background of existing ‘knowledge’. The new behaviour (knowledge) † will be incorporated with other pre-existing behaviour (knowledge). Over time, associations will be enhanced or decreased, and the linked associations and temporal sequences will form our ‘common knowledge’.

Self-talk, the extension of (external) language that forms a major component of what we term ‘thinking’, provides us with the mechanism for reflections on future

†The ABC model draws together knowledge and behaviour—i.e., language *is* behaviour. Thinking may also be regarded as a form of (internal) behaviour. So the expression of a piece of knowledge, either by speaking about it or by thinking about it, is taken to be the performance of some behaviour.

behaviours—‘looking before leaping’. Talking (or thinking) about a behaviour is part of its context. Because of our past learned experiences, we are able to temporally link events in time—‘if *A* happens then *B* will occur’—and form an internal linguistic trace of temporal sequences to connect these events. Then, given a context of *A*, we can follow the sequence to speak of (self-talk) *B*, thus providing a new context and a new link in the chain of a ‘plan’.

The thing to note is that the ‘plan’ is a *situated* (i.e., contextualised) temporal linkage of pre-learned concepts and temporal sequences.

The ABC model is really just a connected series of dynamic ‘look-up tables’ (soft-reflexes) which are self-organised over the whole life-time of the person. There are really no ‘decisions’ or ‘choice’ of representation particulars involved—just a contextually determined flow over learned concepts and behaviours. There is no ‘decision’ from multiple possibilities, but rather the winning nodes on the SOM surfaces choose themselves based on previous experiences and learning.

Innate Knowledge & Behaviour

Just how much of behaviour is innate and how much is learned has been a perennial issue in psychology. The view taken in this thesis is that learning is more important than perhaps previously accepted. In particular, we reject the view that language understanding is innate as Chomsky and other linguists have proposed.

Of course some behaviour is innate—certain reflexes seem to be innate as they bypass the brain, and other autonomic behaviours are pre-determined. Some behaviours which require a rapid response, or are basic to the needs of the creature, would seem to require a hard-wired mechanism. However, innate behaviours are non-adaptive and hence subject to failure in changing conditions. Flexibility and adaptability would appear to be important for the ‘higher’ cognitive functions.

Innate structures, especially of higher cognitive functions such as language, would also appear to suggest a level of complexity in design. The ABC model does not suffer from this problem. The overall structure is relatively simple, and only general structures of self-organising maps, recurrent linkages and so on, are required. As well as being a much more robust and adaptive structure, there is a degree of commonality and similarity in the structures.

There is a difference between innate structures leading to a possibility of behaviours, and innate behaviours. The ABC model provides an ‘innate structure’ which can learn various linkages between sensory inputs and behavioural outputs—the model is *general*. Innate behaviours are, on the other hand, *specific*. The ABC model does not *prescribe* any *particular* behaviour. The system will just self-organise to do the things that fit previously learned statistics and have served the creatures needs in the past.

Dennett (1990, page 153) makes the comment:

We can all agree, today, that there could be no learning at all by an entity that faced the world at birth as a *tabula rasa*, but the dividing line between what is innate and what develops maturationally and what is actually learned is of less theoretical importance than one might have thought. While some information has to be innate, there is hardly any particular items that must be: an appreciation of *modus ponens*, perhaps, and the law of the excluded middle, and some sense of causality. And while some things we know must be learned—e.g., that Thanksgiving day falls on a Thursday, or that refrigerators keep food fresh—many other ‘very empirical’ things could in principle be innately known—e.g., that smiles mean happiness, or that unsupported, unsupported things fall.

We suggest that other than basic survival mechanisms and some form of general structure within the brain (for example, hierarchical self-organising maps, recurrent linkages, external world filters and so on), which need to be pre-wired, all other ‘world knowledge’ is not innate and must be learned.

We thus disagree with Dennett and suggest that *modus ponens*, causality and the law of the excluded middle are also learned behaviours. As well, we need to be careful in how we discuss innate and learned concepts. To be more precise, we would grant that a sensory input that we, as conscious human beings with our common knowledge would interpret as a ‘smile’ from a care-giver, would no doubt elicit a return facial movement in the form of a smile and a lowering of anxieties through changes in the autonomic system in an infant, and that this basic behavioural response is perhaps inbuilt. But the concepts of ‘smiling’ and ‘happiness’, and their connection, is learned

through observation and training.

Combinatorial Explosion

The construction of any system that requires a large knowledge base which is to be searched (using some form of von Neuman-based computer) of necessity must result in an explosive growth in time and complexity as the number of elements increases. Any combinatorial expression, representing the number of possible ways of grouping elements of the knowledge base according to particular rules, must increase dramatically as the base's size increases. As a simple example, the time to sort n elements is at best proportional to $n \times \log n$. As n becomes increasingly large, the time to sort becomes prohibitively long.

The combinatorial explosion problem is then the issue that, given more knowledge, a representationalist approach will be too expensive in terms of complexity, and hence time.

Not only is this issue of more knowledge requiring longer search times not true of human behaviour (see Section E.3), but the ABC model does not suffer from it either. As we explain elsewhere, a hardware-only implementation of the ABC model selects a winning node on a SOM surface in (near) constant time, independent of the number of nodes on the surface. That is, it takes the same time to traverse a 10 by 10 SOM surface as it does a 100000 by 100000 SOM surface. Only local decisions are required, and these may all be performed in parallel.

Problem Of Induction

Most current models of cognition install *induction* as the primary source of new knowledge. With induction, learning is from the particular to the general. Usually, some region of a feature space is taken to cover instances of a particular concept, and the general formulation or description of the region is taken as being a *rule* for the concept. Other, as yet unseen examples of the concept are expected to fall within the designated region (or regions in the case of disjunctions), and thus also satisfy the rule.

Induction is generally performed over *linguistic* terms. For example, instances of the concept *bird* might include terms such as *feathers*, *beak*, *warm-blooded*, *flies* and so on. After experiencing a number of examples of the concept *bird*, the generalisation

$bird \Leftarrow feathers \ \& \ beak \ \& \ flies$

may be induced as the ‘rule’ for bird.

However, induction has a number of problems associated with it. For a start, it assumes that there are linguistic primitives out of which all other concepts may be built. But this is simply not the case, as was shown by Wittgenstein (1963) and Heidegger (1978). There are no *linguistic simples* in the world that are available to *synthesise* other concepts and words via induction.

Further, as we discuss in Chapter 4 in reference to vision, the world is not ready-made with separable objects or events. The infant first needs to *learn* which objects and events are separable, and hence recognisable as individual perceptual concepts before they are able to be labelled.

As Barsalou (1992, page 296) and others point out, there are potentially an infinite number of possible inductions from a set of examples. This leads to the need to introduce some arbitrary cutoff methodology such as Occam’s Razor, Probably Approximate Correct (PAC) learning or Minimum Description Length (MDL).

There are other issues with induction (see, for example, Gregory 1987) which make it a poor learning tool. The view taken here is that induction is a tool and a product of society—it needs to be taught to be used. However, it is *not* the primary learning mechanism of humans and animals.

Generalisation is achieved within the ABC model through self-organisation on SOM surfaces. Similar input values self-organise to excite neighbouring neurons on the SOM surface. This form of generalisation, combined with the bifurcation of attractors through differential output ‘behaviours’ and temporal sequence learning, form the mechanisms of learning.

The ABC model is also able to explain *single example* learning—another difficulty for induction. If the *affect value* is high, the learning rate is greater.

The Knowledge Problem

Models developed in the cognitivist paradigm derive most (if not all) of their problem-solving capabilities from their predefined body of knowledge, knowledge supplied and interpreted by the program’s creator. They make little or no attempt to hypothesise where ‘knowledge’ comes from, and learning is generally limited to the *induction* of ‘rules’ from structured examples, or the *deduction* of ancillary rules.

In general, no mechanism is postulated to expand the body of knowledge. Symbols (and operators) are either ‘basic’ (defined *a priori*) or represented as *necessary* and *sufficient conditions* from other symbols. Further, no thought is given to the *grounding* of the symbols in the world—their *intentionality*.

The ABC model, on the other hand, is predicated on the learning of new behaviour, or to put it another way, gaining new knowledge. The model is consistent with the observations of child development by Piaget, Vygotsky and others in which children gradually acquire new behaviours, and build upon previous behaviours.

We talk further about *training*, the explicit transfer of knowledge from an experienced member of a society to another, less experienced member in Chapter 5.

Marlsburg Synapses Or Incremental Changes

Computation requires sudden changes in values—the value in a certain storage location is suddenly changed following a particular calculation, or a register is updated with the next program instruction. This is a fundamental requirement of any binary computer program based on the von Neuman architecture which requires the transfer of data to/from a central CPU, as well as external memory. Thus any computational scheme for cognition that is implemented in a biological neural structure would appear to require rapid changes in the connectivity between neurons.

However, it is hard to see how this might be achieved biologically within the brain. According to the current orthodoxy on neuron behaviour, there is no way that new dendrites can be produced rapidly, nor new axons, nor even new axon terminals in the brain. The only plausible alternative is to change an existing synapse in some way.

This idea of rapid and substantial synaptic alteration was put forward by von der Malsburg (1981), and considered by Crick (1984) in proposing his ‘searchlight hypothesis’ of attention.

The overwhelming view of neuroscientists is that drastic changes to synaptic weights do not occur—the prevailing current opinion is that synaptic connections (weights) only change in small increments. Given this restriction, it is difficult to imagine how a *computational* device could work at the neural level.

The suggestion of some form of ‘virtual computer’ which somehow uses the *firing rate* of neurons as a basis for computation is highly unlikely and somewhat fanciful.

How such a device would develop and evolve, much less maintain itself, is extremely difficult to comprehend.

The ABC model does not suffer from this problem as it concerns the *flow* of activity over structures of neurons whose synapses are only changing slowly over time.

Emotions/Drives/Motivations

Current cognitivist models of cognition attempt to portray human nature as rational, intelligent and even sublime. Emotions are not considered to be an important component of human cognition, and may be ignored. As stated in Gregory (1987, page 219):

From the earliest philosophical speculations to the present day, emotion has been often seen as interfering with rationality, as a remnant of our pre-sapient inheritance—emotions seem to represent unbridled human nature ‘in the raw’.

The rejection of affective factors (emotions) was one of the components of cognition that was cited as being deliberately de-emphasised in the cognitivists approach to cognition as it might unnecessarily complicate the cognitive-scientific enterprise (Gardner 1985, page 6). But this is clearly an untenable position.

One of the early and influential studies on emotions was done by Darwin ([1872] 1965), who suggested that there are certain fundamental emotions that are expressed in the overt behaviours of humans and other animals. James ([1884] 1994) turned the conventional wisdom of the day around by insisting that the bodily behaviour observed *is* the emotion, rather than the *result* of the emotion. The James-Lange theory (as it became to be known), postulated that when the individual perceives some external event, various bodily responses are initiated (such as avoidance or approach reactions), together with responses of the autonomic nervous system. The combination of these events then constitutes the emotion that is experienced.

Schacter (1971) suggested that rather than individual and different emotional states, different emotional experiences arise out of the same visceral background. The different responses will then be the result of differing cognitive and perceptual evaluations of the subject’s context. That is, emotional experience is a combination of visceral arousal and cognitive evaluation.

It is clear that emotionally tinged experiences are more strongly remembered than more mundane events. We are all able to recount various aspects of our situation and behaviour that were coincident with some momentous event—the assassination of a president perhaps, or the landing of humans on the moon. Even many years later these events are still etched in our memories.

Piaget ([1946] 1962) suggested that the cognitive and affective aspects of behaviour are “two sides of the same coin.” Perception is not simply a passive organisation of information, but an active selection of various alternate activities based on interests, needs, pleasure or pain. Whenever we cognise anything we make choices based on affect. Piaget also claimed that drives, feelings and emotions are *always* present in cognition, and that the two are inseparable. According to Piaget, emotion constitutes the “energetics” of thought while cognition provides the structure (Cowan 1978).

It has also been shown recently that it is difficult to make logical, unemotional decisions purely on pragmatic grounds (Forgas 1996). Experiments show that very routine, simple and predictable judgements tend not to be unduly influenced by emotions, but the kind of problem that requires prolonged, constructive, and systematic thinking is strongly influenced by emotion. Also, the longer people think about problems the more likely their emotional state will affect the outcome.

Aparicio IV & Levine (1994) stress that emotion, mood, motivation and instinct have a greater role to play in perception and cognition than logic alone.

Emotions are an integral component of cognition, allowing preferential treatment to be given to important behavioural requirements of the organism. It allows the animal to learn to avoid dangers, and to strive to maintain instinctive behaviours such as procreation, shelter and nurture. Emotion helps to determine interest and importance.

The ABC model does allow the inclusion of an emotional components to cognition. This is achieved in two ways: first, higher emotionally charged events are learned more rapidly because of the higher gain in the learning rate; and second, the vector linkage from the emotional centres (for example, the limbic system) allows an emotional context to be included in behaviour.

Connectionism & Feed-Forward Neural Nets

The focus to date within connectionism has been on software methods, and little attention has been directed towards a material ('real neuron') implementation. Also, much of the discussion of neural nets is directed to their use as an *engineering* methodology—artificial neural nets (ANN).

The current interest in Feed Forward Neural Nets (FFNN) does not provide a complete answer to cognition. FFNNs do not take learning seriously on two counts: firstly, they require pre-existing categories/symbols as inputs and fail to provide an explanation of how these categories came about in the first place; and secondly, the learning mechanism (back-propagation of errors) assumes that there exists some inbuilt knowledge of a desired outcome for the learning so that an error difference from this desired result can be computed. Further, there is no provision for deciding which categories are important.

The *backpropagation* learning mechanism is biologically infeasible (there is no evidence of backpropagation in real biological networks), is slow in learning, and has philosophical problems in that it requires the *a priori* existence of categories or symbols—that is, the inputs and outputs are generally fixed and semantically predetermined. In other words, FFNNs are simply pattern recognisers—given the existence of patterns, and a knowledge of the pattern structure, find that pattern within a larger spectrum.

Unless we assume that all categories are innate, then pattern recognition cannot be the underlying mechanism of human and animal learning. This point is made by Skarda & Freeman (1987, page 172):

... the neural system we have described is not best thought of as a pattern completion device, although it may do that. The problem is epistemological: we do not know what a completed pattern is (so convergence to it cannot be ascertained as in an error correction device), nor, we suspect, does the brain.

The position taken by the ABC model is that the brain is not supplied with symbols from birth, but performs perceptual categorisation from *tabula rasa*. The process of cognition is self-organisation rather than pattern completion.

One of the criticisms of connectionist approaches to cognition has been that it is difficult to see how local interactions can lead to an overall result without explicit

executive control. The ABC model goes a long way to showing how the coordinated and complex behaviours of humans and animals may be achieved within a connectionist architecture.

The neuroanatomical structures of the cortex are much more specific, with their own characteristic architecture, than those normally studied from a theoretical perspective under the name ‘neural nets’. The ABC model is an attempt to describe the full architecture of the neocortex, as well as an explanation of how this phylogenetically newer component is able to be incorporated into the rest of the brain. It goes well beyond current connectionist models.

Another point to be made is that neural networks in nature are inherently parallel *hardware* devices rather than software devices. There appears to be a fair deal of confusion in some quarters on this issue. For example, Anderson (1995, page 106) states:

... Fodor & Pylyshyn (1988) made some telling arguments against the promiscuous application of connectionism to cognition. The essential criticism they make is one that an engineer would be happy to make: associative neural networks are such an inefficient way to compute that it would be foolish to build a cognitive system like that.

As is shown elsewhere in this thesis, the reverse is the case. Computer simulations of neural networks are indeed slow, as are cognitivist attempts at building realistic models of cognition, but a hardware implementation of the ABC model should be able to show rapid learning capabilities, as is the case with humans.

3.5 What Is Left Out?

In this section we describe features and procedures which are not found in the ABC model. These features have been left out simply because they are unnecessary, or a deliberate decision has been made to eliminate them because it is perceived that they present a barrier to a clear understanding of the brain. However, all may be later accounted for in the ABC model.

Logic

One feature that has been deliberately excluded as an *explicit* component of the model is the process of logic. The view taken in this thesis is that cognition is *statistics*, not *logic*.

Far too much emphasis has been given to logic in models of cognition. The ABC model suggests that logic is a *product* of cognition, requiring much learning and training—and not an integral *component* of cognition.

The identification of logic and thinking has a long history in philosophy, going back at least to Aristotle and including Boole, Russell and early Wittgenstein. Theories of knowledge have tended to show a strong orientation towards the inclusion of logical thinking (Johnson-Laird 1983, Johnson-Laird 1989, Johnson-Laird 1995).

Knowledge in such models is generally assumed to be a formal propositional system, with attendant logical properties such as completeness and consistency. Propositions are processed according to the rules of logic, and new propositions are derived from old ones through the application of various operators, connectives and quantifiers (such as *not*, *foreach*, *forall*, *and*, *or* and *implies*).

Sigmund Freud challenged this traditional view, suggesting that there are highly irrational, powerful influences which affect our behaviour and our perception of the world. More recently, psychologists have shown that human knowledge processing is indeed *not* logical. †

Hampton (1982) proposed that *similarity* rather than logical principles controlled the decisions made by subjects about category membership. Other similar findings include Hampton (1987), Hampton (1988), Osherson & Smith (1981), Osherson & Smith (1982) and Smith, Shoben & Rips (1974).

Another aspect of the non-logical behaviour of humans has been termed *availability* by Holyoak & Glass (1975) (see also Glass & Holyoak 1986). In this study, the time to respond to false sentences (such as *arrows are dull*) was measured. Logic would predict

†Note that this does not imply that it is necessarily *illogical*. For example, in a study by Hampton (1982), subjects were asked to answer questions regarding category inclusion. In one response, subjects stated that *typewriters are office furniture*, that *office furniture is furniture*, but also that it is *not* the case that *typewriters are furniture*. This breaks the transitivity rule of logic: if *all X are Y* and *all Y are Z* then *all X are Z*.

that the response times should be the same for all false statements, but Holyoak & Glass found that response times differed according to the strength of associated meanings needed to form the contradiction.

Anthropologists who have studied preliterate cultures have found that Western logic is not universal. Other cultures have other means of classifying and describing their experiences which, while different from ours, are no-less appropriate to their environment. An example is the study by Vygotsky & Luria (1993) who witnessed the rapid social changes that came about when non-technical cultures, like Uzbek and Central Asia, were suddenly expected to participate in the technically advanced western culture of the new Soviet Union. As well as affecting attitudes and beliefs, Vygotsky felt that social context has a profound influence on how we think, as well as what we think.

Logic is a hard won skill of a society. Philosophers and other intellectuals pride themselves on their use of logic, yet this skill is only achieved after many years of training and practice—it is a learned behavioural tool which has evolved through culture over many generations.

Obviously we are not saying that we need not be logical in our analysis or in our construction of the ABC model, or that there is no such thing as logic, but that logic is not a *basic component* of cognition. However, the long history of a love of logic in Western culture cannot be sustained at the neural level.

Intelligent Behaviour

Also left out of consideration in the ABC model, *in the initial stages*, is any reference to 'intelligent' behaviour. Cognitivists, and the AI pioneers such as Newell and Simon in particular, showed an infatuation with intelligent behaviour such as game playing. Animals, however, including humans, evolved to adapt to the environment, not to play chess. The cognitivists neglected to show how such intelligent behaviour could result from the humble beginnings of animal and child behaviours. They concentrated on mature intelligent behaviour rather than indicating a phylogenetic transitional path or any feasible learning mechanism, and neglected to account for child development to any great extent.

A full understanding of such intelligent behaviour as chess playing is still to be achieved. The recent contest between a human and a computer in a chess championship did not employ such understanding: the two participants used completely different strategies. The machine employed brute force, searching billions of potential moves at each move while the human used gestalt behaviour (perceptual attractors) to recognise a much smaller number of possibilities.

It is the proposition of this thesis that the mechanisms of the ABC model are similar to those used by the human player. However, there is obviously much work to be done before we can claim to be able to reproduce the intelligent behaviour displayed by such human experts.

3.6 Summary

The ABC model combines perception, learning, conceptualisation, language and action (behaviour) in a common mechanism and architecture. Although some behaviour may be innate, learning through self-organisation, association and temporal sequences is seen as the mechanism of cognition.

The model is consistent across the phylogenetic tree, and provides for continuity across reflexes, learned perceptual conceptualisation, language and thought. Recognition (or observation, or perceiving) and learning as seen as coincident, and the model combines perceiving, thinking and action as the coherent coordination of behaviour.

The model is dynamic—the neural weights are *constantly* changing. Memory, our perception of the world, concepts, and so on are thus constantly changing—we are the current sum of our past experiences. Our world-view changes over time—children view the world in a very different light from an adult

The model is not a *computational* model; while the model can be simulated on a serial computer, the ABC model has very little in common with the processes of a serial (von Neuman) computer. This is an extremely important point as it allows us to break away from the *representationalist* view of cognition that has impeded progress for many centuries.

Chapter 4

Vision

4.1 Introduction

The chapter is intended to give a reasonably detailed examination of one of the sensory modalities—vision. We first examine some of the neurophysiology and neuroanatomy of the primate visual system, and then look at some of the current theories which attempt to explain, and in the case of computer vision, reproduce the visual function. We then critically examine the claims and methods made in current theories, and discuss some recent psychological and neuroanatomical findings which bring many issues into question. These biological observations indicate that the current theories of visual recognition cannot be accepted as being viable for human and other primate vision.

The ABC model is then discussed in relation to vision, and some temporal learning experiments using the temporal model are examined. Of specific interest is the means by which arrays of neurons, and their hierarchical and recurrent interactions in general, can be adapted to produce temporal structures which encode spatial information.

The ABC model takes a new approach to visual recognition—rather than being based on a spatial fusion of visual information, the ABC model utilises temporal sequences. The ABC model provides an alternative approach to explaining visual cognition which rejects the traditional computational models, positing instead a dynamic neural model.

Finally, a number of experimental observations are discussed in relation to the model.

4.1.1 Basic Physiology and Anatomy of the Visual System

Before examining the proposed ABC model of biological vision, it is appropriate to briefly examine some of the basic physiology and anatomy of the primate visual system. This rather cursory examination is a reminder of the constraints that must be satisfied by any serious model of biologically-based vision.

The human eye contains some 130 million rods and approximately 8 million cones. Rods, which are monochrome receptors of light, are generally found to be in the majority in creatures, such as the owl, which are active at night. Cones, on the other hand, are selective for different wavelength ranges which gives a percept of colour. † There are three types of cones, each sensitive to a different part of the spectrum. The absorption curves of the three types of cone are illustrated in Figure 4.1, which indicates a peak wavelength of approximately 420 nanometres for the short-wavelength sensitive S cones, around 530 nanometres for the medium-wavelength sensitive M cones, and 560 nanometres for the L cones which are sensitive to long-wavelength light. ‡ Animals with a preponderance of cones tend to be active during daylight (Sekuler & Blake 1990).

The distribution of rods and cones in the retina is very irregular. At the centre of the macula, only cones are found, with no rods. In the periphery, mainly rods are found, with only a few cones. The distribution of rods and cones is shown in Figure 4.2.

Further, the number and distribution of cones within the retina is also very non-uniform. It is thought that of the 8 million cones, only about one million are S cones, with approximately 2.3 million M and 4.7 million L cones (Cicerone & Nerger 1989). § S cones are scarce at the fovea, are at a maximum concentration just outside the fovea, then taper off in numbers with increasing eccentricity. M and L cones, on the other

† However, each individual cone simply accumulates photons within a particular range of wavelengths.

‡ These wavelengths roughly correspond to the colours of violet, yellowish-green and yellow respectively.

§ Kaiser & Boynton (1996) give a ration of L to M to S cones of 32:16:1. See also Zrenner, Abramov, Akita, Cowley, Livingstone & Valberg (1990, page 175) for a plot of the distribution of cone density in the human retina.

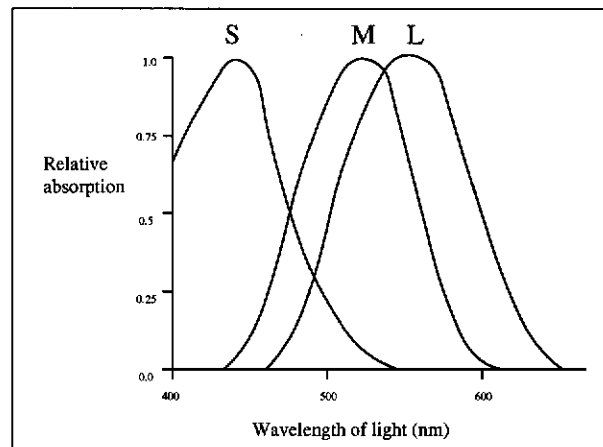


Figure 4.1: Absorption Curves for the Three Cone Types in the Human Eye.

hand, show the highest concentration at the fovea before gradually decreasing towards the periphery (Sekuler & Blake 1990).

The 130 million photo-receptors converges onto about 1 million retinal ganglion cells. Near the centre of the macula, the convergence ratio is about unity, while in the periphery, the ratio is several hundred to one. For cones, a ratio of one-to-one in the fovea decreases to about one ganglion cell for every two cones beyond about 10° eccentricity.

The photo-receptor responses are transferred via bipolar cells to retinal ganglion cells (RGC), whose axons project to the *lateral geniculate nucleus* (LGN). From the LGN they synapse onto relay cells, and then project to area V1 of striate cortex.

The manner in which the three cone types are combined for transfer to the visual cortex provides for one achromatic channel and a pair of chromatic channels. The achromatic channel is obtained by summing the M and L cone responses. The two chromatic channels provide colour-opponent schemes: a *blue-yellow* channel is obtained by summing the responses of the M and L cones, and subtracting the S cones, while a *red-green* channel is obtained by summing the responses of the S and L cones and subtracting the M cone responses (Sekuler & Blake 1990).

The rods and cones are combined in such a way that the ganglion cells are responsive to particular *receptive fields*—small regions of light that exhibit a central region of excitation (or conversely inhibition) surrounded by an annulus of inhibition (excita-

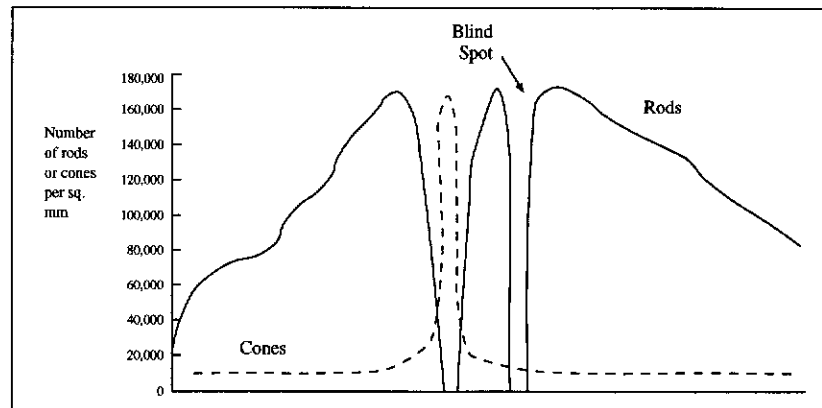


Figure 4.2: Distribution of Rods and Cones in the Retina. (Adapted from Sekuler & Blake (1990, page 54))

tion). Receptive field sizes vary systematically with retinal location—within the macula some are of the order of only 0.01 millimetres, whereas only 10 millimetres away from the fovea the receptive field centres may be up to 50 times larger. At each retinal eccentricity there are multiple receptive field sizes (Sekuler & Blake 1990, page 66).

Two major pathways originate from the retinal ganglion cells, producing a rich set of data pathways that interconnect a multitude of higher visual areas. These are the parvocellular and magnocellular streams. There are striking differences in temporal processing between parvocellular (P) cells and magnocellular (M) cells. P cells, which comprise about 80% of ganglion cells, tend to give sustained responses to a maintained stimulus. M cells, in contrast, respond only transiently to a maintained stimulus, yet respond best to modulations of 20 Hz or greater, continuing to respond up to 60-80 Hz (Van Essen & Anderson 1989).

P cells carry chromatic information by virtue of the spectral opponency of their receptive fields. Because of the relative numbers of cone types, most P cells have red-green opponency, but a small minority have blue-yellow opponency. The M cells exhibit only minor spectral opponency, but behave non-linearly to a spectral contrast (Van Essen & Anderson 1989).

Lesion studies confirm the physiological data (Van Essen & DeYoe 1995, page 393). P-layer lesions lead to a large deficit in colour discrimination, as well as moderate deficits

in form and depth discrimination. M-layer lesions produce deficits in motion, depth and form perception. The two streams contribute jointly to many aspects of perception, but provide separate contributions to the perception of colour and motion.

Of the axons of the ganglion cells that originate from the retina, about 80% project to the LGN while the remaining 20% of the optic tract projects to several structures in the midbrain, the most prominent of which is the *superior colliculus* (Sekuler & Blake 1990, page 100).

The superior colliculus is a phylogenetically older, more primitive visual area than the visual cortex, and in many lower order animals such as frogs and fish, the superior colliculus is the primary visual processing area. However, in vertebrates, the phylogenetically more recent visual cortex has taken over as the more important visual area. The superior colliculus appears to be used to detect objects away from the point of fixation, and to guide orienting movements of the eyes and head towards an observed object. It is not thought to contribute to a detailed visual analysis of objects (Sekuler & Blake 1990, page 100).

Of the cells projecting to the LGN, half go to the same side of the brain as the source retina, and half go to the LGN on the opposite side of the brain. Individual LGN cells then relay the signals to the striate area V1 in approximately a one-to-one manner.

The LGN contains 6 layers. These layers serve to segregate the inputs from the P and M cells, and to align the inputs from the two eyes while keeping them physically apart. The upper four layers (parvocellular, or P) receive inputs from the retinal P cells, while the retinal M cells connect to the lower two (magnocellular) layers. Figure 4.3 (a) shows a representation of the LGN layers, and Figure 4.3 (b) the relative mapping of the visual field to retina and on to the LGN (Sekuler & Blake 1990, pages 102–103).

At the cortical stage there is a vast expansion in the numbers of neurons, with hundreds of cortical neurons for each LGN neuron (Van Essen & Anderson 1989). The visual cortex accounts for about 50% of the total cortex in the macaque monkey, with area V1 (striate cortex, or primary visual cortex) accounting for about 15% alone. For humans, V1 is about twice as large, but because of the ten-fold increase in cortical size over the macaque, it accounts for only about 3% of the total cortex.

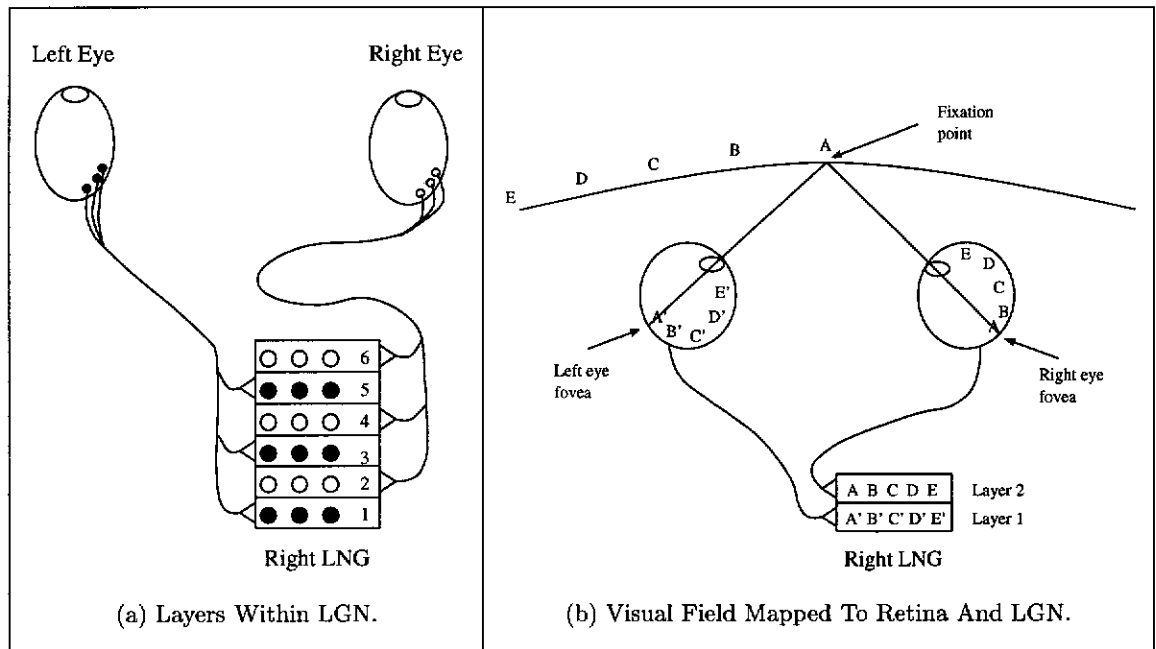


Figure 4.3: Lateral Geniculate Nucleus.

The volume of the cortex dedicated to the macula, the region of highest visual acuity, has been estimated to be 35 times greater than the area of the cortex dedicated to an equal amount of retinal input from the periphery. There are approximately 1.3 billion visual cortical neurons (counting both hemispheres), which corresponds to nearly 600 cells for each LGN input (Van Essen & DeYoe 1995, page 386).

As many as 32 separate visual areas have been isolated in the macaque which are largely or exclusively visual. An additional half dozen or so areas receive strong visual inputs, but also receive major inputs from other modalities. The total number of corticocortical connections may be as high as 305, nearly a third of the total possible number of pathways interconnecting these areas. A more detailed description of the currently postulated areas and connections is given by Van Essen & DeYoe (1995, page 388).

Figure 4.4 shows the sub-cortical (RGC and LGN) and cortical streams for the macaque monkey. The three sub-cortical streams are, in order of prominence, the P, M and K streams. The two most studied, the P and M streams have been described previously. The K (koniocellular) stream is the least understood, despite the fact that K cells

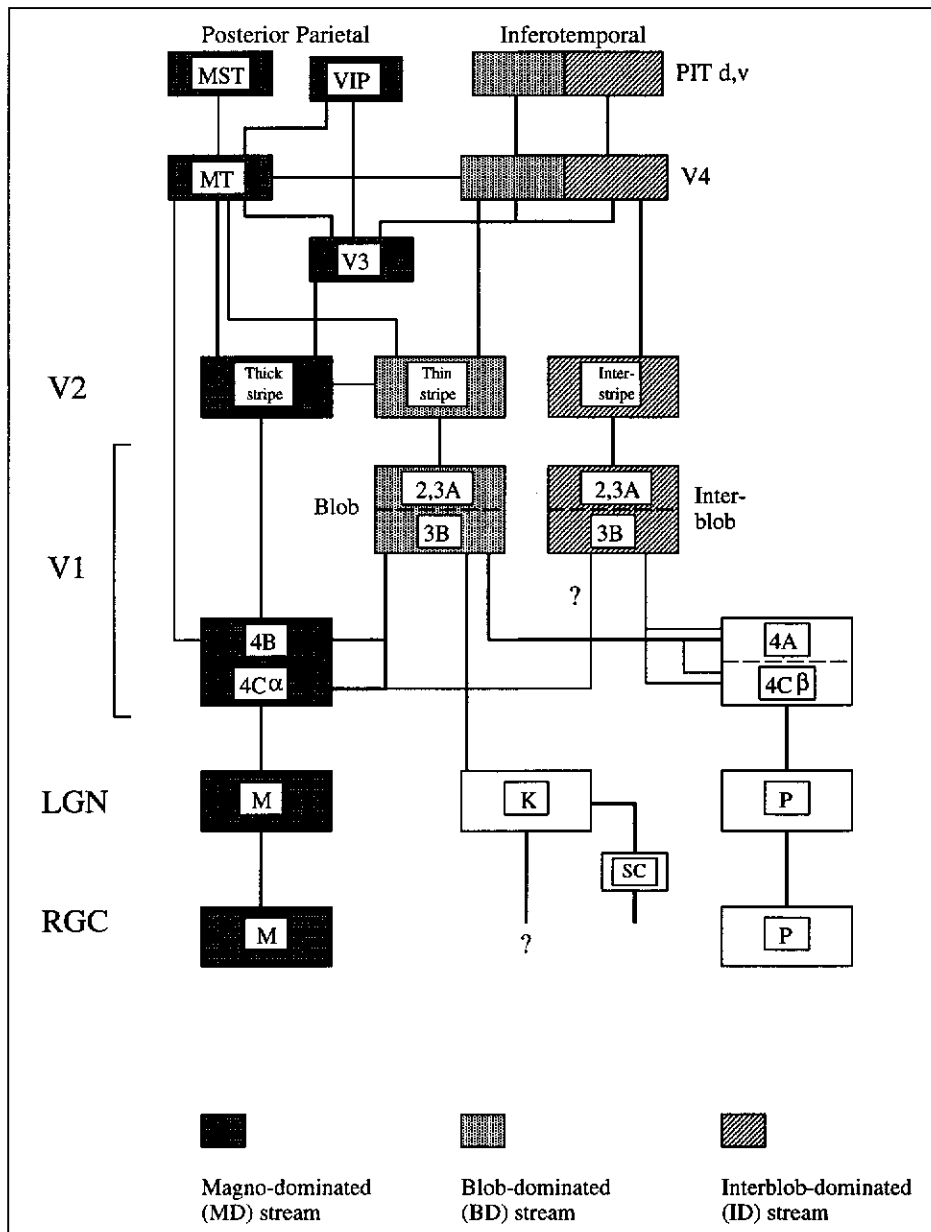


Figure 4.4: Sub-cortical and Cortical Processing Streams in the Macaque.

(Adapted from Van Essen & DeYoe (1995, page 390).)

are as numerous as M cells. K cells project to the blobs of V1, and receive selective inputs from the superior colliculus. K cells have a small size, are sparse in their cortical terminations, and show a sluggish response to visual inputs. For these reasons, they are thought to play only a relatively minor role in the transmission of ascending sensory information (Van Essen & DeYoe 1995).

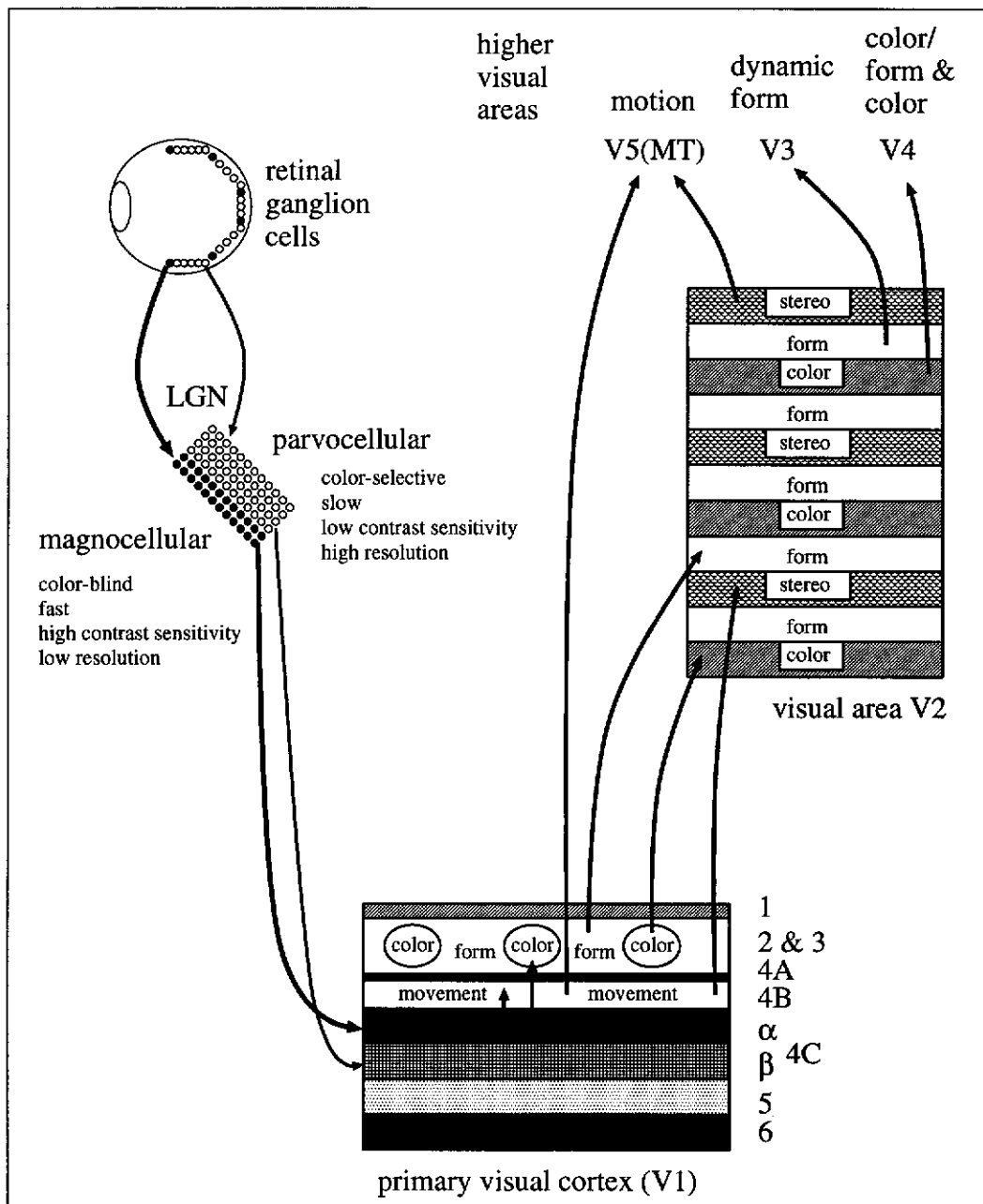


Figure 4.5: Functional Segregation in the Primate Visual System. (Adapted from Livingstone & Hubel (1988). Original copyright 1988 by the AAAS.)

Within V1 and V2, the P and M streams are split and combined into a number of streams which project on to areas V3, V4 and V5 (MT), and then on to other areas. There are also a number of interconnections between areas V3, V4 and MT. These

paths and connections are shown in Figure 4.5.

Most cells in area V1 of the cortex are orientation selective. Each cell responds to a receptive field at a given position in retinal space. Area V1 thus provides a 'map' of the retinal surface in terms of spatial frequencies and angles. Figure 4.6 shows the receptive field responses produced from a long electrode penetration made obliquely to the cortical surface.

As well as the described forward connections, numerous back connections are made between cortical areas. Van Essen & DeYoe (1995) cite a general principle that connections seem to be arranged in reciprocal pairs. For example, V1 projects to V2, and V2 projects back to V1. Most reciprocal pathways are asymmetrical with regard to the cortical layers in which connections originate and terminate, suggesting forward pathways with reciprocal feedback connections. In the visual system, there are strong feedback connections at every level beyond the optic nerve. Lateral connections are also found. Van Essen & DeYoe have postulated 10 hierarchical levels within the visual cortex, with the two sub-cortical levels at the beginning, and projection on to the limbic system following visual processing. Connections occur more frequently between areas at nearby levels of the cortical hierarchy than between areas that are separated by many levels.

The functionality of higher cortical areas is still uncertain. Selectivity of various kinds, (as encountered in area V1 with orientation selectivity), is also found at coarser spatial

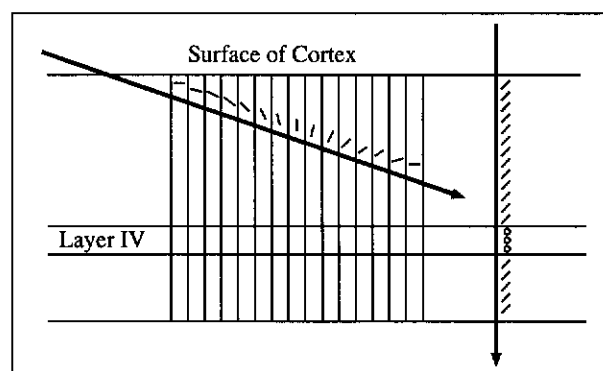


Figure 4.6: Change in Orientation Selectivity in Area V1. (Redrawn from Hubel & Wiesel (1977).)

scales in extra-striate areas. In V4, colour-selective and orientation-selective cells are reported to occur in separate clusters (Zeki 1983). Motion-related selectivities have been found in areas MST, including motion characteristics relating to optic flow such as rotation, expansion and contraction (Tanaka & Saito 1989), while area MT also exhibits motion selectivity. Area V3 has been shown to respond to spatial frequencies and angles (Zeki 1993).

Zeki & Shipp (1988) describe V1 and V2 as segregators, containing separate groupings of orientation, colour and motion selective cells. Signals from V1 and V2 are then sent to three pivotal cortical regions: areas V3 and V3A (the V3 complex), which is involved in dynamic form analysis; area V4 (the V4 complex) involved in colour analysis in association with static form analysis; and areas V5 and V5A (the V5 complex) involved with motion. These pivotal areas may communicate with each other either directly, through recurrent connections to V1 and V2, or through confluent forward connections with areas in the parietal and temporal lobes.

At even higher levels in the hierarchy, neurons in inferotemporal cortex respond to complex shapes, including natural stimuli such as faces and hands (Tanaka, Saito, Fukada & Moriya 1991). For the posterior parietal cortex, the response properties suggest involvement in the analysis of spatial relationships, eye movements, and target selection for visual attention (Colby 1991).

A more in-depth description is given by Zeki (1993) or Hubel (1995). A detailed 'road-map' of the currently proposed connections may be found in Van Essen, Anderson & Felleman (1992). See also Young (1992) for a topological analysis of the connections in the primate visual system.

4.1.2 Eye Movements

Eye movements play a very important role in visual perception. The eyes saccade every 200 to 300 msec, with the saccade period being very brief—only about 10% of the total viewing time. Saccades are both ballistic and extremely rapid, with speeds of up to 1,000 degrees per second (Bridgeman, der Heijden & Velichkovsky 1994). The actual

saccade may take as little as 10 msec, and even a 40° saccade has a duration of only about 100 msec, much too short a time for there to be any guidance of the movement by error feedback from the retina or from proprioceptors. Thus it seems likely that saccades must be predetermined (Howard 1982, page 260).

Visual sensitivity is reduced just prior to, and during, a saccade. This *saccadic suppression* severely degrades motion and position information taken in during saccades. †

Fixations are the pauses between saccades. For the few hundreds of milliseconds of the fixation period, the gaze is virtually still with respect to the head. ‡

The foveal area of a human eye restricts high-resolution vision to about 2° of visual angle. This is approximately the angle subtended by a thumb when the hand is held at arms length (O'Shea 1991).

The traditional view has at least three processes taking place within the 300 milliseconds or so of a typical fixation (Viviani 1990); the analysis of the visual stimulus in the foveal field, the sampling of the peripheral field, and the corresponding planning of the next saccade.

At a higher cognitive level, this view maintains that whenever an observer explores the visual world in order to solve a problem, at least three cognitive operations seem to be involved (Viviani 1990):

- the activation of a set of a priori beliefs about the possible states of the world,
- the breakup of any complex, holistic hypothesis into an appropriate hierarchy of simpler alternatives,
- the translation of these alternatives into an actual sequence of spatial locations whose contents are most likely to disambiguate the alternatives.

†Note also that the spontaneous activity of V1 neurons is suppressed according to the onset time of saccades. The suppression begins about 20-30 msec after the saccade is initiated, and lasts about 200 msec (Duffy & Burchfield 1975).

‡Small positional drifts and microsaccades may occur during longer fixations, but these do not usually exceed 10 minutes of arc.



Figure 4.7: 'Wife and Mother-in-Law' Polysemous Image

Not much is known about the relationship of eye movements to actual cognitive processes, but Viviani (1990, page 385) summarises some results of eye movements which can be related to discriminable mental states.

- when subjects observe the two (or three) perceptual solutions of a polysemous figure which have well-known attractors (such as the eyes and lips in Boring's 'Wife and Mother-in-Law' image [†] shown in Figure 4.7), their eye fixations tend to cluster around the appropriate attractors while perceiving a particular solution.
- in viewing difficult, hidden images such as that shown in Figure 4.8, in which definite contour information is lacking, a similar shift of attention occurs. As soon as the target pattern is (partially) recognised, [‡] the eye quickly fixates the salient features of the new gestalt. The resultant fixations mean that some hitherto meaningless blobs now acquire a representational meaning, and the new, emergent configuration is confirmed.

[†]Also known as the 'Young Woman/Old Woman' image.

[‡]'Pops out' is the term often used. The scene in Figure 4.8 is a dalmation dog approaching a tree. On first viewing, it is very difficult to decipher the image, but having deciphered the image once, it is virtually impossible not to pick out the dog and tree almost immediately.



Figure 4.8: A Classical 'Difficult Image'

- in observing technical images such as chest X-rays or radar displays, skilled professionals immediately concentrate their fixations on certain details which are only meaningful and important in the context of an expert, conceptual description of what is represented (Carmody, Nodine & Kundel 1980). Laymen, on the other hand, tend to scan the whole image in a random, uniform way. †

When subjects look at faces, the gaze mostly concentrates on the eyes and mouth, even after recognition has occurred (Yarbus 1967, Groner, Walder & Groner 1984). This concentration appears to be in order to ascertain the attitudes and feelings of the target person, and it is often postulated that scanning these two features of the face, being the most important for discriminating the feelings of the observed person and thus very important for survival and social intercourse, is somehow hard-wired into the process of face recognition.

We examine the evidence for spatial integration (that is, integrating retinal information across saccades) in Section 4.2.4.

A brief overview of current and previous models of vision, especially in relation to the computer vision literature, is presented in Appendix E. The various problems that

†Similar differences in eye movements between experts and novices occurs in the analysis of chess positions (De Groot 1978).

these models face are also discussed.

4.2 Biological Considerations

There are many examples of experiments which tend to question the traditional view of vision (discussed in Appendix E). Churchland et al. (1994), and several preceding papers by Ramachandran (1985, 1986, 1988, 1992) and Ramachandran & Anstis (1983, 1986), report on a number of experiments which bring into question the serial hierarchical view in which there is a one-way flow of information from *early processes* (such as edge detection, shape from shading, stereo matching) to *later processes* (pattern recognition). The experiments provide evidence for a much richer, more interactive perceptual system, and provide strong neurobiological and psychological plausibility to the need for an overhaul of the traditional views.

In this section we briefly report these experimental results, and direct the reader to the original articles (in particular Churchland et al. (1994)) for full details.

4.2.1 Hierarchical Processing and Priming Effects

The traditional flow diagram for a typical computer vision system resembles that shown in Figure 4.9. The process is a top-down hierarchical one, with only forward interactions between sub-processes. Unfortunately this model is not tenable for biological visual systems for a number of reasons.

One such reason is the observation that many recognition tasks are *primed* by familiarity and context. A number of experiments dealing with eye movements are significant. Compatibility effects result in reaction times which are shorter when stimulus and response are spatially homologous than when they are not (Hallet & Adams 1980). Preparatory effects are seen when a prior signal pre-cues a required movement, reducing its latency. For example, a non-specific visual warning before a target onset reduces saccadic latencies (Ross & Ross 1980).

There are a number of priming effects that are observed which involve figural and

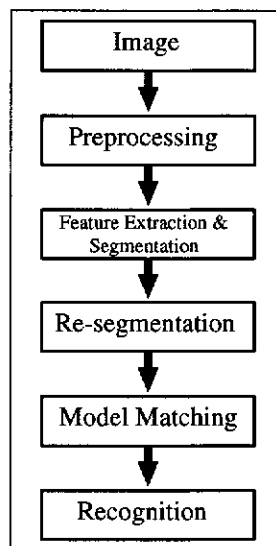


Figure 4.9: Typical Structure of a Computational Vision System.

linguistic stimuli. For example, a letter string can be presented in the peripheral field to facilitate the naming of a visually similar word which appears later in foveal vision (Rayner, McConkie & Ehrlich 1978).

Similar priming was found when simple line drawings were used instead of words (Pollatsek, Rayner & Collins 1984). Subjects were shown a line drawing of an object in peripheral vision. During a saccade to the object, the initially presented picture was replaced with another picture that the subject was instructed to name as quickly as possible. Strong facilitation (150 msec) compared to a control condition occurred if the second picture was identical to the first. Facilitation also occurred if the two pictures were conceptually similar, for example two different pictures of a horse (90 msec) or visually similar (e.g., a ball and a tomato).

Priming is also found with familiarity—the ‘experienced eye’. Many examples may be cited, such as chick sexing, the interpretation of x-ray photographs, or experienced bird watching (Gibson 1969). Familiar visual objects are processed faster and with more accuracy than are unfamiliar objects. Numerous studies have demonstrated this *perceptual learning*, both in humans and animals. For example, Humphrey (1974) was able to train monkeys to differentiate various other species (such as pigs and cows), and to even recognise individual members of each species. Also, Krueger (1975) has

reviewed a number of ways in which familiarity affects perception.

4.2.2 Adaption to Discordant Stimulation

It is generally assumed that the sub-cortical sensory systems responsible for the initial coding of stimulus attributes are 'hard wired' in the human neonate. For instance, the retinotopic projection of ganglion cells to the visual cortex is fixed before birth, and is normally not affected by subsequent environmental influences. However, other sensory mechanisms, such as the orientation detectors of the visual cortex, pass through a critical period in early infancy. During this development phase, environmental influences may indeed modify the orientation detectors, as was shown in the numerous experiments performed on animals which were raised in artificial environments (see Howard 1982, pages 115–119). For example, Blakemore & Cooper (1970) report an experiment in which kittens were raised for several months in an environment which consisted entirely of vertical or horizontal stripes. Careful behavioural studies showed that the cats suffered a permanent loss of visual acuity to lines at orientations other than the one to which they had been exposed. Similar experiments with normally-reared adult cats do not produce comparable results, so that following development, the orientation detectors remain largely immune to such influences. †

Similarly, the basic machinery of the motor system is laid down in the embryo or in early infancy. Nevertheless, the spatial coordinated behaviour of adult human beings is remarkably flexible and readily adjusted to suit changes in stimulus conditions. This flexibility seems to arise in the processes which relate one sensory system with another, or which coordinate sensory events and motor responses, and not in the basic sensory or motor mechanisms themselves.

A number of studies of the adaption of the visual system to discordant stimulation are discussed in Howard (1982, pages 480–520). The classical study is that of Stratton,

†Mize & Murphy (1970) suggest a note of caution in generalising these findings. They found that rabbits have a full complement of innate orientation-selective cells, and that rearing with vertical lines has no effect. However, Blakemore (1970) has suggested that plasticity of orientation detectors is required in animals with binocular vision. While humans and cats do have binocular vision, rabbits do not.

who, for a number of days, wore spectacles which inverted and reversed the visual scene (Stratton 1897a, Stratton 1897b). Stratton found that his actions gradually became more smoothly coordinated over time, although each one had to be mastered separately. He found himself gradually adjusting to the conflicts produced by the inverting lenses, and people and objects that were in his surroundings began to look real, rather than incongruously reversed. Early on in the period of adjustment to the inversion of the visual image, Stratton reported that there appeared to be a *double representation* of the felt position of a limb being viewed, but that the older representation weakened progressively, especially during activity.

Anstis (1992) has studied the adaption to a negative, brightness-reversed world through the use of a modified video. He found considerable initial difficulties in recognising even facial expressions, but found that he was able to adapt over a period. †

Various studies have shown adaption to displaced vision with the use of attachments such as a prism (Howard 1982, page 493). Following a period of inaccuracy, the subject is able to adapt and correctly locate objects. When the prism is removed, a further period of adaption to reverse the prism effect is required.

Numerous other examples of imposed transformations of the visual stimulus array are described by Gibson (1969, pages 193–213), including displacement, inversion, false colouring, minification, magnification, enhanced binocular disparity, and displaced auditory stimulation. In all cases for which the sensory alteration was systematic and continuous, the subject's perception underwent some form of adaption in the direction of a realignment to their prior unmodified world-view. Further, when the modifying instrument was removed, a further period of reverse adaption was required before their perceptions returned to normal.

†Interestingly, his abilities in a number of perceptual tasks improved over time, but they “felt subjectively like inferences rather than direct perceptions”. Further, he felt a “reality mismatch” between the negative images and the sounds normally associated with them.

4.2.3 Invariance

Many computer vision systems assume invariance to translation, rotation and scale. However, there is good evidence that at least rotational and scale invariance within vision are not evident in human visual perception. For example, the recognition of upside-down faces is extremely poor, even for very familiar faces (Yin 1969). Thompson (1980) showed that upside viewing was extremely poor at determining the expression (smiling, frowning) on a face. For example, a display of two upside-down images of Margaret Thatcher contained one in which the eyes and mouth were actually printed the right-way up, opposite to the rest of the face, while one was simply shown upside-down. Subjects were unable to accurately determine the expression of those which were simply displayed upside-down, but were able to determine the expressions of the altered images, thus indicating that recognition did not utilise a global transformation.

Scale invariance has been shown to be culturally determined. There is much anecdotal evidence for this. For example, a report of natives reared in the close confines of a jungle, who initially fail to scale objects appropriately when they first experience open spaces of a plain, and see an elephant on the horizon.

A thorough cross-cultural study of the influence of culture on visual perception was undertaken by Segall et al. (1966). The researchers studied the responses of subjects from various societies (13 non-Western peoples and 3 of Western background) to a number of visual illusions (see Figure 4.10). Their conclusion, and that of other similar studies (for example, see Segall, Dasen, Berry & Poortinga 1990, Hudson 1960) clearly indicates that visual perception is influenced by the environment and culture experienced during development. The differences in experiences point to the conclusion that to a substantial extent we learn to perceive. †

In a test of spatial invariance, Nazir & O'Regan (1990) exposed subjects to a completely new and unfamiliar pattern, but ensured that it impinged on the retina in only a single

†In order to show that these effects are not simply as a result of two-dimensional figure illusions, Brislin & Keating (1976) constructed a three-dimensional version of the Ponzo (perspective drawing) illusion. Their results confirmed the earlier cross-cultural results with two-dimensional figure versions of the illusion.

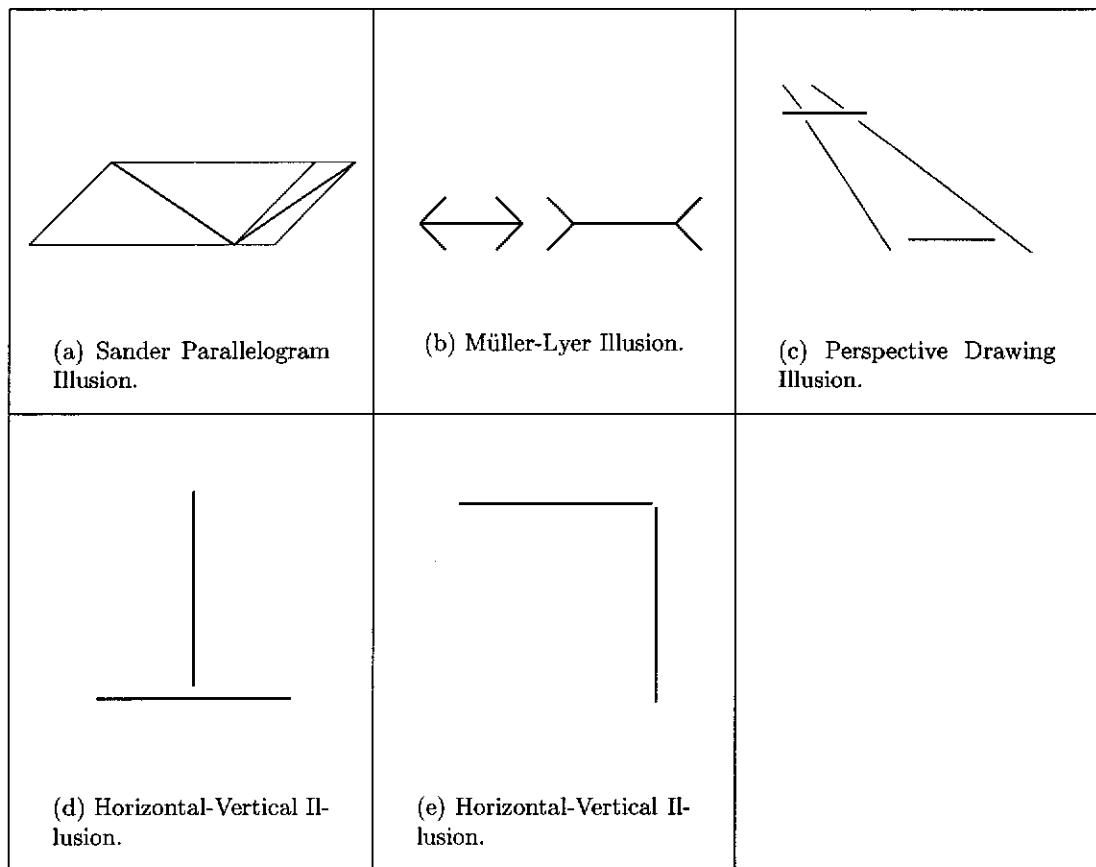


Figure 4.10: Visual Illusions Tested for Cultural Differences.

retinal location. Following learning, the same pattern was presented at other retinal locations. The subjects were unable to recognise the objects at the new locations, some in fact denying that they had ever seen the shape before. After a few presentations, however, subjects *were* able to make a correspondence with the initial pattern.

This suggests that some form of global operator for translational invariance is not found, but rather that the perceived translational invariance of objects is due to individual learning episodes at many retinal locations during the long training period of early life, a view shared by Hebb (1949, pages 47–48).

4.2.4 Saccades, Fixation and Spatial Fusion

Many papers have been devoted to the the question of how the brain maintains stability of internal image representations following saccadic eye movements—a trans-saccadic

fusion of visual information (see, for example, Bridgeman et al. 1994). An implicit (and sometimes explicit) assumption of most current vision systems is that the internal representation of the retinal image is somehow maintained across saccades, with the fovea providing a detailed sub-image to *fill-in* the total perceived image. For example, Potter (1983) proposes a form of integrative visual buffer where information from successive fixations is pasted together to provide a coherent alignment of the individual snapshots.

O'Regan (1992, page 462) points out the 'imperfect' nature of the image we obtain of the world.

- The retina of the human (and other vertebrate) visual system is constructed with *obstacles* in the light path (blood vessels, neurons and axons making up the early stages of the visual system, and even the body of the rods and cones themselves, which appear to be “in backwards”). The receptive system seems to be *inverted*.
- A “blind spot” formed by the optic nerve exiting the retina.[†]
- A severely non-uniform *spacing* of both the rods and cones in the retina (including *within* the foveal area). See Figure 4.2 and Section 4.1.1.
- A non-homogeneity of the colour receptors (cones) across the retina, as well as a thinning of the yellowish macular pigment, both producing a non-uniform colour image (O'Regan 1994).
- Optical aberrations within the eye lens (O'Regan 1994).
- A power difference of two diopters for red and blue light in the eye lens (O'Regan 1994).
- Saccadic eye movements which smear and displace the image.
- Torsion (small cyclorotations around the retinal visual axis (Howard 1982)).

[†]The angle subtended by the blind spot is relatively large—about 4° of visual angle (dva). It is generally assumed that the brain somehow “fills in” the missing information into the region of the blind spot, but according to O'Regan this hypothesis has not been tested seriously—but see Ramachandran (1992) and Section 4.2.7.

- Various small correction movements such as micro-saccades, micro-drifts, small vergence movements and micro-tremors (Zuber, Stark & Cook 1965).
- Further eye, head and body movements which not only move the retina relative to the world, but are also responsible for retinal image slip during fixation.

If anything, this list understates the case. Howard (1982) includes an extended discussion of the various movements, deviations, slips and twists observed in the behaviour of the retina.

Yet despite all of these difficulties, we still seem to have a ‘perfect’ view of the world, an image that is stable and complete. We are not aware of these ongoing processes and imperfections.

Various solutions are proposed. For example, it is suggested (Bridgeman et al. 1994) that information for a fusion could be provided by structural ‘cues’ in the visual world, ocular muscle input, proprioceptive input from extra-ocular muscles and motor system commands.

The displacement of the retinal image is assumed to be accounted for by an *adjustment* provided by extra-retinal signals. Each incoming retinal view is placed in the correct ‘position’ on the internal representation by allowing for any displacement caused by saccadic, eye or body movements. This is illustrated in Figure 4.11. The dotted box indicates the extent of the ‘internal representation’, and the image information provided by the three saccades indicated is positioned correctly using the displacement information. † All of the eye and body movements are somehow added up to produce a correcting adjustment (the Δx , Δy in Figure 4.11) so that the foveal image is placed correctly in the overall internal world image, “guaranteeing a seamless visual percept and the ability to accurately locate objects in our environment” (O’Regan 1992, page

†The actual situation is somewhat more severe than that indicated in Figure 4.11. The foveal area (corresponding to the dotted area in the figure) subtends an angle of only 2° , whereas the full visual field (and hence supposedly the internal image) is purported to be as much as 210° (Miura 1987). The postulated extra-retinal signals must account for all of the changes in the retinal frames across saccades. It would need to compensate for eye movements, head and body movements, variable focus of the eye, movement in targets and so on.

467). The process is termed “trans-saccadic fusion” by O’Regan & Lévy-Schoen (1983).

These traditional views assume that any internal representation has metric properties like those found within the world. They pre-suppose that a picture-like image is retained in the brain. However, there are a number of other problems with this *internal screen* model (O’Regan & Lévy-Schoen 1983, Turvey 1977, Baber 1983, Irwin 1993a). These include:

- A difficulty incorporating depth information into the representation. How would eye convergence, accommodation and focus be taken into account?
- What mechanism could produce the fusion given the extreme difficulty of obtaining a precise registration because of widely differing resolutions, colours, and extent (due to the non-linearity of the retina)?
- The extra-retinal signal must be perfectly accurate or else errors will build up to cause poor registration.
- What mechanism could account for head and body movements?

Most contemporary vision models make the assumption of spatial integration. They assume that information from successive fixation images is somehow brought together

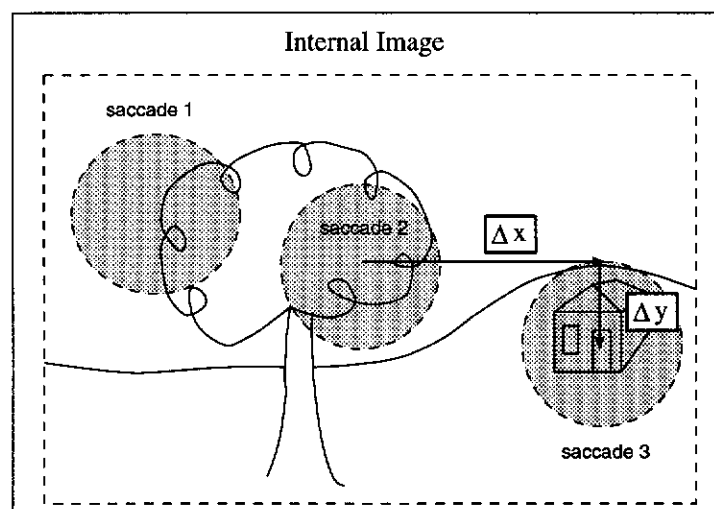


Figure 4.11: Internal Image Representation.

into a metric-preserving internal representation. The representation may be explicit, such as a photograph-like representation in some area in the cortex, or else as a more implicit representation in which higher-order representations are joined in a metric-preserving manner. In many cases of computer vision models, the issue is not even mentioned. These models begin their early processing with a full world scene, which is then segmented to provide the initial primitives for further processing.

We could always reject the photograph-like representation for reasons of parsimony. The visual system seems to go to a lot of effort to tease apart numerous components of the image in the retina, LGN and early striate cortex (as described in the initial sections of this chapter) for it to be put back together again as a replica of the outside world, ready for subsequent processing.

Spatial integration of higher-order representations has no psychological basis, and would appear to be an immense, if not impossible computational task *even if* the appropriate 'extra-retinal signal' were available.

It seems clear that the various vestibulo-oculo reflexes and other extra-ocular neural signals provide general directional sense and orientation, but this is a very different proposition from insisting that they are able to compensate for retina, head and body movements sufficient to ensure a computational spatial registration of images across saccades.

More importantly, there is good experimental evidence that spatial fusion is just not performed (O'Regan & Lévy-Schoen 1983, Turvey 1977, Baber 1983, Irwin 1993a, Irwin 1993a). Thus we are led to reject the notion of *spatial integration* of retinal images across saccades.

4.2.5 The World As Outside Memory

An alternate view to the "internal screen" model described above has been proposed by a number of people (Helmholtz 1925, Hebb 1949, Gibson 1950, Gibson 1968, MacKay 1967, MacKay 1973, Hochberg 1984, O'Regan & Lévy-Schoen 1983, O'Regan 1992, O'Regan 1994) In this view, the brain does not record internal representations (icons)

of objects in the world, but rather uses the world itself as an “outside memory store” (O’Regan 1992, page 464).

O’Regan has conducted a number of critical experiments which cast serious doubts on the “trans-saccadic fusion” model. Figure 4.12 illustrates the setup of one experiment. Each stimuli consisted of two halves, the second half being the same for all three stimuli. The two halves, when superimposed, form the three words at the bottom of Figure 4.12. During each trial, the two halves of each stimulus were presented in sequence, and in the same physical location in space. However, the first was presented before a saccade, with the second presented after the saccade. Over a number of trials with various stimulus durations and delays, subjects were *unable* to ‘fuse’ the two apparently random patterns into a recognisable word (as would be expected if the ‘internal screen’ hypothesis was valid). Other researchers have also confirmed that no fusion appears to take place (for details, see page 469 O’Regan 1992).

Several carefully planned and independent studies have confirmed this finding (Rayner & Pollatsek 1983, Irwin, Yantis & Jonides 1983). These results appear to rule out the strong interpretation of a visual buffer, including the perceptual availability of the result.

Experiments in which line drawings are exchanged during saccades also indicate “that an integrative visual buffer, if it exists, plays little part in the perception of line draw-

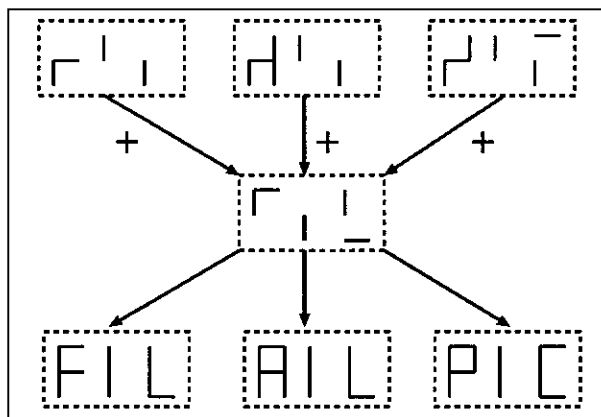


Figure 4.12: Trans-Saccadic Fusion Experiment. (Adapted from Fig. 3. O’Regan and Lévy-Schoen (1983))

ings” Pollatsek et al. (1984). Their conclusion is that “it seems unlikely that an integrative visual buffer is ever used to combine information across saccades.”

The very fact that scene changes at the movies do not cause momentary confusion also supports the argument against spatial fusion. If spatial integration was required for scene understanding and recognition, we would need to scan the screen for some moments to take in the full contents of a new scene.

A tactile analogy is given by MacKay (1967, 1973). When we explore the tactile sensations of an object with our eyes closed, we do not need to obtain a *full* and detailed tactile map of the object to identify it. A single touch or a series of cues may be sufficient. Further, the surface details corresponding to the spaces between our fingers do not necessarily need to be ‘filled in’.

It is the process of active *exploration* of the tactile sensations and the recall of previously learned sensations that enables us to recognise, and thus be able to name, the object. If an initial sensation is insufficient for recognition, further exploratory touches at other locations of ‘expected’ high discriminatory value will be undertaken until sufficient evidence is obtained. We recognise the object through a series of tactile experiences rather than first obtaining a full tactile description. In this case, there is no thought of postulating a tactile mechanism which compensates for the movements of the hand as it ‘fills-in’ the tactile representation of the object—the equivalent of the extra-retinal signal in vision.

And so it is with visual and other perceptual sensations. As stated by O’Regan (1992, page 472), “... ‘perception’ is *getting to know* or *verifying* the sensations caused by possible actions.”

There is no need for a full metric (internal replica) representation of the object—all that needs to be stored is a series of linked (temporally connected) cues. The object is *in the world* for further sampling if required. Other contextual associations will also tend to support or inhibit the perceptual cues that are currently being experienced. Note that we are not suggesting that *no* representation is stored, but that the ‘representation’ is not a spatially fused iconic form. Rather, it is a temporal sequences of sensations.

Awareness is an active process. If we are unaware of something, then it is not part of our current conscious thinking. It is only if context or other urgings (such as hunger) cause us to actively seek out some object or phenomena, or some sensation (such as motion) brings an object to our attention, that we become aware of it.

As put by O'Regan (1992, page 473)

Since "seeing" involves both *interrogation* of the visual field, and also *apprehension* or *integration* or *comprehension* within the current mental framework, one would predict that a person would fail to see something either (a) if he or she does not interrogate or wonder about the appropriate aspect of the visual field or (b) if he or she is unable to integrate the obtained sensations into his or her mental framework. In particular, even if you are directing your eyes directly at something, unless (a) you are (at least unconsciously) wondering about it, and (b) you are able to apprehend it, you will not have the impression of "seeing" it.

A number of experiments have been conducted which demonstrate this point. For example, Haines (1991) describes the case of pilots using a 'heads up display' in a flight simulator who would not see a perfectly visible air-plane parked in the middle of the runway, and would drive right through it. Neisser & Becklen (1975) studied subjects who were asked to view two superimposed action sequences on a screen, and found that they would only see what they were attending to.

In a series of experiments in which items on a screen were changed during a subject's saccades, McConkie (1979, 1990) was able to show that even obvious and large objects could be removed, changed in colour, or shifted without the subject being aware of the modification. Unless there was some cognitive reason for the subject to attend to an object (such as watching for a change), its presence or absence in the scene was not noted.

A number of careful experiments involving the reading of text on a screen provides additional support. The screen viewed by subjects contained masked text, (for example, all words X'd out), except for those words within a moving window which is tied to

the subjects foveal view. This is illustrated in Figure 4.13. As readers move their eyes along the line of text, the window moves with the eyes so that actual text is always at the fovea, while the window is surrounded on both sides by masked text. The changes occur during each saccade. However, the subject is completely unaware of the masking, and maintains that the whole text remains unchanged and available (Rayner et al. 1978, Erlich & Rayner 1981, Pollatsek & Rayner 1989, Morris, Rayner & Pollatsek 1990, O'Regan 1990). We discuss reading further in Section 4.8.9.

The perception that we have visual access to all objects in our field of view *at once* is an illusion. We really only see what we are attending to at the time, other than some diffuse sense of other 'objects' being in the periphery. If we wonder about what a particular 'objects' might be, we then direct our conscious attention towards it by either saccading to it and thus bring it into full vision, or relying upon a previous memory of the object which can be cued by some imprecise sensations. For example, as I am writing this paragraph, I can direct my attention to the right periphery of my vision and obtain a blue sensation. I remember that I placed a blue book on my desk a few moments ago in roughly that position, and a saccade in that direction confirms my recall. Other vague sensations in the periphery are ignored unless my flow of attention is directed towards them.

I can actively explore the environment to fill in my expectations and inquiry, or my attention can be directed to a location because of some prompting sensation, such as

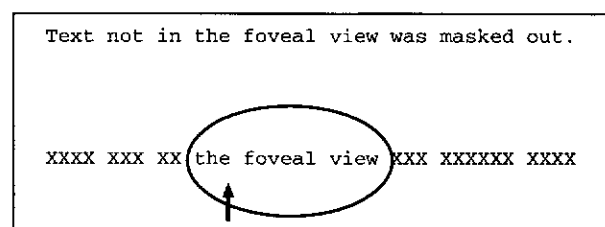


Figure 4.13: Perceptual Span in Reading. The oval indicates the 'perceptual span', and the arrow the centre of the foveal view. In a typical (English reading) subject, the perceptual span is about 17-18 characters, with about 2-3 characters to the left of fixation and about 15 characters to the right.

visual motion or a sound. Seeing is the action of interrogating the environment.

Visual perception is such an intensely rich sensation of *total* external presence that most people find it difficult to accept that it is not as it appears. The illusion results from the fact that we have *some* sensation over a very wide visual field, detailed information only in a very small foveal area, and an attentional mechanism that can range across the whole area very quickly. It is the flow of attention (that is, the temporal sequences of foveal inputs) that is the key to understanding visual recognition.

Churchland et al. (1994) propose four reasons for our strong perception of a whole-scene visual representation:

- the ability to repeatedly *visit* stimuli in the scene,
- a short-term *semantic memory* (of a few seconds duration) that maintains a general ‘feeling’ of the overall context without creating and maintaining the point-by-point detail,
- the *objectification* of our sensory perceptions so that they are represented as being about some object in space,
- a *predictive* element within pattern recognition based on context.

The rejection of a metric basis for vision is a difficult one. Our Western cultural heritage is very heavily influenced by metric-preserving representations such as perspective in paintings, maps, and photography. Further, metric-preserving metaphors are strong in Western thought as opposed to some other cultures for whom space is not a linear Cartesian geometry. †

The visual mechanisms proposed here are consistent with this view of the world as a “memory buffer” that is accessed by visual behaviours (O’Regan 1994, O’Regan 1992, O’Regan & Lévy-Schoen 1983). The relative stability of the world allows it to be used as an outside memory, thus obviating much of the need for detailed internal representations.

†See, for example, a discussion by Whorf (1971) on the different treatment of space by the Hopi language. Talmy (1983) also discusses how languages structure space.

4.2.6 Semantic Integration

Given that spatial fusion is unlikely, another possibility is that the visual system operates at a more abstract level. O'Regan & Lévy-Schoen (1983) have put forward the hypothesis of a level of representation more general than the analog visual code available on the retina as a possible explanation of the perceived phenomenal stability and continuity of the visual world. Their suggestion is that our mental representation of a visual scene is essentially semantic rather than 'photographic-like'.

The representation is based on labels ('tags' is their term). The image of a chair is coded as the label 'chair', with spatial relational terms (such as 'near', 'far', 'in front of', and so on) providing enough relative-position clues to guide eye movements should further spatial information need to be extracted from the visual field.

This semantic representation has the obvious advantage of not requiring complex shifting or aligning of successive 'snapshots', but has serious epistemological problems.

This use of semantic labels takes no account of the source of the labels, especially in relation to animal and child visual recognition. If visual objects need to be labelled with semantic terms, what do children use before they learn the label (or name or referent) of a particular object? Do they use a place marker (if so why bother with a label subsequently) or are they unable to perceive the item until they have learned the referent? Do things pop into view when we name them?

Or is the label for each object innate? This implies that some form of object pre-recognition is found in the brain for all objects that are not only recognisable now, but also the names for all new technologies yet to be discovered—clearly another epistemological dead-end.

As we will show later with the ABC model, referents certainly do take a part in conceptualisation, but they play no fundamental part in perceptual conceptualisation or recognition. The acts of recognition and semantic naming are separated. Just as with the Evening Star/Morning Star problem of philosophy, referent and concept must be separated. This applies as much to objects ('chair') as it does to relationships ('behind', 'under', 'above', etc.). The ABC model does enable a linkage between the *perceptual*

concept (the neural attractors formed by the perception of an object) and the associated *referent attractors* (the neural attractor formed by the perceptual learning of the referent—the sound of the referent or the visual perception of the name of the object in written form), but the process proposed for visual recognition does not involve or require the labelling of objects.

4.2.7 Fill-in of Blind Spot and Scotomas

It has long been known that we are not visually aware of the blind spot on each of our eyes, the area within the retina where the optic nerve exits the retina for the brain. It has usually been held that some form of *fill-in* of the surrounding retinal array is substituted for the missing area.

In an experiment by Gilbert & Wiesel (1991), a small patch of the retina was destroyed and recordings made from the area of visual cortex to which this patch would normally project. As expected, the cells were initially silent, but within a few minutes, the same cells now had receptive fields which extended beyond the zone of the lesion; that is, they responded to visual stimuli outside the scotoma.

Gilbert & Wiesel obtained the same results with an artificial scotoma, a gray, square area of constant luminance surrounded by random noise. On steady fixation, the artificial scotoma was ‘filled-in’ by the surrounding noise. Cells which initially had receptive fields within the gray square now had much larger receptive fields, and included regions outside the square. Ramachandran (1992) has shown similar results with artificial scotomas, and has shown that these (as well as the blind spot) are indeed ‘filled in’ with surrounding sensory information. Using various bars and other shapes, Ramachandran found that it takes some 6 to 7 seconds to ‘complete’ the ‘filling-in’ of the figure. He further suggests that this fill-in occurs at an early stage in visual processing, and that it does not result from retinal receptor fatigue.

As well as the results regarding the fill-in, these experiments imply that the receptive fields are not fixed anatomical entities formed by retinal receptors funnelling information on to single cells in the cortex, but are rather dynamic in their behaviour.

4.2.8 Global Influences on Perception

There are several experiments which illustrate a global or contextual influence on visual perception, in contrast to the traditional view of most psychologists and computer vision researchers who generally assume that perception occurs in stages. There are very few psychophysical experiments that provide direct evidence for the traditional view, and the experiments discussed in this section indicate that the so-called early vision modules are not autonomous and do interact significantly with each other. They further suggest that higher-order processes *do* affect our perceptions.

bistable motion perception

Figure 4.14 illustrates an experiment which shows the role of top-down influences (such as attention) within perception. A set of four dots of light are alternately illuminated across the diagonal. The black dots marked A in the figure are first illuminated, then they are turned off and the open circles marked B are illuminated, and so on in a continuous cycle (Ramachandran & Anstis 1983, page 84). A number of these four-dot constructs are maintained on a screen, and the oscillating lights in each is seen as a movement of light between pairs of dots.

The striking thing is that subjects do not see a mixture of horizontal and vertical movements, but one or the other (illustrated by the arrows in the two parts of the figure). The perception is one of global apparent motion. The display is essentially bistable just as with the Necker cube. Also, like the Necker cube, subjects can focus their attention to change the direction of oscillation, but at higher oscillation rates this is very difficult.

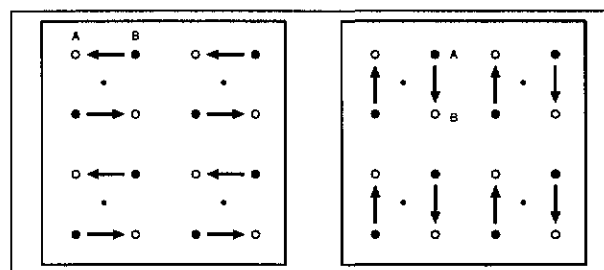


Figure 4.14: Bistable Motion Perception.

This result strongly suggests that global or higher-level considerations do interact with so-called low-level operations such as motion perception.

“Necker cubes and other bistable figures are often used to illustrate the point that perception is really a hypothesis on the state of affairs in the world rather than a passive response to sensory stimuli” (Ramachandran & Anstis 1983, page 84).

The subjects perception of the cube changes dramatically as the brain switches between alternate *interpretations* of the figure.

virtual occlusion

A variation on the previous experiment is shown in Figures 4.15. In these experiments, a shaded square is found on the right-hand side. In Figure 4.15 (a) all group A dots blink on and off, but only the upper and lower B dots alternate off and on. The perception is that the middle A dot moves behind the “virtual” occluder.

However if A contains only one dot, as in Figure 4.15 (b), the perception is not found. What is seen is a single dot blinking on and off, with a square to the right. The surrounding subjective motion is required to provide the perception.

cross-modal interactions

Using the same single dot and shaded square, with the addition of a tone sounded in each ear simultaneously with the blinking of A and the virtual blinking of B, (as indicated in Figure 4.15 (c)), subjects do indeed see the single dot move to the right

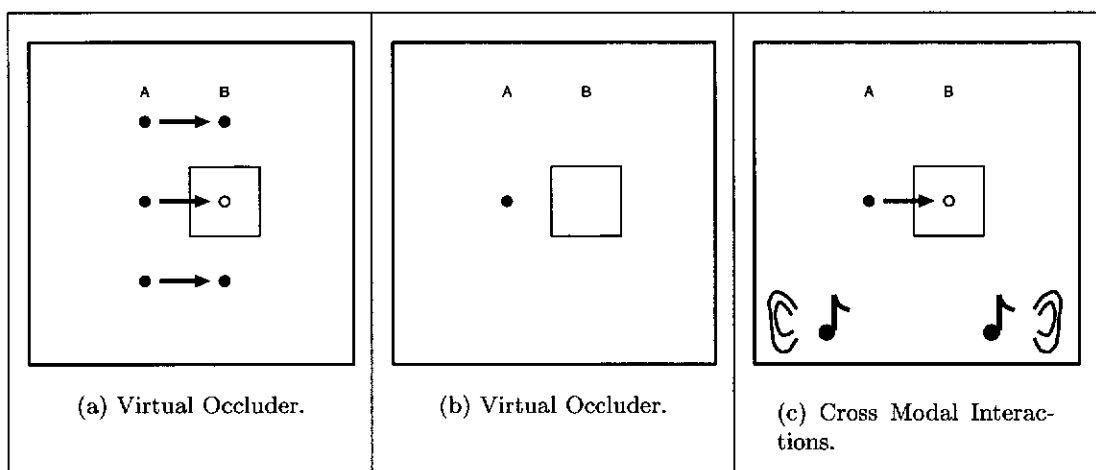


Figure 4.15: Virtual Occlusion.

behind the occluder (Churchland et al. 1994, page 31).

The integration of auditory and visual information “pulls” the dot in the direction of the sound movement. A similar (but weaker) effect can be induced if the light blinking is accompanied by a corresponding left-right vibration stimulation to the hands.

This experiment is a very convincing demonstration of the interactive and integrated nature of vision, as opposed to the traditional straightforward ‘image interpretation’ view. While integration of auditory and visual information is expected at some stage in the brain, this demonstration shows that this integration occurs in what was considered to be early visual processing.

motion correspondence and the role of image segmentation

Figure 4.16 illustrates an experiment in which four pacmen and four full circles are alternatively displayed on the left and right of a screen. The effect as perceived by subjects is that of a (virtual) foreground square shifting left and right, alternatively occluding two sets of circles. Further, if a regular grid of dots is superimposed in the screen, the dots within the subjective square also appear to move with the illusory surface, while the dots outside the square appear to remain stationary. Yapping pacmen are not seen (Ramachandran 1985).

This experiment indicates that global interpretations can dominate local constraints, and that segmentation and the correspondence problem are not performed purely at a

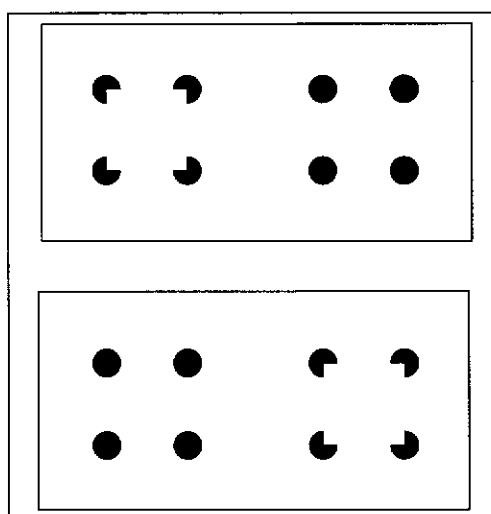


Figure 4.16: Occluding Square.

local level, early in the so-called hierarchy.

shape-from-shading

Ramachandran (1992) cites a number of experiments in which screen images of shaded 'spheres' are perceived as being either concave or convex. The perception in many cases may be reversed by an act of will, and there seems to be a built-in assumption that the light source is above.

If an image contains a number of 'spheres', those with shading at the top will appear to be concave, while those with shading at the bottom will appear to be convex. Turning the image upside-down will reverse the perception for individual spheres. Tilting the head in conjunction with the image retains the shape perception, but turning the image by 90° relative to the head so that the shading is from the sides (no matter what the orientation of the head itself) makes it harder to differentiate the spheres into concave and convex shapes.

Thus the assumption of overhead lighting and shape-from-shading is primarily, if not exclusively, dependent upon retinal rather than world coordinates, a somewhat counterintuitive result since it implies that when you tilt your head the visual system 'assumes' that the sun is stuck to your head.

Other experiments involve concave masks which are lit by sources at various directions relative to the mask. The perception is always one of a convex face, even when the illumination is from below. This perceptual persistence shows a strong top-down effect in a supposedly low-level task, namely shape-from-shading.

stereoscopic depth perception

Stereo fusion is usually regarded as an early vision task, yet an experiment described by Ramachandran (1986) indicates that stereo vision also makes use of top-down global information. An image is stereoscopically fused on the basis of *subjective contours* rather than actual low-level details.

figure-ground

Experiments by Peterson & Gibson (1991) indicate that figure-ground separation does not precede shape recognition as postulated in the traditional models.

4.2.9 Neuroanatomy and Neurophysiology

The experiments and observations in this section describe more the hardware implications and connections of any cognitive model of human and animal vision.

backprojections

The structure of mammalian cortex shows multiple forward and backward projections of axons at several levels. For example, in the monkey cortex, the various forward axon projections [†] are equivalent to, or outnumbered by, back projections (Van Essen & Maunsell 1983, Van Essen & Anderson 1989). These backprojections in most cases appear not to be merely reciprocating feedforward connections, but are more widely distributed and include distribution to some areas from which they did not receive a forward projection.

Rockland (1992b, 1992a) and Rockland et al. (1992) describe the connections between areas of visual cortex of the monkey as shown in Figure 4.17.

motor structure connections

There are strong connections to areas associated with motor actions. For example, the cat superior colliculus (SC) includes projections from twenty-five cortical areas (Harting, Updyke & Lieshout 1992). The superior colliculus is thought to have an important role in directing saccadic eye movements (as well as ear movements in animals with orientable ears).

[†]Forward in the sense that the axons project from regions which are closer (in synaptic distance) to the sensory periphery to regions more synaptically distant; for example, projections from V2 to V4.

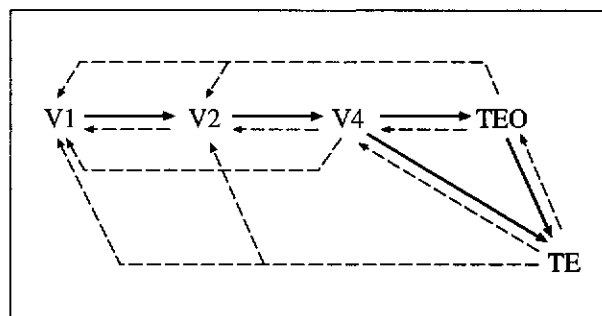


Figure 4.17: Feedforward (solid lines) and Backprojection Connections (dashed lines) in the Monkey. (From Rockland et al. (1992)).

Further, nearly every area of mammalian cortex has some projections to the striatum (basal ganglia), an area thought to have an important role in integrating sequences of movements and voluntary eye movements.

As well, neurons which are sensitive to eye position have been found in the LNG (Lal & Friedlander 1989), area V1 (Trotter, Celebrini, Stricanne, Thorpe & Imbert 1992, Weyland & Malpeli 1989), and area V3 (Galletti & Battaglini 1989).

illusory contours

A number of results show that neurons in the visual cortex of monkeys respond to illusory contours. † For example, von der Heydt, Peterhans & Baumgartner (1984) report that neurons in visual area V2 of the macaque respond to illusory contours, and Grosf, Shapley & Hawken (1992) have shown that some orientation selective cells in V1 respond to a class of illusory contours. The detection of illusory contours depends on some interpolation across a span of the visual field, suggesting higher level operations are backprojected to lower levels.

cross-modal interactions

Most current theories of vision only deal with the dominant exteroceptive function of vision where we perceive the properties of the visual world, and fail to take account of the other perceptual modalities. That this interaction may be important is demonstrated in the phenomenon ofvection, the illusory sensation of self-motion induced by movements in peripheral vision (Brandt, Dichgans & Koenig 1973). While very little is known of the linkage between the exteroceptive and proprioceptive functions of vision, its importance should not be underestimated.

Insufficient work has been done on the interaction of the visual and auditory modalities. The experiment described on page 199 and illustrated in Figure 4.15 is one exception which indicates a close linkage between these two modalities.

There are a number of other experimental results which show a strong linkage between various modalities. For example, the responses of cells in V4 to a visual stimulus can be modified by somatosensory stimuli (Maunsell, Sclar, Nealey & DePriest 1991), and a similar task-dependent modification for cells in somatosensory cortex (area

†The question must be asked as to whether these contours are actually illusory. In many cases, spatial frequency filters *will* record them as being nearly equivalent to actual contours. They are only *illusory* given higher-level (learned) knowledge of the actual scene and objects involved.

S1) has been found by Fuster (1990).

4.3 Evolutionary Considerations

The experimental results discussed in previous sections suggest that the process of vision is not a simple isolated and hierarchical stream from sensations to perception, but that other sensory modalities and motor systems play a significant role in what is literally seen.

As Churchland et al. (1994) point out, one must give some prominence to the biological ‘purpose’ of vision in our considerations of its mechanisms. Improved motor control is surely the evolutionary rationale of vision—to enhance the creatures chances of survival and reproduction through being able to visually monitor the world around it, allowing it to better seek out appropriate nourishment, to better avoid predators and other dangers, and to enhance its chances of finding an appropriate mate.

What is important to a creature at any given moment will depend upon the context, the immediate environment and the creatures current state and needs. Rather than taking in a full iconic view of the world as proposed in current vision models, it would seem that only immediately relevant information is explicitly required by the creature. The creature will tend to foveate to items of interest, and although unattended objects are represented in some minimal fashion (sufficient to guide attentional shifts and eye movements), they are not literally seen in the sense of “visually experienced”.

Interactive vision is exploratory and predictive—visual learning allows an animal to predict what will happen in the future. Further, recognition can be faster and more accurate if the animal can make exploratory movements, particularly of its perceptual apparatus, such as whiskers, ears, and eyes. New dangers and opportunities must be learned and incorporated into existing behavioural patterns.

The traditional view of vision assumes that the connection to the motor system is made only after the scene is fully elaborated. The idea is that the decision centres make a decision about what to do on the basis of the best and most complete representation of the external world. However, many situations may require action long before a

complete analysis of the visual terrain can be made—motor action may be required on the basis of preliminary and minimal analysis.

Churchland et al. (1994) remind us that “natural selection and reinforcement learning share a certain scientific appeal; to wit, neither presupposes an intelligent humunculus, an omniscient designer, or a miraculous force—both are naturalistic, as opposed to super-naturalistic. They also share reductionism agendas. Thus, as a macro-level phenomenon, reinforcement learning behaviour is potentially explainable in terms of micro-mechanisms at the neural level.”

Using active reinforcement learning, the brain is able to build a network more suitable to the current environment, including predictive mechanisms to direct attention to what is worth looking at given one’s interests, and a close linkage between perception and appropriate behaviour. Sophisticated visual perception evolved, not as an end in itself, but in the service of better motor control.

4.4 Active Vision

Most current and previous vision models use static recognition to identify component parts of a ‘scene’. The system is first trained with objects to be recognised. The initial stage of this might be the extraction of parts of each object which are used, in conjunction, to subsequently recognise the object. Objects may either be recognised singly (object recognition) or as one of several objects (scene recognition).

Early models used object attributes only, whereas more recent models include spatial relationships between the object parts, and between the objects themselves (see, for example, Dillon 1996). Most, however, attempt to recognise the objects from information taken *at one instant of time*.

The supposed task of most current vision systems is to find a description of the three-dimensional spatiotemporal world by transforming static two-dimensional data. The goal of such a system is to “infer 3-D surfaces, volumes, boundaries, shadows, occlusion, depth, colour, motion” (Shapiro 1987, page 389).

The approach taken in this thesis is that visual recognition is a mechanism that involves temporal sequence learning and recognition. Rather than recognition occurring at one time instance, it is spread over several time intervals, and the recognition occurs as a result of the learning and subsequent identification of temporal sequences. Temporal learning is usually described as applying to one-dimensional input streams. Although the visual domain is two-dimensional, [†] the same method of temporal learning can be used to solve the visual recognition problem.

A number of other researchers have also taken a more behavioural approach to vision. Their view is that the visual system is “more readily understood in the context of the visual behaviours that the system is engaged in, and that these behaviours may not require elaborate categorical representations of the 3D world” (Ballard 1991, page 58).

The active vision paradigm (Blake & Yuille 1992, Bajcsy 1986, Bajcsy 1988, Ballard 1991, Giefing, Janßen & Mallot 1992) is based on the ability of the perceiving system to purposively change—for example, to change the fixation point, or the focal length, or even the location of the system within the world. Rather than attempting to squeeze every bit of information out of every image, the active system adapts its behaviour in order to obtain information that is important at the moment. The paradigm is concerned with a continuous flow of images rather than a succession of static images. [‡]

Vision is far from static—to quote from Bridgeman et al. (1994, page 247):

Our eyes are almost constantly moving. Body and head movements change the position of the eyes relative to the world. Even subjects trying to remain as steady as possible while using normal postural supports to keep the head in place show appreciable head movement. And even during steady fixation with the head immobilised, a variety of small eye movements (e.g., microsaccades, drifts, and tremors) and larger ones (e.g., saccades, pursuits,

[†]We ignore, for the moment, the effect of stereoscopic vision in providing information for the third dimension, and concentrate instead on the 2-dimensional visual information that impinges on, say, a single retina. Further, our contention is that stereo fusion is a learned process.

[‡]The question of active attention and gaze control is not considered here—see, for example, Olshausen & Koch (1995). Also see Treisman & Gelade (1980), Brown (1990), and Ahmad & Omohundro (1991).

and vergence movements) constantly change the position of the eyes relative to the world.

Fischer & Weber (1993, page 553) summarise the observations that need to be explained in any theory of vision: †

Visual perception and cognition, movements of the retinal image, and eye movements are so closely related that one wonders how one field can be studied without taking into account the others. The facts are straight forward:

(i) During natural viewing conditions, a normal adult subject makes 3-5 saccades in a second, separated by periods of 200-300 msec during which the eyes do not make large or fast movements. These periods are usually called “fixations” but this terminology is avoided [. . .] because [. . .] periods of no eye movements are not necessarily periods of active and attentive fixation, but can as well be periods where the eyes simply do not move.

(ii) If the retinal image as a whole is prevented from moving (by successful voluntary attempts not to move the eyes or by technical means), vision is blurred rapidly, and the perception of the retinal image eventually fades away completely within 10 seconds.

(iii) The highly inhomogeneous structure of the primate retina, with an extremely high density of receptor and ganglion cells in the centre, a specialised fovea, and a rapid decline of the cell densities to the periphery, makes it almost impossible to have a homogeneous and simultaneous percept of the total visual field without somehow moving the fovea to different positions and somehow acquiring and integrating information from these successive “looks”. The existence of a fovea requires both eye movements and periods of fixation, i.e. the active suppression of saccadic eye movements.

(iv) As a result of the complicated interaction between afferent, central,

†This rather long quote is included here as it sums up the important points extremely well.

and efferent neural processes we perceive a complete and stable visual field, which can serve as a frame within which we see motion and within which we move ourselves or parts of our body.

Vision is a real-time process, not the passive process postulated by traditional vision systems. Typical short fixation times in the human visual system are about 0.25 seconds, and cortical neurons typically fire at rates of 10 spikes per second, giving 2.5 spikes per fixation (Ballard 1991, page 71). Input to the visual system, then, is a series of short duration ‘views’.

4.5 Scanpaths

It is well documented that fixations tend to cluster around places with high gradients of change in the luminance distribution (Mackworth & Morandi 1967). Moreover, the sequence of saccadic scan patterns for humans is both regular and idiosyncratic (Noton & Stark 1971b, Noton & Stark 1971a). As stated by Noton & Stark (1971a, page 35):

... every person has a characteristic way of looking at an object that is familiar to him. For each object he has a preferred path that his eyes tend to follow when he inspects or recognises the object.

The eyes seem to follow fairly regular pathways, visiting the features of an object in a somewhat cyclic manner, rather than criss-crossing the object at random (Noton & Stark 1971a, page 37). For example, in one experiment performed by Yarbus (1967) subjects were asked to view a photograph of a bust of Queen Nefertiti. Recordings of their saccades showed that fixation on a feature, say her eye, was usually followed by a fixation on the *same next feature*, such as her mouth. The saccade path showed regularities, moving from one ‘feature’ to the next in a consistent sequence.

Work done by Noton & Stark (1971a) found similar results, and showed that “each scan path was characteristic of a given subject viewing a given picture. A subject had a different scan path for every picture he viewed, and for a given picture each subject had a different scan path.”

Noton & Stark (1971a) proposed a mechanism involving a 'feature ring'.[†]

... as a subject views an object for the first time and becomes familiar with it he scans it with his eyes and develops a *scanpath* for it. During this time he lays down the memory traces of the feature ring, which records both the sensory activity and the motor activity. When he subsequently encounters the same object again, he recognises it by matching it with the feature ring, which is its internal representation in his memory. Matching consists in verifying the successive features and carrying out the intervening eye movements, as directed by the feature ring.

Our proposal is essentially the same, except that the 'feature ring' is replaced by temporal sequences which are stored on the SOM surfaces of a SOMA network. Visual scanning via saccades is then seen to be similar to the more *symbolic* processing of, say, characters or words, described in Chapter 2 and Appendix B. Each fixation point will provide a vector (via the visual filters) to be learned, just as each letter provided a vector. These vectors are then passed to the temporal learning system in a consistent sequence, in much the same way as with language.

If the theory is correct, then we should expect that the saccade trace laid down during learning of a new object should be reproduced when that same object is being recognised,[‡] and this expectation was confirmed by Noton & Stark (1971a, page 40). They found that in about 65% of cases, subjects reproduced the same learned scanpath when subsequently recognising an object.[§]

The mere existence of such regular scan paths suggests a process of learning. The finding that different subjects had different scan paths for a given picture, suggests that the scan paths are not the result of peripheral feature detectors that control eye movements independent of the recognition process. Any such detectors might be expected to operate similarly for all subjects. Conversely, the observation that a given subject had different scan paths for different pictures suggests that the scan paths do

[†]Their feature ring was essentially a circular linked list.

[‡]Assuming similar contextual inputs and a similar purpose for the recognition.

[§]The 65% required a strictly serial ordering (their feature ring). In our proposal of learned temporal sequences we are not restricted to the linear sequences of a ring.

not result from some fixed habit of eye movements. †

As well, the fixed scan paths suggest that the eye-movement motor components involved in perception are not independent movements to move a pattern over the retina (in order to fill in the details left out by a limited fovea), but are rather an integral part of the memory sequences on which recognition is based.

Figure 4.18 (a) shows a typical scanpath found by Noton & Stark in which subjects viewed an adaption of a line drawing by Paul Klee. Figure 4.18 (b) shows an idealised scanpath for this particular subject for this particular picture.

Groner et al. (1984) explored the idea of ‘local scanpaths’ in the sense of Stark et al. (that is, in reflecting consistent patterns of successive fixations), and ‘global scanpaths’ (reflecting the distribution of fixations over a larger time scale irrespective of their immediate succession). In global scanpaths, fixations don’t follow each other, but rather reflect a tendency to concentrate somewhere in the course of the exploration process, representing more of a ‘search’ process to satisfy an expectation of the subject. They

†The actual mechanism of selecting ‘features’ as saccade points is not discussed here. There is some evidence that the mechanism may (in part) be culturally determined (Segall et al. 1966, Segall et al. 1990). See also Arbib (1995b), Treisman & Gelade (1980), Brown (1990) and Ahmad & Omohundro (1991) for discussions of active attention and gaze control.

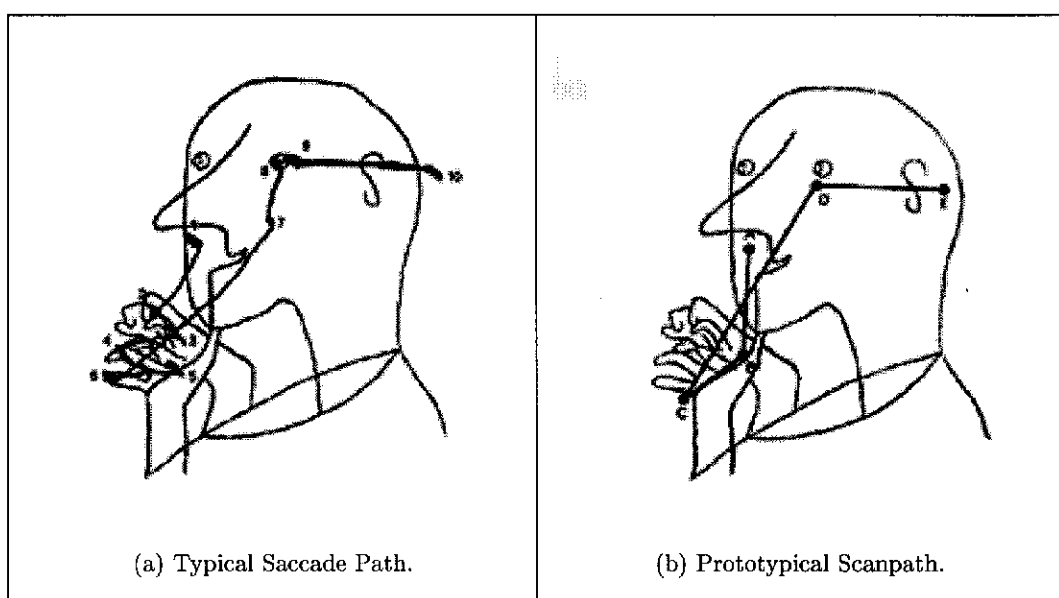


Figure 4.18: Saccade Scanpaths.

found the possibility of discriminating individual styles in face scanning by considering only the relative frequency of fixation triplets. Walker-Smith, Gale & Findlay (1977), in studying face recognition, also observed short sequences (two or three saccades) that were common to both recognition and examination phases.

Groner & Menz (1985) found that while most subjects utilised both local and global strategies, there were some subjects who used just one or the other strategy, and even some subjects who used neither strategy.

In their original articles, Noton and Stark claimed that in 65% of cases the same sequence occurred in both memorisation and recognition of a pattern. However, they gave no quantitative criteria for recognising a scanpath. In another study examining the role of scanpaths in the recognition of random shapes, Locher & Nodine (1974) applied a more precise criteria. They found the presence of scan paths in over half of all eye-movements regardless of the shape complexity.

To provide further quantitative measurements of scan paths, Hacısalihzade, Stark & Allen (1992) have modelled sequences of visual fixations as Markov processes, with the sequences abstracted as character strings. †

Stark & Choi (1996) also employ a string editing technique in which regions of (2D) space are quantised and allocated a character label. All fixations falling within each area are recorded as having the label of that area. Statistical techniques may then be used to compare strings of learned and recognition sequences. The technique is very useful in defining similarity and dissimilarity between eye movement patterns. The results are impressive and confirm the presence of regular, learned scanpath sequences in eye movements, and a high degree of similarity in eye movements between the initial viewing period and subsequent recognition.

Other results indicate that for an object that can be taken in with a single saccade, the subject's attention moves around the picture but the fixation point remains fairly steady near the centre of the picture. This process involves internal shifts of attention

†Note that this is essentially the approach taken using the ABC model, with self-organising and recurrent nets used rather than Markov processes, and with vectors obtained from local filters instead of string characters.

with little or no eye movement. We have not considered this supposed internal attention component in detail here—see Sperling (1960), Olshausen & Koch (1995), or Posner, Cohen & Rafal (1982) for a review.

One potential objection to scanpath theory is the well-known phenomenon of instantaneous recognition. Simple visual scenes and familiar faces can usually be recognised in one glance. Further, information about facial expressions can be identified from a 20 msec view as reliably as from much longer exposures (Viviani 1990).

4.5.1 Visual Imagery Scanpaths

Recent experiments have also shown that visual imagery also exhibits eye movements consistent with scanpaths (Brandt & Stark 1997, Stark & Choi 1996). Subjects were allowed to view a picture of a checkerboard-like grid for a certain period, during which time their exploratory eye movements were recorded. They were then asked to visualise that pattern for several seconds while looking at a blank screen. String editing analysis confirms a high degree of similarity in eye movements between the viewing period and the imagery period. Their conclusion is that eye movements during imagery are not random, but seem to relate to the content of the visualised image, and that mental imagery uses mechanisms similar to vision.

There is also support for the concept of a *shared representational medium* for imagery and perception from other fields. For example, Farah, Peronnet, Gonon & Giard (1988) relate supporting electro-physiological studies in which event-related potentials are recorded during both imagery and viewing, while Goldenberg, Podreka, Steiner, Willmes, Suess & Deecke (1989) report on Single Photon Emission Computer Tomography experiments which relate regional changes in cerebral blood flows.

The eye movements recorded during imagery are distinct in certain characteristics from eye movements during viewing—longer average fixation periods and smaller amplitudes. However, there was a high degree of correlation between the visual sequences and the subsequent imagery sequences. Given that there was no external world picture to drive or influence the eye movements, Brandt & Stark (1997) conclude that the constrained

scanpaths are generated from an internal cognitive model.

4.6 ABC Temporal Experiments and Vision

It has been the view of some researchers that the role of saccades is to *fill in the details* due to the limited size of the fovea. The view presented here suggests that the role of saccades is to provide the temporal sequences for the visual system to learn and to recognise.

Rather than the task of vision being the construction of a detailed internal representation of the world, in this section (and in the ABC model) we propose that the recognition process of vision is the retracing through previously learned visual sequences (via saccades) until an attractor is reached that ‘identifies’ the object.

The model of early vision proposed here uses neural structures that are similar to those proposed in Chapter 2 and Appendix B for one-dimensional sequence learning. For vision, however, the domain is two-dimensional—the system is required to learn not only the input sequences, but also a 2D scan pattern to follow in order to ‘parse’ and hence ‘recognise’ an object.

The early vision system builds retinotopically organised maps of environmental features such as spatial frequencies, motion and colour. Temporal sequences of these feature vectors are learned and provide the means of recognising an object. The recognition information is obtained in the local context in which objects appear.

The eyes saccade over the objects in an image of the world, touring around its coordinates in some manner. At each fixation point, various ‘filters’[†] extract information which corresponds to the neural activity associated with the ‘view’ at that instant in time. The retinal information is transformed via receptive fields, and separated into various sub-modalities (spatial frequencies, colour and motion), until several distinct vector representations are available at areas V1 and V2. The learning of the sequences of vectors occurs separately in areas V3, V4, and V5 (and then subsequent areas). The temporal sequence of these restricted ‘part-image’ vectors is learned by the system and

[†]See Appendix C for a full description of the visual filters proposed for the ABC model.

associated with the appropriate world object.

In learning (and reproducing) temporal sequences of vectors, the proposed system is consistent with, and provides a mechanism for the scanpath theory of Noton & Stark (1971a). Thus, rather than the term scanpath we will use scan-sequence. While a major portion of the vector corresponds to the components of the scene at the fovea, this is not exclusively the case. The larger receptive fields of the periphery are also included.

Recognition is not separate from the learning process. That is, as the eye saccades over an object, the incoming sequence of vectors will be 'recognised' and some behaviour/action initiated. The behaviour could involve the production (speech) of a verbal 'label' that has been previously associated with the object—a behavioural demonstration of recognition.

Notice that this method provides a mechanism for *relational learning*. Spatially adjacent parts of an object will tend to be adjacent in the temporal sequence of inputs—spatial relationships become temporal relationships through saccading. †

Consider an experiment which attempts to recognise some simple regular geometrical shapes—an equilateral triangle, a square and a circle. To keep the exercise small, we will use only three colours for each object—red, green and blue, with a white background, and only one standard size for each shape.

We first need to consider how the external world will be converted into an appropriate vector for learning by the model. Consider a gross simplification of the Wilson Modified Line-Element Model (Wilson & Gelb 1984) that is used in the full ABC simulation model and discussed in some detail in Appendix C.

If a mythical creature inhabited a world of these geometrical figures, and some shapes were good to eat while others had to be avoided, then it is possible that the creature might have evolved a series of spatial frequency filters similar to those shown in Figure 4.19.

At the centre of the creature's retina, the light sensitive cones are interconnected so that

†Relational learning is discussed further in Section 4.8.1.

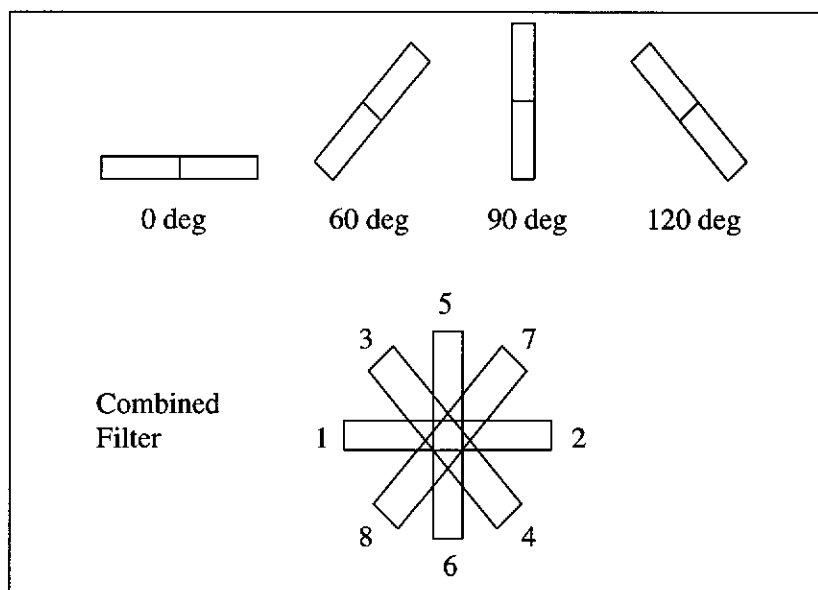


Figure 4.19: Simple Spatial Filters.

it can detect various lines and colours in linear ‘filters’ which subtends certain angles to the horizon as indicated in Figure 4.19. The creature does not have a 30° filter, or a 150° filter as they are not required in *this* world. Whenever the creature looks at an object, the light impinging on these filters sends a signal to the creature’s visual system to enable it to recognise the shape. Each filter provides some data to be included in an input vector that will form the input to a SOMA recognition system. Note that these fictitious spatial filters are double sided; that is, each side of each filter supplies one vector element. †

The order in which the vector elements are allocated is indicated by the numbers on the combined filter in Figure 4.19. That is, the left half of the horizontal filter supplies a value for the first element of the vector, the right horizontal half a value for the second element, and so on.

Suppose that the creature evolved two filters of the same general type. The first provides a measure of the existence or not of a line boundary within its domain. If a line

†The filters as described are rather simple and perhaps simplistic. The filters used in the full ABC simulation model employ spatial frequency filters of various sizes and orientations at each point of the retinal image.

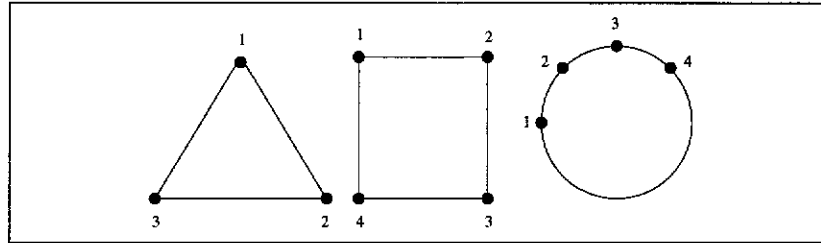


Figure 4.20: Saccade/Fixation Points.

boundary falls on a component of a filter then the corresponding vector element is 1, otherwise it is 0. These line spatial frequency filters thus provide an eight ‘bit’ vector.

The second filter is sensitive to colour—a very simple version of the colour filters discussed in Appendix C. For this colour spatial filter, if a filter section is mostly filled with a particular colour, then the corresponding elements are given the appropriate colour code according to the following table. This filter thus produces a sixteen ‘bit’ vector.

white	00
red	01
green	10
blue	11

We restrict ourselves to only considering a limited set of fixation points for each object, as shown in Figure 4.20. For these fixation points, the values given to the respective

Figure	Saccade Point	Vector
triangle	1	0 0 0 1 0 0 0 1
	2	1 0 1 0 0 0 0 0
	3	0 1 0 0 0 0 1 0
square	1	0 1 0 0 0 1 0 0
	2	1 0 0 0 0 1 0 0
	3	1 0 0 0 1 0 0 0
	4	0 1 0 0 1 0 0 0
circle	1	0 0 0 0 1 0 0 0
	2	0 0 0 0 0 0 1 1
	3	1 1 0 0 0 0 0 0
	4	0 0 1 1 0 0 0 0

Table 4.1: Vector Values for Line Filter

vector elements is shown in Table 4.1 for the line spatial frequency filter, and Table 4.2 for the colour spatial filter.

We restrict ourselves further by only saccading in a clockwise direction. This is not a restriction of the system—the system could learn the fixation points in any order—but it does reduce the time and complexity of the exercise. We thus train the system on clockwise relational saccades (1-2-3-4-1- . . .) for each object. †

The output from the system is a set of concepts which will indicate the shape and

†One could question if the use of particular fixation points is appropriate, and how does this relate to saccades and fixation in the human visual system. This question is discussed in Section 4.5 in relation to the learning of scanpaths.

Figure	Saccade Point	Vector
red triangle	1	0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1
	2	0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
	3	0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
green triangle	1	0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0
	2	1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
	3	0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0
blue triangle	1	0 0 0 0 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1
	2	1 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0
	3	0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0
red square	1	0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0
	2	0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1
	3	0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0
	4	0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0
green square	1	0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0
	2	1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0
	3	1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0
	4	0 0 1 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0
blue square	1	0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 0 0
	2	1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 1
	3	1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 0 0 0 0
	4	0 0 1 1 0 0 0 0 0 1 1 0 0 0 1 1 0 0 0
red circle	1	0 0 0 1 0 0 0 0 1 0 1 0 1 0 1 0 1 0 0
	2	0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 1 0 1 1
	3	0 1 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1
	4	0 1 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0 1
green circle	1	0 0 1 0 0 0 0 1 0 1 0 1 0 1 0 1 0 0 0
	2	0 0 1 0 0 0 0 1 0 0 0 0 1 0 1 0 1 0 1
	3	1 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 1 0 1
	4	1 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0 1 0
blue circle	1	0 0 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0
	2	0 0 1 1 0 0 0 1 1 0 0 0 1 1 1 1 1 1 1
	3	1 1 1 1 0 0 0 1 1 0 0 0 1 1 0 0 1 1 1
	4	1 1 0 0 0 1 1 1 1 0 0 0 1 1 0 0 1 1 1

Table 4.2: Vector Values for Colour Filter

colour of the recognised object. For example, the data supplied to the system consists of a concept and the temporal sequence that, when recognised, would indicate that this concept has been seen. The concept *triangle* includes triangles of all colours, whereas the concept *red* includes all red shapes.

We have not yet described where the concepts, such as *red* or *triangle*, come from. In this temporal learning experiment, the concept data (including labels) is supplied, and thus the ABC model appears to be using supervised learning. However, within the full ABC cognitive model all elements are learned, including the concept labels. The 'labels' are attractors formed by similar learning in the *auditory* modality, which are subsequently associated with the corresponding visual concepts.

We can see this by comparing Figure 3.5 with Figure 4.25, and this can be readily seen as shown in Figure 4.21. The sound attractors of the spoken label (sign) are associated with the visual attractors of the object (or event or concept). Conceptual 'cues' are obtained from association layer links to other modalities, especially the auditory modality which provides the sound 'naming' cues.

The data supplied to the temporal learning system to define the various concepts used in the experiment are:

`triangle Tr@1 ReTr@1 Tr@2 ReTr@2 Tr@3 ReTr@3`

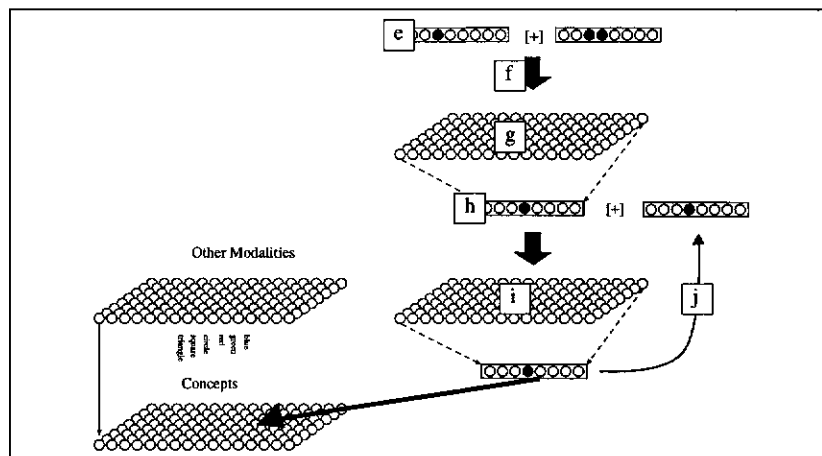


Figure 4.21: Vision System Compared with Vision Sequence-Learning Experiment.

```

triangle Tr@1 GrTr@1 Tr@2 GrTr@2 Tr@3 GrTr@3
triangle Tr@1 B1Tr@1 Tr@2 B1Tr@2 Tr@3 B1Tr@3
...
red Tr@1 ReTr@1 Tr@2 ReTr@2 Tr@3 ReTr@3
red Sq@1 ReSq@1 Sq@2 ReSq@2 Sq@3 ReSq@3 Sq@4 ReSq@4
red Ci@1 ReCi@1 Ci@2 ReCi@2 Ci@3 ReCi@3 Ci@4 ReCi@4
...

```

The string `Tr@1` is a shorthand way of indicating the appropriate vector for the line filter at saccade point 1 of a triangle, `ReTr@1` is shorthand for the actual colour filter vector that would be generated were the creature to fixate on point 1 of a red triangle, and so on. Note that these ‘symbols’ are the equivalent of the characters and words used in the more ‘symbolic’ examples.

The input vectors for the temporal sequences are supplied in pairs, the first for the line filter vector, the second for the colour filter vector. Thus the first line of the input data shown above is a trace of the three saccade points for a red triangle, the second a trace of a green triangle, and so on.

To ensure that the recognition can begin at any saccade point for each object the input data values are wrapped-around in a continuous stream; for example, for the triangle, the sequence 1-2-3-1-2-3-... is used in training. Training thus associates the continuous ‘visual’ streams produced by the line and colour vectors with the appropriate concepts.

The system structure for this example is that shown in Figure 4.25. The two input vectors produced by the line and colour filters are concatenated for mapping to the first SOM surface. The output vector is trained to select one or more concepts as indicated. The concepts are shapes *circle*, *square* and *triangle*, as well as colours *red*, *green* and *blue*. If a *blue square* is recognised, both the *blue* and *square* concept nodes should be selected.

Once the system is trained, we are in a position to see if it is able to recognise objects within this micro-world. Consider the two objects shown in Figure 4.22—a red triangle and a blue square.

The input used to test the system is that which would have been produced by the

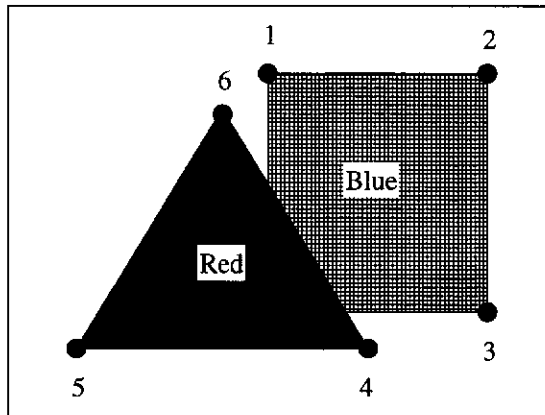


Figure 4.22: Occluded Objects.

creates early visual system in saccading around the objects say from points 1, 2 and 3 on the square to points 4, 5 and 6 on the triangle, all in a clockwise direction to be consistent with out training data. The temporal sequence pairs

Sq@1 B1Sq@1 Sq@2 B1Sq@2 Sq@3 B1Sq@3 Tr@2 ReTr@2 Tr@3 ReTr@3 Tr@1 ReTr@1

are supplied as input and the output concept vector examined to see which concept has been selected.

As an example of a more difficult recognition case, consider the 'scene' as shown in Figure 4.23. Here a green triangle and a red square partially occlude a blue circle. In this case the vectors produced by the filters at points 2 and 5 will not be as per the training examples but will be 'mixed'. Figure 4.24 shows the placement of the filters

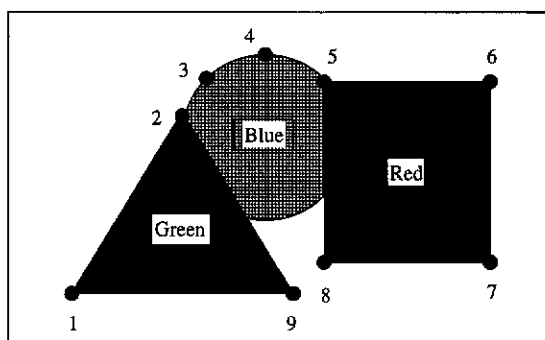


Figure 4.23: Occluded Objects with Mixed Filters at Joins.

when the creature has fixated on points 2 and 5 respectively. The input filter vectors for these points are:

Point 2	Line SF	0 0 0 1 0 0 1 1
	Colour SF	w b w g w g b g 00 11 00 10 00 10 11 10
Point 5	Line SF	0 1 1 0 0 1 0 0
	Colour SF	b r b r w r w b 11 01 11 01 00 01 00 11

When the trained system is tested on the temporal sequence produced by fixating on points 1 to 9 as indicated in Figure 4.23 the output concept vector scores were observed as shown in Table 4.3.

It is clear from this rather simple example that the system *has* been able to recognise the appropriate objects. The output vectors at each recognition stage could be linked to some form of action depending on the recognition: if green triangles are good to eat, a (learned) approach action would be initiated, but if blue circles were dangerous then an avoid action could be undertaken when the blue circle is detected. Of course, this example is rather simplistic, but the general principles are the same as for the full model.

Figure 4.26 illustrates a possible network that would both learn the temporal sequence of 'features' and the sequence of motor control vectors to provide anticipatory *next* saccade points. Let us assume that the eye is currently fixated at a point $P_1(t)$ on a new object that has just been presented to the subject. During the initial learning phase, visual inputs from $P_1(t)$ enter the system (at time t) and progress through the

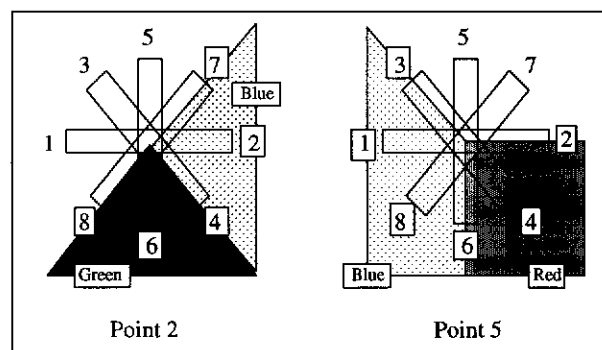


Figure 4.24: Occlusion Filters.

Point No.	Triangle Score	Square Score	Circle Score	Red Score	Green Score	Blue Score	Concept
1	0.62	0.00	0.00	0.00	0.59	0.00	green triangle
2	0.00	0.00	0.77	0.00	0.00	0.83	blue circle
3	0.00	0.00	0.89	0.00	0.00	0.96	blue circle
4	0.00	0.00	0.92	0.00	0.00	0.99	blue circle
5	0.00	0.00	0.92	0.00	0.00	0.99	blue circle
6	0.00	0.95	0.00	0.98	0.00	0.00	red square
7	0.00	0.95	0.00	0.98	0.00	0.00	red square
8	0.00	0.95	0.00	0.98	0.00	0.00	red square
9	0.95	0.00	0.00	0.78	0.86	0.00	green/red triangle
1	0.99	0.00	0.00	0.00	0.95	0.00	green triangle
2	0.99	0.00	0.00	0.00	0.95	0.00	green triangle

Table 4.3: Recognised Concept Scores.

SOM surfaces to be available for association with some eye motor vector. The passage through the system takes time Δt . During this period, the eye make a further saccade to a point $P_2(t + \Delta t)$, and it is the eye muscle vector for this second fixation point that is associated with the initial inputs at time t .

During a subsequent recognition phase, visual inputs from fixation point P_1 can then

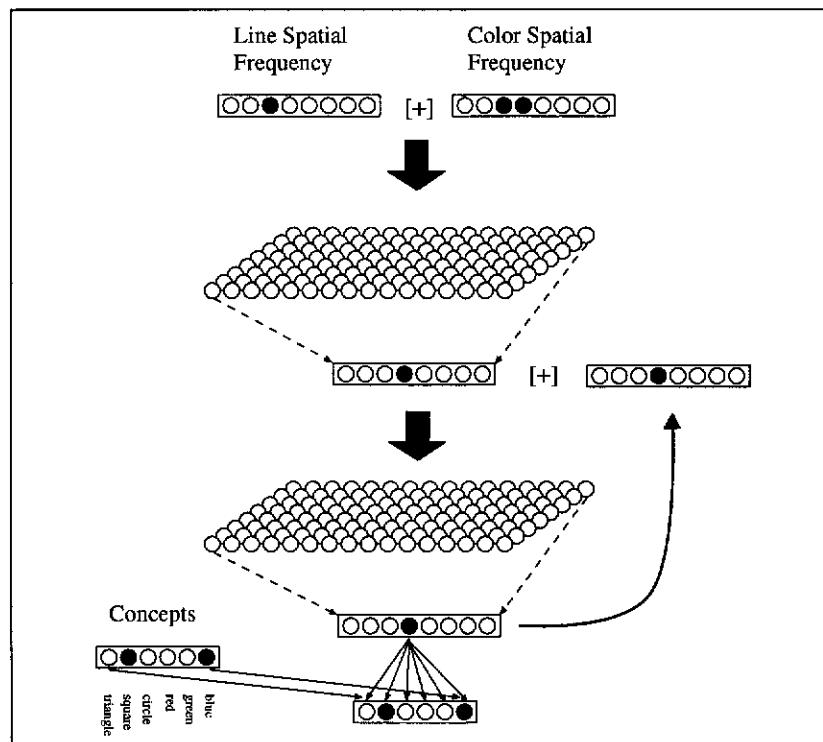


Figure 4.25: Vision Network.

generate the appropriate eye muscle vector required to saccade to fixation point P_2 .[†]

Ballard (1991) summarises:

As humans we have the compelling experience of living in a three-dimensional visual movie. The world appears vividly colourful and stable. One temptation is to propose models of perception that capture this phenomenon in very explicit ways, say as a pictorial memory buffer. If the explicit buffer seems too crude, one can posit elaborate data structures that are equivalent in the sense that they contain the information necessary to construct such a picture.

However, when one examines the *mechanisms* of human and animal visual

[†]Another possibility is that the eye muscle vectors are recorded in another, 'parallel' LAPS structure. The inputs to both the visual LAPS (recording vectors for spatial frequency, colour and motion), and the saccade LAPS (recording eye muscle positions and movements) could be associated in a manner described in the following chapter.

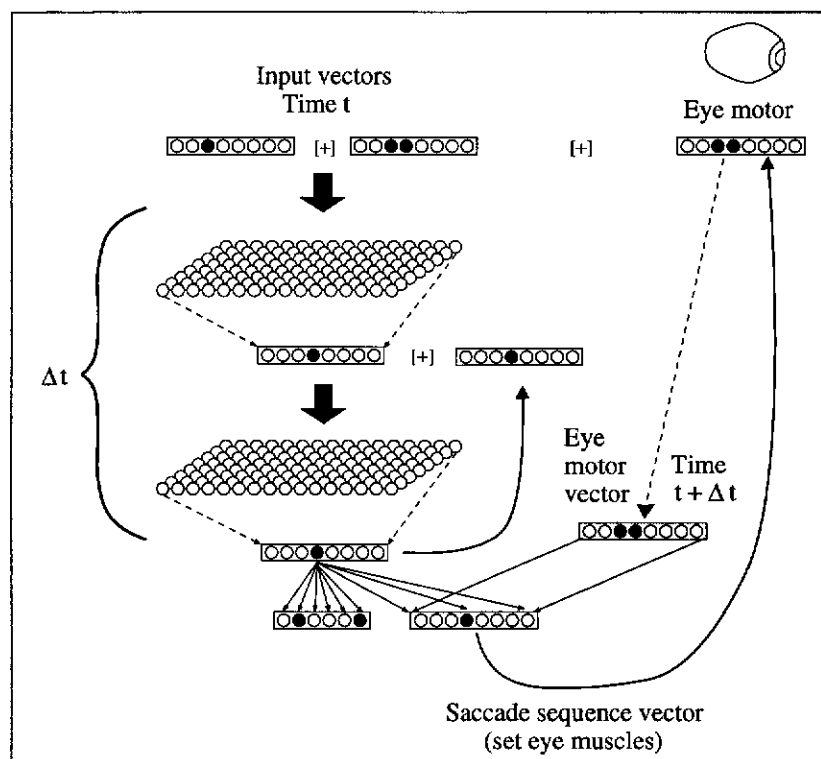


Figure 4.26: Vision Network with Learned Saccades.

perception in detail, or tries to build anthropomorphic robots, it quickly becomes apparent that the way the apparatus works at this level of abstraction, e.g., the fast sequential saccadic searches, is incompatible with phenomenological notions of invariance and stability.

...

Vision depends on the world being sufficiently stable so that behaviours can be executed on demand. Perhaps it is this ability to conduct behaviours that make assumptions about the world that provides the illusion of stable perception.

4.7 ABC Model Avoids Problems of Traditional Models

The ABC model avoids most of the problems mentioned in previous sections, problems which condemn the traditional symbolic computational models. A major characteristic of the ABC model is that it positions cognition as a dynamical system rather than computational, thus eliminating the multitude of issues associated with the symbol processing model and representations.

ABC is a model of the full task of vision, from early receptors to perceptual conceptualisation, labelling and behaviour. It is a consistent, self-similar architecture which is supported by the neuroanatomical evidence. The model is consistent with the scan-path observations of Noton and Stark, and regards visual recognition as a dynamic process of temporal learning and scan-sequence confirmation rather than the 're-cognition' proposed by the traditional models. As well, the strong generalisation effect of the SOM surfaces allows for some latitude ('fuzziness') in finding exact matches for scan-sequences, and local context allows for alternate decisions.

The model is low-level in that it only requires local Hebbian learning, while at the same time allowing for high-level feedback and context, thus providing a mechanism for priming and other top-down influences such as illusory contours. A massively parallel means of 'information' flow is described, with serial processing through temporal

sequences and feedback.

The arbitrary formalisms of inductive and deductive logic are not required in the underlying model, but may be explained in terms of learned behaviour. ABC is able to separate perceptual learning—learning to recognise and take appropriate behaviour with sensory inputs only, without the use of labels or language—and labelled (or supervised) learning. In this model, ‘parts’ are simply learned and labelled bifurcations of previously recognised objects, not fundamental components.

The non-uniformity of retinal sensitivity does not present a problem as ABC does not attempt to ‘reconstruct’ an outside metric iconic view. All that is required is a series of vectors (at least one at each fixation point) which go to make up a temporal sequence. The ‘evidence’ of the rods and cones in the receptive fields is used to construct each vector element value. It does not matter that the point sources of data (rods and cones) are irregular in their distribution, so long as they are relatively *stable* (i.e., have fixed locations) over extended periods in the life of the animal. The rods and cones are combined to give various vectors such as spatial frequency and orientation (of various sizes and angles), colour and motion. In a sense, the choice of external world filters is arbitrary and at the whim of evolutionary pressures. Provided that each vector supplies information to associate with, and discriminate other inputs, it can be employed on the same underlying ABC structural model.

Spatial fusion is not required—rather the model uses temporal integration to achieve recognition. The fill-in of the blind spot is also not an issue, as the inputs that are perceived to come from the blind spot are simply an averaging of the surrounding SOM values. Issues such as shading cues, being learned, will indeed retain their learned orientation—‘vertical’ relative to the head—associated with their usual body position.

The model does not require arbitrary representations or shape primitives, and does not require any form of segmentation. It requires only a retinal frame of reference.

The use of self-organising maps accounts for the re-learning and re-linkage of sensations across modalities, explaining the phenomenon described in the discordant stimulation experiments. For example, placing a prism in front of the eyes for an extended period will mean that the appropriate SOM surfaces in the vision modality will need to re-map

the new sensations, and these new (displaced) visual inputs will need to be re-associated with other sensory inputs such as proprioception.

The system is indeed concept-based rather than object-based, allowing labelling of things which could in no way be recognised as separable objects in the traditional schemes.

4.7.1 Recognition

Recognition in this model is of two types; instantaneous and temporal. Instantaneous recognition is achieved in essentially one fixation, and results when sufficient information is available in a single vector input to allow appropriate output behaviour to proceed. For example, one would expect commonly viewed objects (or concepts such as facial expressions) to be readily recognised because of the learned associations between their input vectors and the linked recognition behaviour. †

Temporal recognition, on the other hand, requires a number of fixation vectors to form a temporal sequence for recognition—the scan-sequence. It is the sequence rather than the individual vectors that leads to recognition.

Once an input vector impinges on a SOM surface near to a previously learned vector, then the generalisation on the SOM surfaces will allow the trace to lock-on to a scan-sequence and follow it for confirmation. Each subsequent confirmation loop will in turn give a predictive recurrent vector which will better locate the scan-sequence, and so on until the associated *concept* attractor is reached, leading to the appropriate behaviour previously associated with that attractor—such as perhaps orally naming the object.

†At this stage in the development of the ABC model, we restrict ourselves to a single winning node on each of the SOM surfaces. A future version might remove this restriction, thus opening the possibility of multiple attention threads.

4.8 Further Discussion and Implications of the Model

In this section we discuss a number of issues which are consistent with, or explained by, the ABC model.

4.8.1 Relational Learning

Minsky & Papert (1969) described a number of figures which are impossible to discriminate for any system that does not involve relational factors. Examples of these types of shapes are shown in Figure 4.27. If, for example, only local feature filters such as corners and line segment 'ends' were available (as per Figure 4.28), then all figures include four corners and two ends.

Many have taken this to provide a powerful argument against the belief that an adequate model of human perception can be constructed which takes independent samples from local feature detectors. However, the temporal sequencing of samples from local isolated feature detectors *is* able to differentiate figures such as these.

At each fixation point, a pair of values is stored—the local feature, either a corner or a line end, and an indicator of the eye position. The eye position indicators are given as numbers 1 to 4 on Figure 4.28. Each of the four shapes in Figure 4.27 then has

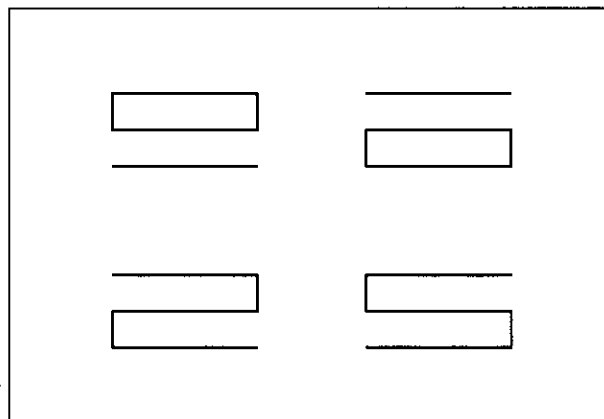


Figure 4.27: Minsky-Papert Figures.

a different temporal sequence of vectors if scanning of the objects is consistent. The sequence vectors for each of the four shapes are indicated in Figure 4.28. A consistent scan of the images, then, will enable them to be discriminated.

As indicated earlier, the use of eye positions is potentially justified as neurons sensitive to eye position have been found in the LNG (Lal & Friedlander 1989), V1 (Trotter et al. 1992, Weyland & Malpeli 1989), V3 (Galletti & Battaglini 1989) and also in the prefrontal cortex (Fuster 1985).

4.8.2 World Filters

The initial component of the ABC model is a series of 'filters' which provide the outermost connection to the external world. For the human visual system, the likely filters are spatial frequency, colour, motion and possibly spatial frequency and colour combined. These at least are the filters chosen for the model implementation.

In the model, these filters are taken to be fixed. However, there is good evidence that at least some of the filters of higher animals undergo some form of self-organisation during development.

For example, Blakemore & Cooper (1970) reared kittens in an environment consisting entirely of horizontal or vertical stripes. The kittens were allowed normal binocular

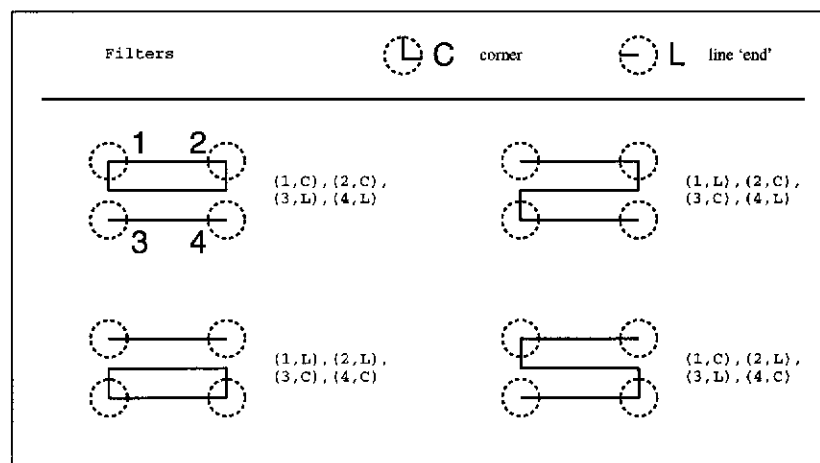


Figure 4.28: Minsky-Papert Figures Recognised Using Temporal Learning.

vision, but their visual environment contained no corners or edges, and they were even prevented from seeing their own body by a wide black collar. Neurophysiological experiments performed after development was completed indicated cells of visual cortex (area 17) that were mostly binocular, and in most other respects like those of a normal animal. However, the distribution of preferred orientation of the cells was quite abnormal, with no cell having an optimal orientation within 20° of the perpendicular to the environmental direction, and with very few cells responding to orientations within 45° of it. They did not attribute this to a passive degradation of certain cortical neurons due to under-activity, but instead concluded that the visual cortex had *adjusted* itself during maturation to the nature of their visual experiences.

This finding is consistent with a self-organising map learning the actual spatial frequencies experienced during the development period of the cat. We postulate a SOM surface as a component of the external world filter for spatial frequencies. Given that the newborn kitten has early receptive fields for all spatial frequencies, only those that it actually experiences during its development period will get mapped to the SOM surface of the external world filter. If the learning rate for this surface decreases to zero over the development period, then the distribution of spatial frequencies experienced during development will become 'hard-wired'.

4.8.3 Receptive Fields in Periphery

The full vector taken in at each fixation point is strongly biased towards the fovea, but the larger receptive fields within the periphery are still represented. This is illustrated in Figure 4.29. This figure is incomplete as there are many more overlapping receptive fields at numerous distances from the fovea which is represented as the small black dot at the centre.

Kuffler (1952) found that receptive fields vary in size across the retina. In the foveal region, the receptive fields are very small with a field centre of a few minutes of arc, whereas at the periphery the fields are much larger with a field centre of 10 to 20° .

Objects are usually learned at (or close to) the fovea, although it is possible to recog-

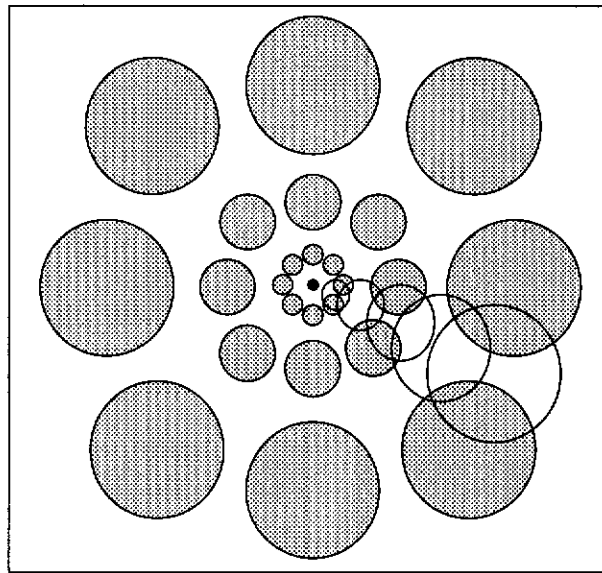


Figure 4.29: Receptive Fields.

nise objects in the periphery using the appropriate peripheral receptive field vector components. However, the vector components associated with the periphery are used to draw attention to objects not currently being fixated. This is especially the case for instances of motion and flicker in the periphery.

Multiple vectors are learned and associated with the same label (cf., Nazir & O'Regan 1990), thus allowing recognition of the object from various viewpoints.

As an aside, particular components of the input foveal vector associated with the periphery may be used as the basis for learning relational terms such as 'above', 'left', 'below', and so on. For example, a high value found in the vector element associated with a peripheral receptive field above the fovea (indicating that something is *above* the current fixation point) may be used to initiate the contextual concept of 'above' (or 'below').

4.8.4 Treisman Figures

The ABC model is able to explain the constant versus linear search time experiments of Treisman and her coworkers (1980). In these experiments, subjects were asked to find particular target symbols in a scene with a number of distractors. For example,

Figure 4.30 (a) illustrates a scene in which a target ‘white square’ is surrounded by black square and white circle distractors. (Figure 4.30 (c) shows a similar case only this time using colour). The reaction time for finding the target object in these “conjunctive search” tasks is linear in the number of distractors. The task is conjunctive in the sense that the subject is required to search for a combination of features found in the distractors—here white and square.

In contrast, targets which have a feature different from the distractors tended to ‘pop-out’, giving a reaction time that is independent of the number of distractors. Treisman & Gelade (1980) termed this task a ‘feature search’.

Treisman & Gelade (1980) showed that for a wide range of stimulus attributes, attention seems to be shifted serially across fixations when a subject is attempting to conjoin independent attributes of a stimulus.

As we discussed in Section 4.7.1, the ABC model can explain *instantaneous* recognition provided the vector obtained from the fixation has unique values for one or more of the vector elements. The feature search condition of Treisman is such a case—unique vector elements are present, and these correspond to a particular direction so that a subsequent saccade can be made directly to the target. The unique vector elements result from a particular target feature being found in one (or several adjacent) receptive fields, even in the periphery. Target features for this ‘pop-out’ are restricted to the external world filters—spatial frequencies and angles, colour and motion. That is, for a ‘feature search’, only the location in the periphery where the independent feature is found will a filter be exciting the corresponding neurons, and so attention will be directed to it in constant time.

The *linear* scan times are required in cases where unique vector elements are not found within any receptive field. To find the target in such cases, the subject is forced to saccade (possibly at random), but with not necessarily one fixation per distractor object, to bring target closer to the fovea. Because of prior learning, recognition is best at the fovea, and so the target must be brought to the fovea so that the learned associations and recognition behaviour can be undertaken.

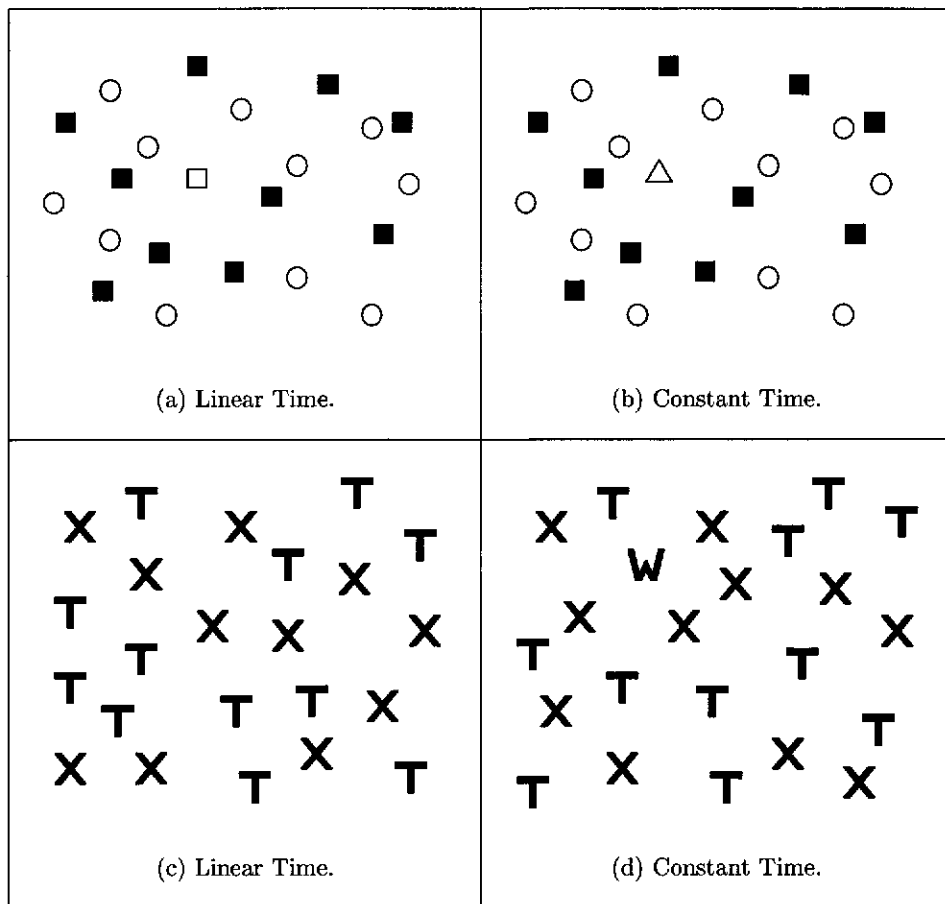


Figure 4.30: Treisman Visual Search.

4.8.5 Integration of Local Features

The traditional explanation for the eventual resolution of ‘difficult’ images such as Figure 4.8 (the dalmation dog on page 180) is that the “percept must result from the synthesis of contour information gathered from over a large portion of the picture. Looking at only local regions of the picture, as cortical cells do, would never lead to the recognition of the scene and the central object within that scene. Rather, we see the scene and the object because of some more widespread, or global, comparison of local contour information” (Sekuler & Blake 1990).

The ABC model suggests a different explanation. The so-called *global* contours are simply not there, and recognition is achieved instead through finding a sequence of smaller foveal images (in which some local *illusory contours* may be generated by spatial

frequency filters) such that the vectors generated by the local images *approximates* those corresponding to the actual object or parts thereof (here the dalmation dog), and the sequence corresponds to a learned scan-sequence for that object. The generalisation processes on the SOM surfaces will enable the ‘noisy’ vectors to *latch on* to the sequence used for recognition of the object, and once this partial recognition of a ‘possible’ object is achieved, then the subsequent scan-sequence may be followed for confirmation.

When the recognition of the object is ‘confirmed’, recurrent contextual information (top-down knowledge) makes it easier to find additional cues and to latch on to a scan-sequence. †

4.8.6 Polysemous Figures

Polysemous images such as the famous ‘Wife and Mother-in-law’ image shown in Figure 4.7 (on page 179) present a problem for traditional vision theories. While the low-level stimulus information reaching the eye (such as ‘contours’) remains constant, the interpretation of the image may fluctuate between a picture of a coy wife and one of the ugly old mother-in-law.

However, an explanation is forthcoming if we consider scan-sequences. The alternating interpretations may be achieved through the alternation of several different scan-sequences, one for each interpretation of the image. If the scan-sequence associated with the wife is currently being employed by the subject, then the wife interpretation will be the accompanying percept, whereas following the scan-sequence associated with the mother-in-law will produce a perception of that figure.

4.8.7 Reinforcement and Learning

The ABC model employs a reinforcement learning strategy. Some have maintained that conventional reinforcement learning faces a serious difficulty. The assumption that the brain creates and maintains a picture-perfect visual scene at each moment creates the

†Note that at a viewing distance of 40 cm, a 2° fovea corresponds to a diameter of approximately 1.4 cm.

problem of just how the brain determines which of the many objects and features in its current world view are the appropriate ones to be reinforced. How does a naive brain determine which “stimulus” in the richly detailed stimulus array is to be credited following a reward—which synapses should be strengthened?

Churchland et al. (1994) suggest that some ‘attention’ mechanism is the solution. They suggest that perhaps the brain has been hard-wired by evolution to bias attention towards properties which are relevant to the survival of a particular species. Their point is that an attended feature of the visual environment (either directly or via some iconic or working memory) is more likely to be causally associated as being relevant for reinforcement than when a rich-replica visual world is used as the representation.

Our view is similar, with fixation visual sequences taking the place of the attention mechanism. The primarily foveal views of each fixation form the component to be reinforced, and it is generally only when an object or event is at, or near, the fovea that it is ‘identified’. The larger receptive fields of the retinal periphery are less involved with the recognition process (although may still participate) and serve more to bring items of interest to the ‘attention’ of the creature and cause it to foveate to the area of interest. While the vector formed at each fixation has a preponderance of information from the well sampled foveal areas, it still includes information from the periphery that may be used for recognition.

The neonate will initially shift attention in a more or less random manner (ignoring any innate reflexive behaviour), and the regularities of the external world, along with any exploration by the organism, will result in linkages forming between the various sensory modality inputs and behaviour through associative reinforcement. The organism will tend to reproduce options which were successful in the past, further strengthening any associations (Whitehead & Ballard 1990, Whitehead & Ballard 1991). The approach is that of standard operant conditioning in which a positively reinforced response is more likely to be reproduced when similar conditions are subsequently encountered (Kandel & Schwartz 1983, Henton & Iverson 1978, Rescorla & Wagner 1972, Mackintosh 1974).

Reinforcement learning has had a number of strong critics in the past, mainly as a result of the seeming plethora of information that a creature has available to it, and

the above-mentioned need for some means of separating out the relevant details for associative reinforcement. Some have even gone as far as suggesting that reinforcement learning cannot be the mechanism appropriate to the sophisticated learning typical of higher cognitive organisms (e.g., Chomsky (1965, 1980) and Fodor (1981)).

The ABC model does not suffer from this inability to associate appropriate sensory inputs for reinforcement. In the ABC vision model, for example, the foveal vector (including any peripheral receptive field information that may be relevant) is exactly the type of attentional detail needed to form a relevant causal link. Only the specific detail of the current (and perhaps some preceding) temporal sequences obtained from attentional fixations is associated with the reward or punishment of the reinforcement signal.

Churchland et al. (1994) suggest that a further skepticism of reinforcement learning as a cognitive learning model results from neural net modelling. Neural nets appear to scale poorly as the number of dimensions of the input space increases to supposedly approximate the rich requirements of a realistic visual scene, resulting in inappropriately long training times and thus rendering reinforcement learning as impractical for complex task domains. However, this criticism of the supposed slowness of reinforced learning in neural nets misses a number of crucial points, and shows a lack of understanding of biological neural structures.

Biological neural structures (and the ABC model) are massively parallel *hardware* systems. The update time is constant no matter how many input lines there are as only local (hardware) interactions are involved. In fact, the more lines the better as it enables better discrimination. A *serial simulation* of a neural net on a von Neuman computer, however, will scale poorly as the number of input lines increases, but this is not relevant. This is a limitation of the simulation vehicle (the serial computer), not of the neural model. A suitable hardware simulation, as discussed elsewhere in this thesis, would not suffer from this deficit.

Another issue ignored by the critics is that learning with artificial neural networks depends upon a number of parameter settings, such as the learning rate. In biological neural nets this is also the case. Faster learning is achieved with a *high-gain* resulting

from a high *value-based* (or ‘emotional’) component. Learning not to put a hand in a fire a second time is an example—the initial pain will result in a sufficiently high gain that the appropriate learned association is achieved very rapidly.

The final point ignored is that for humans and animals, learning is performed over the whole development period and even into adulthood. For humans especially, this is an extended period.

4.8.8 Stabilised Retinal Images

If the eye is prevented from saccading across an image by some means (either by immobilisation of the eye by some paralysing drug or by means of some special optical device), then the image will fade and even disappear entirely (Prichard 1961). The image can be revived by a slight displacement or by modulating the brightness of the stabilised stimulus.

The traditional view is that physiological saturation of the retinal receptors is the reason for the fading, but this view is disputed by MacKay (1991, page 92), who maintains that while satiation no doubt plays a significant part, it does not abolish completely the primary cortical response.

MacKay suggests that the reason for the fading is a loss of covariation in the stabilised signals. The ABC model suggest an alternate but similar reason—if the visual system does not follow any scan-sequence but continues with the same input, then we would expect to lose the ‘perception’ of the image space. The temporal sequence flow would be interrupted, and we would lose the percept of continuity in world.

4.8.9 Reading

The processes of reading constitute an extensive topic that requires more resources than we can allocate here in order to cover it adequately (Pollatsek & Rayner 1989, Morris et al. 1990, Erlich & Rayner 1981, McConkie, Kerr, Reddix & Zola 1988). However, we can make a number of observations in passing, and suggest that this topic might be a suitable one for further investigation in relation to the ABC model.

The ABC model suggests a mechanism for reading. In reading a line of text, each fixation provides a vector of the various spatial frequencies and orientations. These are associated with particular characters early on in the learning-to-read process, and eventually with combinations of characters to forming word (or a number of words in the case of words which frequently appear together). By learning to associate the particular spatial frequency and orientation values from the various receptive fields in the fovea and near-fovea, appropriate behaviour (recognition actions) can then be undertaken.

The model suggests that word recognition is initially sequential character recognition (c-a-t spells cat), but that eventually whole word recognition is achieved. The foveal and extra-foveal learning of whole-word patterns occurs to the right for English readers, to the left for readers of Arabic, and below for readers of Chinese.

The words themselves may be used as prompts for a saccade to the next word. During reading, a saccade often 'lands' the fovea near the third letter of the word, and small corrective saccades are made when this is not satisfactory. Some suggest that this could imply that the eyes are aiming at a target, and hence that at least some crude visual processing has guided the saccade (McConkie et al. 1988, Rayner, Slowiaczek, Clifton & Bertera 1983). However, the ABC model provides an alternative explanation—just as eye-movements are a component of the proposed scan-sequence mechanism, here too an appropriate saccade eye movement may be learned for each word so that the eye is able to skip to a convenient location on the next word. This is consistent with the finding that a word may be skipped more often if it is highly predictable in the text—in effect the multiple words of the frequent combination are treated as a single unit (Morris et al. 1990).

This is illustrated by an example—if a subject is asked to count the number of 'f' characters in:

```
the final location of the saccade
is a function of the full window size and
the total of the distances depends on
the finite size of the words
```

most people respond with 4, but the total is actually 8. Subjects don't count the 'f' characters in the four 'of' words as these words are skipped—the word 'of' is essentially redundant in these cases and so the word is not read as a separate entity.

If a word is encountered for which there is not yet a direct-linked 'meaning' association, then the subject must resort to character-by-character 'phonetic' reading. This view is consistent with the current 'dual-route' explanation of reading (Coltheart, Curtis & Haller 1993). The use of the generalising capabilities of the SOM surfaces also allows for an explanation of the 'analogy' process as proposed by Glushko (1979) and Marcel (1980), who argue that non-words are read aloud by the activation of word lexical entries that are orthographically similar.

The previously discussed masking experiments of reading, in which all text except for that at the fovea (and near-fovea) is masked, is consistent with our scan-sequence proposal (Rayner et al. 1978, Erlich & Rayner 1981, Pollatsek & Rayner 1989, Morris et al. 1990, O'Regan 1990). A moving window of actual text accompanies the saccadic movements of the subjects, who remain unaware that all other text is masked and maintain that the whole text is present and available.

Words in the periphery (whether masked or not) are required for continuous reading. If the non-fixated portion of a scene is totally absent, and there is nothing at all on the screen in its place, the viewer will not move his/her eyes at all. Subjects will only move their eyes if there is some kind of stuff in the periphery to serve as an eye movement target (even if it is blurred or masked). O'Regan has informally confirmed this with reading (O'Regan 1997)—there is no feeling of a whole sentence being present, just a single word. On the other hand, if all the other words are present but masked out, then reading proceeds almost normally.

4.9 Conclusion

The ABC model of vision, described in this chapter, is very different in many respects from the traditional hierarchical recognition theories. Recognition within the model (including predictive, what-next recognition) depends on richly recurrent neural net-

works, and involves associations between the various modalities. This rich recurrence, especially with continuing multi-cortical area input, challenges the conventional conception of a chiefly unidirectional low-to-high processing hierarchy.

There is no real distinction between learning and recognition in the model, both taking place concurrently. The model is dynamic and continuous, and provides very good agreement with experimental findings.

The linkage of self-organising maps, association layers, and recurrency employed in the ABC model shows much promise in providing a real alternative to the established computational paradigm and to finding a better understanding of the processes of vision.

Chapter 5

Further Discussion and Implications

In this chapter, we discuss a number of issues relating to cognition in more detail, and also examine some implications of the ABC model. Limitations of time and space prevent us from attempting a thorough investigation of these areas, but the examination, albeit sometimes cursory, does indicate the power of the ABC model in explaining a number of seemingly disconnected phenomena and opens up many avenues for future research.

5.1 Language

The study of language is a very important one in the study of cognition. The use of language appears to give such a huge boost to human capabilities over those of other animals that a detailed and integrated theory of language is absolutely essential to any theory of cognition. This message is emphasised by Luria (1982, page 27):

Language, in the course of social history, became the decisive instrument which helped humans transcend the boundaries of sensory experience, to assign symbols, and to formulate certain generalisations or categories.

or again (Luria 1982, page 30):

Humans differ from animals because of the existence of human language—a system of codes that designates external objects and their relationships, and helps to arrange these objects into certain systems of categories. This system of codes leads to the formation of abstract thinking, to the formation of “categorical” consciousness.

In his discussion of the works of Vygotsky, Bruner (1987, page 15) states that:

For Vygotsky, the chief instrument of integration and order in human mental life is language, language used in the service of other higher mental functions. But language is not to be taken in Saussure’s (1955) sense as a system of signs. Rather, for Vygotsky, language is a powerful system of tools for use—for use initially in talk, but increasingly and once inwardness is achieved, in perception, in memory, in thought and imagination, even in the exercise of will.

In the past, the study of linguistics has often been conducted in isolation from other disciplines. For example, Chomsky and others have claimed that there are no general principles of human learning, and that each domain (language, vision, mathematics, and so on) has its own particular constraints.

While we agree that language would seem to be biologically constrained, because only humans use language, it is the form of that difference between humans and animals that is of interest. Basic assumptions that have held for a number of decades (including the widely held view that the human language facility is innate), are questioned, and a new proposal put forward based on the ABC model.

In Chapter 3 we outlined the ABC model of language. In this section, we examine various other issues regarding language, and extend our discussion of the ABC model.

5.1.1 Current Theories of Language

Before we begin our discussion of language, it is instructive to briefly examine the history of grammar in Western society. We take this overview from Stork & Widdowson (1974).

The study of grammar began as a philosophical enquiry into the structure of the Greek language by Dionysus Thrax in the first century B.C. Emphasis was placed on written rather than spoken language. The parts of speech (noun, verb, participle, article, pronoun, preposition, adverb, conjunction and so on) were established by Thrax as part of his description of written Greek.

The Romans, who held the Greeks in high regard, took a more prescriptive view, attempting to describe Latin in terms of the Greek categories, despite obvious differences between the two languages. † Thus early attitudes to grammar took a prescriptive view, and grammar had already become a set of rules prescribing 'correct usage' by the end of the Middle Ages.

The Renaissance saw a revival of Greek and Latin scholarship, and scholars tended to take the classical ideas on grammar as universal truths. Grammar became synonymous with knowledge, logic and correctness.

The French grammarians of the Port Royal school developed the idea of universal grammar, and maintained that all human beings use a common thought structure and hence common language structure. Descartes (1596–1650) was one of the leading figures in this movement, combining classical prescriptive grammar with mathematical ideology.

Grammar as a set of normative rules has been dominant in the theory of languages in Western society virtually ever since. The notion of the universal validity of traditional grammatical categories probably reached its zenith in England around the middle of the eighteenth century, and it was not until the end of the nineteenth century that any form of questioning of these assumptions was made by some linguists. An alternate view, that language is a constantly changing communication system arbitrated by the

†For example, Latin has no system of articles and less verb forms than Greek.

mass of people who use it, came to be considered.

Studies into 'exotic' languages, including languages with no written form, forced linguists to examine languages from the standpoint of their own structure rather than from the categories of traditional grammar. The structuralist movement in linguistics approached the classification of parts of speech by using form and sequence, rather than the traditional categories. Various forms of structural grammars were developed, being mostly text based; that is, based upon a corpus of data and observations. Structuralism was strongly influenced by behaviourism, particularly the notion that language skills were largely a question of stimulus and response and the formation of habits (behaviours).

In the second half of the twentieth century, a radically different form of structural grammar was proposed by Chomsky—the transformational generative grammar. Chomsky's initial rejection of the behaviourist view was based on the proposition that language was *creative*, that humans are able to use and understand sentences that they have never heard or used before. Further, the number of possible sentences in the language was deemed to be infinite. Thus for Chomsky, the aim of linguistics was to explain the underlying system which makes it possible for a language to be infinite. In this sense, Chomsky was returning to the views of the past, specifically those of the Rationalist Grammarians of the sixteenth century, and to those of Plato in his insistence that grammar is innate.

Many think that a grammar is essential; for example, Terrace (1985, page 1012):

Psychologists, psycholinguists, and linguists are in general agreement that using a human language indicates knowledge of a grammar. How else can we account for a child's ultimate ability to create an indefinitely large number of meaningful sentences from a finite number of words?

The view put forward by the ABC model, however, is that there is *no* innate grammar in the brain, but rather learned temporal sequences based on a lifetime of social interactions with other members of society. These language exchanges form (self-organise) our brains so that the most strongly supported sequences and structures (in terms of

neural weights) are those of the ‘grammar’. Our learned ‘grammar’ is contained within the speech of the majority of others with whom we interact. We learn ‘grammatical’ sequences because we pick up these sequences in the speech of others during our period of development. Language is entirely a learned phenomenon in the ABC model.

This view has been stated before in regard to connectionist models of language, such as Rumelhart & McClelland (1987, page 246):

... we have shown that a reasonable account of the acquisition of the past tense can be provided without recourse to the notion of a ‘rule’ as anything more than a description of the language.

We believe that the current computational paradigm prevalent in linguistics is not appropriate to cognition or language. This is especially the case for the innate (programmed) nature of transformational grammars. The ABC model returns to the empiricist notions held by the behaviourists which suggested that language is a communication skill that is learned by each individual.

5.1.2 Chomsky and Language

The main stream of linguistics has been dominated by the work of Chomsky for the past four decades. His influence has been immense, and although his views on language have evolved over the years, they remain fixed on the essentially Platonic view of an innate language facility (Chomsky 1957, 1965, 1972, 1975a, 1980, 1988). See also Pinker (1984, 1989).

As stated by Bates & Elman (1996, page 1849), Chomsky “has argued for 40 years that language is unlearnable; he and his followers have generalised this belief to other cognitive domains, denying the existence of learning as a meaningful scientific construct.”

Chomsky and his followers believe that language must be innate. For example, Chomsky (1967, page 81) made the observation that “a grammar is no more learned than, say, the ability to walk is learned.” Their belief is that children have a built-in language capability which allows them to select a particular grammar appropriate to their

society. A knowledge of the class of human grammars is genetically endowed, and the task of the child is to initially choose the appropriate grammar, which then provides the child with the particular rules associated with that grammar. To Chomsky, grammar is an idealisation of language (*competence*) as compared to the the actual use of language (*performance*).

Some of the reasons for holding this belief are given in the discussion below. In this section we also discuss other issues relating to the *transformational grammar* view of language.

poverty of stimulus

It is claimed that the language input to the child is inadequate for it to learn a grammar—the so-called ‘poverty of stimulus’ argument (Chomsky 1965, Lightfoot 1982). Some specific points described by Seidenberg (1996) are:

The input to the child is degenerate, consisting of both grammatical and ungrammatical sentences that are not labelled as such. It is also variable: children are exposed to different samples of utterances but converge on the same grammar. The input does not include reliable negative evidence, that is, evidence about which structures are not allowed by the grammar; logical arguments suggest that in the absence of such evidence there must be strong innate constraints on the possible forms of grammars (Gold 1967). Finally, languages exhibit properties for which there is no positive evidence in the input.

The implication is that the utterances to which children are exposed cannot account for their subsequent language usage. Language is then seen as essentially unlearnable, and so it must be a kind of human instinct (Pinker 1994). The language learning that the child performs is then restricted to the acquisition of a lexicon and the setting of a few parameters required of the *universal grammar* module in order to select the appropriate grammar.

We, however, believe that this view of the poverty of the stimulus is no longer tenable. As we discuss later in Section 5.1.5, children are able to distinguish word boundaries by statistical means after only *two minutes exposure* to language-like se-

quences, indicating a very powerful learning facility of the kind suggested by the ABC model.

Language learning in the ABC model is not directed toward the learning of a grammar as such, but rather the learning of temporal sequences. The attainment of the 'mathematical' purity of a grammar is not required. The ABC model is a theory of actual language *performance* rather than one of an underlying *competence* and grammar.

As language learning in the ABC model is effectively based upon the acquisition of usage statistics (acquired in the process of interacting with other members of society), then these statistics will be heavily biased by the temporal sequences used by that society. The occasional 'ungrammatical' sentence will not make any difference and will be overwhelmed by the majority of 'grammatical' sentences. However, in certain sub-groups of society in which a so-called ungrammatical speech pattern is the norm, children will learn this pattern as it forms the majority statistics for that group. All children learn from those closest to them, and so will take on the idiosyncratic 'grammars' of their immediate society.[†]

There is no requirement for sentences to be labelled as being grammatical or not—it is not the grammar that is being learned but the sequences. Chomsky and his followers make the mistake of assuming that 'rules' are required to be learned, and so any unlabelled and ungrammatical sentence would of necessity be learned as an invalid and extraneous 'rule' of the grammar. But in the ABC model, the learning is *not* of rules, but of temporal statistics. The very nature of statistical learning means that negative evidence is not required.

A number of studies show that there is no direct evidence that language users actually manipulate rules in their heads. These studies include Ervin-Tripp (1966), Slobin (1971) and Hockett (1954). MacWinney, Leinbach, Taraban & McDonald (1989) provide a discussion of whether language uses rules or cues, and describes a connectionist approach to learning the correct usage of German articles—a non-trivial computational

[†]The author is aware of such a case of a child who learned an incorrect grammatical behaviour from his family. The child was subsequently 're-programmed' by a benevolent teacher who inflicted a nasty pain to the upper arm whenever the child used the incorrect form. The behaviour changed over time, and the child subsequently learned and used the 'correct' grammatical form.

task. They come down strongly in favour of the 'cues' alternative, and show that the connectionist model mirrors the developmental errors made by German children.

The fact that children are exposed to numerous interactions with a number of members of a society accounts for the fact that children tend to converge to a similar set of language sequences. But this is not the same as saying they converge on the same grammar. The statistical modifications to their neural connections resulting from their exposure to similar language usage will mean that they are able to perform *similar* comprehension and language production behaviours.

The argument by Gold (1967) purporting to show that innate constraints must be found in order for language to be learned, is in fact not relevant to the ABC model. This argument is only appropriate to string-based symbolic learning of (the rules of) grammars on serial binary computers—none of which is appropriate to the ABC model. We examine the relevance of Gold's theory in more detail in a later section.

learning mechanisms

The behaviourist approach to language put forward by Skinner (1957) suffered a devastating critique in a book review by Chomsky (1964). This single event is still regarded as one of the most important in ensuring that the views of Chomsky gained the ascendancy, and is on a par with the criticism of connectionism by Minsky & Papert (1988)—both ensured the ascendancy of cognitivism by delivering a knock-out punch to their opponents. Unfortunately, both were given far too much credibility and both impeded any real alternative viewpoint afforded by their opponents from being taken seriously. Also, unfortunately, both were wrong.

Skinner and the behaviourists regarded reinforcement learning as the key to an understanding of cognition, and this view dominated much of the theoretical discussion in psychology for several decades (Nye 1992). The behaviourists held that learning formed an associative link between a stimulus and specific responses. The association could be strengthened by positive reinforcement (either some positive or pleasant outcome, or the removal of some negative or unpleasant event). It could also be weakened by negative reinforcement (either through the introduction of some object or event considered unpleasant, or through the removal of something pleasant). Reinforcement was given when an appropriate response followed the stimulus.

The book by Skinner was an attempt to explain language behaviour using associative rules. In the review, Chomsky was able to show that the simple stimulus-response associations suggested by Skinner would not be able to perform some kinds of computations thought to be required for language usage. The computing mechanism proposed by Skinner was not powerful enough to do the computations that Chomsky believed language users perform—specifically, learning a grammar.

The learning mechanisms used by the ABC model (and Elman’s SRN model) are much more sophisticated than the simple chain of stimulus-response pairs criticised by Chomsky, and are thus not open to the same criticisms. In fact, as we show in a later section, the learning capabilities of an SRN is at least that of a binary computer. We expect the LAPS component of the ABC model to have similar, if not better learning capabilities as indicated by some of the experiments performed in Chapter 2.

rules

Dreyfus (1992) points out that “Chomsky and the transformational linguists have found that by abstracting from human *performance*—the use of particular sentences on particular occasions—they can formalise what remains, that is, the human ability to recognise grammatically well-formed sentences and to reject ill-formed ones. That is, they can provide a formal theory of much of linguistic *competence*.”

The ABC model maintains that these ‘rules’ are mere *descriptions* of the (idealised) language behaviours of certain individuals and *not* the *mechanism* of language understanding and production. We take up a similar point in Section 5.4.2 when we criticise the cognitivist misuse of the concept of rules in everyday cognition. It is our contention that rules and computation are not appropriate in a model of cognition.

As put by Wittgenstein (1960a, page 25):

In general we don’t *use* language according to strict rules—it hasn’t been taught us by means of strict rules either.

universal grammar

Chomsky argues that all (human) languages share a common basis—a *universal grammar*. Learning a language then requires only the setting of a few *parameters* to select a particular grammar from amongst this set of alternatives—to in effect, tune the general mechanism to the grammar required. But as Arbib & Hill (1988, page 58) point out:

Each language has idiosyncrasies of syntax that fill far more pages of the grammar books than do those general principles subsumed by ‘parameter settings’

To Chomsky, the knowledge of a language is essentially within the innate rules of a grammar, and it is the child’s task to determine the appropriate set of rules prescribed for each language. This view holds the grammar as the primary notion. But within the ABC model, the knowledge of a language is found in the learned sequences, and the task of the child is to learn (to as close an approximation as possible) the structures of the utterances of their care-givers and confederates.

In the ABC model, language production is determined by the learned neural weights. Through learning and thus modification of the synaptic weights, the brain encodes a large number of probabilistic constraints derived from prior experience. These constraints include simple and complex contingencies between different types of information, as we saw in Chapter 2.

Those few aspects which are similar or identical between languages must relate to commonalities in the lives of the inhabitants of the earth. Chomsky’s universal grammar would then be a form of *description* of these general properties of language. Attempts to define these universal grammar rules and components, however, have not been successful. For example Crystal (1992, page 85) states:

The universalist ideal is to be able to make succinct and interesting statements that hold, without exception, for all languages. In practice, very few such statements can be made: The succinct ones often seem to state the obvious (e.g., all languages have vowels); and the interesting ones often seem to require considerable technical qualification. Most of the time, in fact, it is clear that ‘absolute’ (or exception-less) universals do not exist.

competence and performance

The *competence hypothesis* holds that a grammar is at the core of human language. The grammar prescribes the rules of the language, but in producing real language, people are restricted by so-called memory constraints, interruptions and other factors which limits their ability to produce correct grammatical constructions in their language. For example, Chomsky (1965, page 3) states:

Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance.

To many linguists, actual language performance is seen as an imperfect presentation of an underlying 'perfect' system. Much is excluded from actual language usage in the expression of this idea. As Seidenberg (1996, page 1599) states, "the clear cases that are often cited in separating competence from performance include dysfluencies and errors. In practice, however, the competence story also excludes other factors that affect language use, including the nature of the perceptual and motor systems that are used; memory capacities that limit the complexity of utterances that can be produced or understood; and reasoning capabilities used in comprehending text or discourse."

Further, the competence theory also takes no account of the communicative functions and mechanisms of language, and completely ignores any account of the statistical and probabilistic aspects of language.

The competence theory pre-supposes a heterogeneous speech community. Hymes (1971) looks at performance and competence from a cultural point of view, and suggests that the reality is of differential competencies within a so-called heterogeneous speech community:

Cazden (1966) has reported on the relevance of social class in the development of mastery of a full range of functional language varieties. In general upper social status children are more advanced linguistically than lower social class children. ... Within the present view of linguistic competence, there is no way to distinguish between abilities of pure speakers and of poor speakers. Sentences from either type would be referred to a common grammar.

Hymes (1971, page 8) contends that in some fundamental sense, the competence of the various speakers may be very different:

Types of English in general constitute a continuum, perhaps a scale, from

reduced to more expanded varieties, somewhat cross-cut by the different uses and adaption of the same original material. A linguist analysing data from a community on the assumption “once English, always English” would miss and sadly misrepresent the actual competence supposedly expressed by his grammar.

We would go further and suggest that the performance of each individual is entirely idiosyncratic, but with various sub-groups of society exhibiting linguistic behaviours that are similar.

Language is a socially determined phenomenon, with the acceptability of various forms of speech determined by implicit consensual acceptance and usage. As put by Hymes, “communities may draw a line of acceptability within expression; where the line is drawn may change over time within the community, and may differ between communities with identical grammars. Indeed, the drawing of the line of acceptability may be a function of personal culture.”

Bernstein (1961a, 1961b) also argues that within what is often regarded as a heterogeneous speech community, there may exist broadly based social differences in language usage. At a broad-brush level, the middle class has different linguistic structures and behavioural functions as compared to the working class. These views are supported by empirical evidence showing social class differences in the speech and writing of adolescents (Robinson & Rackstraw 1967). Similar views and observations are expressed by Ingram (1971) in relation to child learning.

On empirical grounds, the competency theory seems inappropriate. For example, one of the premises of an English transformational grammars as developed by linguists is that humans are capable of accepting (parsing) and understanding central-embedded sentences to any depth. However, Ingram (1971) states that there are experiments which show that most English speaking people will accept sentences with one central embedding, but not two. Bach, Brown & Marslen-Wilson (1986) report that empirical studies have shown that sentences with three or more centre-embeddings are universally hard to process and understand. This number is certainly not indicative of a need for an infinitely recursive linguistic ability. †

†Christiansen & Chater (1994) describe connectionist experiments in the learning of (limited) centre-

As Christiansen (1994) points out, the competence/performance position renders linguistic theory immune to all empirical falsification. In modern linguistics, the paradigmatic method of obtaining data is through intuitive grammaticality judgements. But on what basis are these judgements considered realistic? By denying and ignoring any evidence which contradicts their judgement and their grammatical theories, the views of linguists are “empirically impenetrable to psycholinguistic counter-evidence.”

The reasoning of lack of memory and so on may seem to be appropriate for language *production*, but how does it account for language *understanding*. Not all linguistic behaviour that we hear is ‘grammatical’, yet we are able to understand it. The sentence may sound odd, but the meaning may be clear. What then is the purpose of rules, and how might the ungrammatical sentence be parsed by the grammar?

There are even cases where creative writers bend the ‘rules’ for effect. Or again, a foreign speaker of English might retain the grammatical constructs of their native language, and make frequent grammatical mistakes. The utterance may not only be outside the rules but actually *proscribed* by them. Yet such violations may often go unnoticed, and the meaning of the speaker understood.

generativity

One of the essential properties of a grammar is said to be *generativity*, its ability to produce and comprehend an infinite number of sentences. A linguist might suggest that as there are such a huge number of possible sentences, *it makes practical sense* to assume that some kind of logical syntax exists—it would be hard to figure out how language could function without some global rule-like operations, however implemented.

While we feel that infinity is a bit beyond the human capacity, we concur that humans are able to generate sentences whose strictly identical form they have not produced before. However, each of the components of that sentence *will* invariably be familiar to the speaker.

This means that a meaningful *context* is available to the speaker, and as we demonstrated in Chapter 2, context and prior temporal connectivities between words seems to be all that is required to generate a new sequence, one that may not have been previously ‘heard’ or generated by the speaker. Context, and the ‘rule-like’ learned embedded sentences, and a related although less common construction—cross-dependencies. Their results indicate that SRNs are able to handle these sentence constructs.

temporal associations are the key.

semantics & syntax

Slobin (in Smith & Miller 1966) has suggested that a model of language acquisition that did not take meaning and context into account was not very convincing. In recent times, there has been an increasing realization of the centrality of semantics for language use.

The Chomsky generative grammar formalism, on the other hand, pays no real attention to meaning, instead suggesting that meaning can be obtained through a knowledge of syntax in association with a lexicon. Just how this lexicon and meaning is obtained is not specified.

Within the ABC model, words and their meaning are intimately connected through associative links and temporal sequences.

Another requirement of meaning and rules is pointed out by Dreyfus (1992):

To have a complete theory of what speakers are able to do, one must not only have grammatical and semantic rules but further rules which would enable a person to recognise the *context* in which the rules must be applied.

Context is not considered in the generative grammar model, but it is an an integral part of the ABC model.

5.1.3 Formal Learning Theory

In textbooks on language learning, we are usually informed that a formal demonstration has been given which shows that the learning of systems like human language cannot be based solely on the type of input data children receive (Gold 1967, Wexler & Culicover 1980).

Gold (1967) studied the theory of grammatical inference, and showed that, even for simple classes of languages, no procedure (statistical or other) exists that could learn a language without nontrivial a priori assumptions. The task was to identify a 'correct' grammar for an unknown target language, given a number of examples of the language. The input to the learning procedure consisted of strings of symbols representing sentences within the language.

In one version of the Gold paradigm for inductive inference, the language learner is presented with the *text* of a language, i.e., an infinite sequence of strings made up of all and only strings from the language. The learner is said to learn a language if, on any text for it, the learner's guess *converges* to the same language being presented (Kapur 1993).

Gold found that an infinite number of strings were required to learn the grammar. He introduced the notion of *identification in the limit*; that is, the grammar will be identified in the limit as the number of strings tends to infinity. Within computational learning theory, *identification in the limit* is now one of three major established formal models for learning from example (inductive inference).[†] Learning in the limit views learning as an infinite process and provides a learning model where an infinite sequence of examples of the unknown grammar is presented to an inference algorithm in order to learn the form of the grammar (the rules).[‡]

So far so good, but the problem is that this learning theory is not appropriate to human cognition. It may be useful when dealing with strings on a serial binary computers (such as database applications), but as we show elsewhere in this thesis, that is not the model appropriate to human cognition.

As pointed out by Bates & Elman (1996), Gold's theorem "is relevant only if we make assumptions about the nature of the learning device that are wildly unlike the conditions that hold in any known nervous system."

As stated elsewhere in this thesis, the world is not made up of linguistic 'simples' from which all language terms are formed. The input to the learning device is then not a set of prescribed symbols, and in general, human learning is not a process of induction over symbols.[§]

[†]The other two are the *query learning model* by Angluin (1988), and the PAC (*probably approximately correct*) learning model of Valiant (1984).

[‡]For surveys of work in this model, see Osherton, Stob & Weinstein (1986) and Angluin & Smith (1987).

[§]This is not to say that people do not at times use inductive inference, or more likely *abduction*, in their reasoning. However, induction over pre-established symbols is not the primary mechanism of learning.

Rather, the processes of learning are self-organisation on SOM surfaces and temporal sequence recording, using Hebbian modification of neural weights. The inputs are sensory vectors, and the mechanism is *statistics* and not *induction*.

In the ABC model, language is learned in association with already existing perceptual attractors which provide grounding and meaning for the linguistic terms. The ‘meaning’ (neural connections) of these terms are not fixed, but change over time in accordance with usage.

The task of the language learner is then to link these meaningful terms into an approximation of the sequences used by other speakers who interact with the learner, not to find prescribed rules. Further, the learner does not have to go through discrete stages in forming a symbolic rule, but can incrementally approach the temporal sequence statistics through experience. In the ABC model, meaning and relevance is constantly available to guide the learner.

As we show in the following section, the SRN model is indeed a powerful statistical learner. Although we have yet to confirm our judgement, we feel that the learning mechanisms in the ABC model will be even more appropriate to language learning because of the strong generalisation capability of the hidden layer nodes.

The view that the English language is a formal system reaches its apotheosis in the work of Montague (1974). It is our contention that natural languages are not formal systems. Linguists produce such systems as *descriptions* of language, but to call a natural language a formal system is to confuse the description with what it describes.

5.1.4 Learning in Neural Networks—Theory

One of the issues concerning language has been the claim by many linguists (such as Chomsky and Fodor) that a connectionist approach could not account for the learning of language. Some (for example Fodor & Pylyshyn (1988), Fodor (1988), Pylyshyn (1984) and Pylyshyn (1989)) go as far as to suggest that cognition must consist of essentially logic-like transformations over language-like representations. In this section we mention a number of points which suggest otherwise.

Kremer (1995) has shown that an Elman-style recurrent network (SRN) is *as powerful as any digital computer with finite memory*. A similar statement was made in 1967 by Minsky in relation to general connectionist networks. His claim was that “every finite-state machine is equivalent to and can be ‘simulated’ by some neural net” Minsky (1967, page 55). Minsky’s statement related to networks with no restrictions on connectivity—Kremer’s result, however, deals exclusively with SRNs.

Kremer’s result applies to what an SRN can do in principle, not necessarily in practice. In fact, Mozer (1993, page 11) suggests that the SRN architecture “seems sufficiently powerful *in principle* to handle arbitrary tasks *In practice*, however, many have found that gradient descent is not sufficiently powerful to discover the sort of relationships that exist in temporal sequences.”

Other theoretical learning results concerning recurrent neural networks include: Seidl & Lorenz (1991) have proven that a recurrent neural network can approximate a nonlinear dynamical system; recurrent neural networks have recently been demonstrated to have the ability to learn simple grammars (Cleeremans et al. 1989, Elman 1990, Elman 1991, Fahlman 1991b, Giles, Miller, Chen, Chen, Sun & Lee 1992, Jordan 1989b, Pollack 1991, Rumelhart & McClelland 1986, Williams & Zipser 1989); recurrent neural networks have also been shown to be able to learn deterministic context-free grammars by using an external ‘continuous stack’ (Das, Giles & Sun 1992); Zeng, Goodman & Smyth (1994) have shown how analogue second-order networks are able to learn deterministic context-free grammars; and Giles, Horne & Lin (1995) have shown that large finite state machines can be learned by limiting the class of machines and the type of recurrent network.

The theoretical treatment of recurrent neural networks still has a long way to go. We are very keen to examine the learning mechanisms of the ABC model, but this must be left to further research. The experimental results reported in Chapter 2 point to a very efficient and versatile learning mechanism.

5.1.5 Statistical Language Learning

In an important recent development, Saffran, Aslin & Newport (1996) have shown that eight-month-old infants are able to use simple statistics to discover word boundaries in speech. Moreover, they are able to do so after being exposed to unbroken strings of nonsense syllables for only two minutes. As pointed out by Saffran et al. (1996, page 1926), this suggests that “infants have access to a powerful mechanism for the computation of statistical properties of the language input.”

This seems to provide strong evidence counteracting the ‘poverty of the stimulus’ argument (Chomsky 1965). As pointed out by Saffran et al., linguists have generally assumed, following Chomsky, that an ‘experience-independent’ mechanism of language learning is required because of the supposedly incomplete and sparsely represented language input compared to the child’s eventual linguistic abilities. Further, the child is seen as using highly complex forms of language production in a short period. Given the supposed difficulty of acquiring appropriate linguistic information simply from experience, most theorists have adopted the view that the child has an innate language mechanism and have downplayed the role of learning. The suggestion is that learning does not play a primary role in the acquisition of the more complicated aspects of language, primarily the learning of a grammar.

Notable exceptions to this view include research on prenatal exposure to maternal speech (DeCasper, Lecanuet, Busnel, Granier-Deferre & Maugeais 1994), and early postnatal preferences (Mehler, Jusczyk, Lambertz, Halsted, Bertoncini & Amiel-Tison 1988). Both of these studies show a strong learning capability in infants.

The results of Saffran et al. indicate that infants possess a powerful learning mechanism which is suited to learning the types of structures exemplified in linguistic systems. The performance of the infants in the study was particularly impressive when the impoverished nature of the familiarisation speech stream is considered—it contained no pauses, intonational patterns, or any other cues that, in normal speech, probabilistically supplement the sequential statistics inherent in the structure of words. Their findings “... raise the intriguing possibility that infants possess experience-dependent mechanisms that may be powerful enough to support not only word segmentation but also

the acquisition of other aspects of language” (Saffran et al. 1996).

As Seidenberg (1996, page 1601) points out, these studies show that “infants naturally and automatically encode statistical aspects of care-giver speech without overt guidance or reward. †

5.1.6 ABC and Language

Learning based on the frequencies and distributions of environmental events is emerging as an essential aspect of cognitive development.

(Seidenberg 1996, page 1601)

The approach to language taken in the ABC model is completely different to the transformational grammar views of Chomsky.

In the ABC model, a child learns to use a language, not identify a grammar. The so-called rules of grammars (as constructed by linguists) are seen to be human constructions that are *approximations* to, and *idealisations* of, the actual usage of words and language by real people. The emphasis is placed on how language is acquired and used, rather than on an idealised competence grammar.

In the sense of controlling the production of speech and of language comprehension, grammars do not exist. Every individual has their own idiosyncratic temporal sequences, based on an individual history of interactions with the environment and other speakers. It is these learned temporal sequences, as well as context and other factors, that determine their language usage.

temporal learning

The ABC model learns temporal sequences of sounds within words, and then sequences

†Other studies showing the strong learning ability of infants are Jusczyk (1997) and Morgan & Demuth (1996). The study by Jusczyk, for example, indicates that newborn infants have a preference for their native language. An explanation for this observation consistent with the ABC model would suggest that the process of self-organisation of sounds into the particular language phonemes begins while the child is in utero.

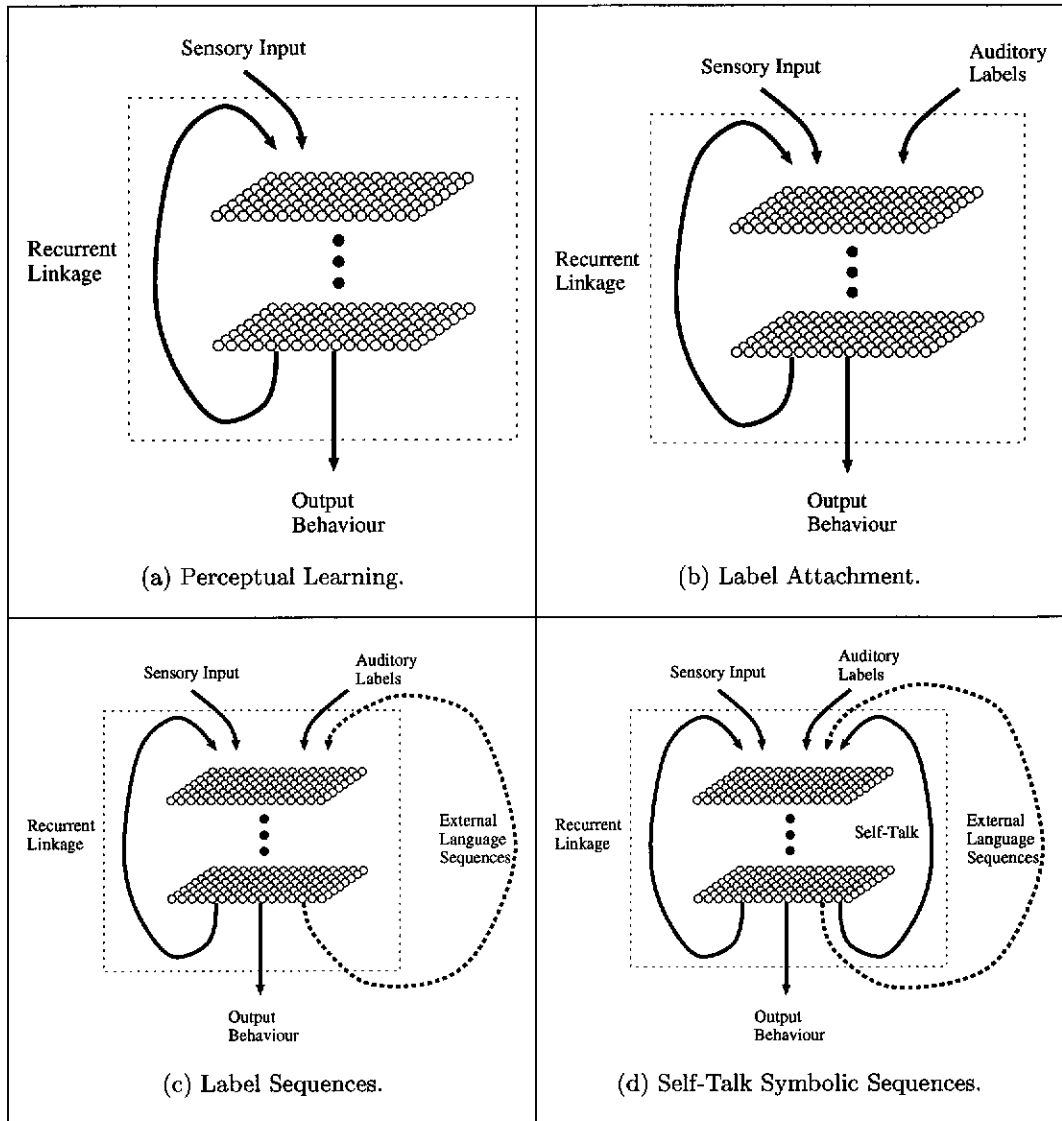


Figure 5.1: The Development of Language in the ABC Model.

of words within longer sequences (sentences). The fact that the language input may include ungrammatical utterances is of no concern because the model is not performing grammatical identification, but is recording usage statistics—the occasional ungrammatical sentence will be counteracted by the majority of ‘correct’ sentences. †

However, if ungrammatical utterances are made consistently by those people who

†Note that in the ABC model, there is no such thing as a ‘grammatical’ sentence or an ‘ungrammatical’ sentence. No usage of language is ‘correct’ in the absolute sense, because all language usage and structure is arbitrary. The only sense in which a sentence may be ‘grammatical’ or not is via societal agreement.

provide the majority language input to the child, then this version of the language will be learned. The model is thus able to account for regional dialects and sub-culture language forms.

Pinker (1996) maintains that unlike sequences of sounds within words, (which are finite in number and therefore able to be memorised), word sequences cannot be memorised because they form an open set. Moreover, Pinker maintains that “grammar does not merely sequence words, but relates each sequence to a meaning through hierarchical, and cross-referenced data structures.”

Pinker maintains that learning words and learning grammar are different computational problems. The statistical learning procedures that have to date been applied to grammar do not behave even remotely like people, but instead guess the next word of a string in a highly simplified artificial language, rather than converting meanings to real sentences and vice versa. But as we were able to indicate in Chapter 2, the ABC is able to generate multi-word sequences based on context.

Pinker’s statement that grammar relates sequences to a meaning is not the case in the ABC model—meaning is directly linked through the perceptual and linguistic attractors which in turn are linked to behaviour (including self-talk). Thus we are able to ‘think’ about objects and events with ‘meanings’ intimately connected. Further, in the ABC model, structure results from meaning, not meaning from structure as suggested by Pinker and other linguists.

sensorimotor behaviours

For the ABC model, the process of language acquisition by an infant is preceded by the learning of sensorimotor behaviours—behaviours which directly link sensory inputs to behavioural outputs. The child is able to reach out and touch an object, for example, and has started to differentiate objects and events in the world. The processes involved in the ABC model in these early sensorimotor behaviours are indicated in Figure 5.1 (a) (in which we have simplified the model and included only representative SOM surfaces).

This differentiation is made on the basis of differential outputs—the sight of a toy elicits a particular cluster of behaviours, whereas the sight of a cup of milk requires a somewhat different behaviour. What started out for the infant as a world devoid of separable objects and events soon becomes (self-) organised into various ‘concepts’

within the brain, which result in accompanying motor behaviours—the ‘explore-toy’ concept and the ‘enjoy-milk’ concept for example.

The view that language is built upon a previously learned cognitive basis is not new. For example, Locke (1978, page 115) states that “this general sense of semantics [as general conceptual structures] must antedate the acquisition of language, both phylogenetically and ontogenetically, and may well serve as the biological foundation on which language is erected.”

McNamara (1971) suggests that “infants learn their mother tongue by first determining, independent of language, the meaning which a speaker intends to convey to them, and then working out the relationship between the meaning and the expression they heard. In other words, the infant uses meaning as a clue to language, rather than language as a clue to meaning.”

As well, McNamara (1972) argues that the child maps language onto prior cognitions, and Slobin (1973) lists cognitive prerequisites for the acquisition of language.

labelling

The initial use of language for the child is the attachment of labels to particular referents, most usually particular objects or events. The usage of language is encouraged by parents, and reinforced within the child by positive behaviours on their part. Labels at this stage are just another property of the object or event.

Linguists insist that children must *learn the meaning* of words like ‘cat’ and ‘eat’, but we suggest that they are already aware of the ‘referent’, that is, the object or event associated with the word. The process of early language learning is to associate labels with the known perceptual ‘meanings’.

As we discussed in Chapter 3, these initial labels are invariably over-generalisations as the domain of the referent has not yet been sufficiently narrowed by parental and other speaker guidance. A number of similar events or objects may be included, which although separable from the point of view of an adult, represent the same concept to the child. It thus requires the same behaviour—the expression of the output verbal behaviour associated with the concept.

Language usage is social in that the child is encouraged to communicate. The child will actively seek to have their needs met by the use of communicative sounds. But the

learning is essentially automatic—the child learns the sequences both of sounds within words, and the frequencies and associations of particular words in sentences, without the need for any active participation. The ABC learning mechanism is active *all* of the time, and does not require any *effort* on the part of the child.

Much of the child's early speech is directed toward maintaining contact with caregivers and getting others to do things (Gleason, Hay & Cain 1989, page 176).

Figure 5.1 (b) indicates the learning of labels to be associated with previously learned sensorimotor behaviours.

symbolic conceptualisation

Once the basis for referencing and language is bootstrapped on top of the perceptual attractors, the linguistic component and the perceptual component interact with each other to provide richer forms of attractors. It is impossible to separate the two—the process is really one of coordination of both language and perceptual mechanisms.

Whereas initially, labels referred to particular objects or events, their usage now is extended to include social conceptualisations.

According to Luria (1982, page 45), the word as the basic unit of language has two functions in the reflection of objective reality. Luria classifies word meaning associations into two systems—meaning and sense.

Sense refers to the unique system of relations and associations that may be pertinent to a particular speaker in a particular situation—basically the same as *connotative* meaning. In contrast to this, *meaning* is that usage of a word which is the same for all speakers of a particular language—a *social-communicative* meaning. These meanings form a relatively stable system of associations and parallels our usage of *denotative* meaning. Thus the word's system of meanings includes both objective and subjective systems of situational and conceptual associations.

The initial and most basic usage is the word's *referential* function. The use of the word in this case is to substitute for an object, an act, a property, or a relationship. This usage is extended when the child learns to deal with things that are not necessarily present. This ability to reference things apart from their physical presence permits the child to extend their use of words to include *symbolic* references.

Within the ABC model, this ability is achieved when word usage is extended from

particular sensory inputs to internal conceptualisations. For example, if sufficient and varied dogs are seen and labelled, the word *dog* will come to refer to the internalised perceptual attractor associated with the general case rather than a particular dog.

This shift from the particular to the general enables the child to act internally, to be able to use internal concepts in self-talk. We discussed the mechanism briefly in Chapter 3—the child, through play, is able to ‘pretend’ and attribute properties of actual objects and events to abstract models, such as using a stick as a pretend horse.

Luria calls this second function of the word its *categorical* meaning. This designates the word’s abstracting and generalising function which results from the word’s capacity to analyse objects, and is accomplished by “introducing [the object word] into a system of complex associations and relations” Luria (1982, page 37). Further, “a word is a *unit of thought*. It is a unit of thought because the powers of abstraction and generalisation are the most important functions of thinking” (Luria 1982, page 38).

Consequently, it is as a result of the word’s second function that humans are able to achieve “that *leap from the sensory to the rational world*, which is essential for human consciousness” (Luria 1982, page 41). It also allows humans to transmit knowledge from one person to another so that the experience of previous generations can be acquired.

Luria maintains that the meaning of a word continues to develop through ontogenesis although it may retain the same particular referent. As the word meaning changes, it is accompanied by psychological changes so that the evolution of word meaning is also the evolution of the capacity of consciousness to reflect reality.

While initially the chief characteristic of word meaning is affect, before moving on to refer to a concrete image, it is in the later stages of language learning that a system of abstract and social connections are learned that stand behind the word. It is in this latter stage that the word enters into a system of categories and acquires what Luria refers to as its *paradigmatic* character, implying its “meaning is situated in an hierarchical system of abstract oppositions” (Luria 1982, page 52).

The abstraction of words away from particular referents thus enable the incorporation of ‘symbol-processing’ within language. Words become more abstract, but may still maintain associations and linkages to particular referents.

These symbols are then able to be linked into temporal sequences to produce language. The content of the utterance is based on meaning, context and the transitional

associations that have been learned for word transitions—the so-called ‘grammar’. Further, self-talk—the internalised alternative behaviour to speech—is then able to be developed using the same symbolic linkages. These stages are indicated in Figures 5.1 (c) and (d).

language and thought as motor behaviour

In contrast to the claims of Chomsky, we propose that there *are* common learning mechanisms in the brain and that these basic learning and motor behaviour mechanisms have been recruited for language. The learning and motor action mechanisms within the ABC model are universal.

Locke (1978) supports this view in stating that “it would seem that speech and language respond to the same neurological principles that govern all motor behavior.” Lieberman (1984, page 1) goes even further, stating that “aspects . . . of human language derive from neural mechanisms which also structure all other aspects of human cognitive behaviour, and which derive from homologous mechanisms that structure the cognitive behaviour of other animals.” Lieberman (1984, page 35) also suggests that the structure of the neural mechanisms imply that they “first evolved to facilitate motor control [and] now also structure language and cognition.”

Also, Allott (1991, page 124) maintains that “language was constructed on the basis of a previously existing complex system, the neural motor system”, and further that “language is completely analogous to skilled motor action.”

For these researchers, amongst others, language is viewed as but one expression of a set of organisational principles which underlie all neurological function and its overt expression in behaviour.

The mechanisms of the ABC model suggest that language is a motor behaviour, just as much as hitting a tennis shot. The connection with speech is obvious as speech requires the manipulation of voice and diaphragm muscles. But the ABC model also suggests that the same learning mechanism used to learn the tennis stroke is used in learning the temporal sequences of language.

The ABC model goes much further in suggesting that linguistic *thinking* (self-talk), being internalised language, may also be considered as a motor behaviour.

symbol-grounding & intentionality

Bechtel (1994) suggests that “intentionality is the characteristic that mental states enjoy of having content, or being *about* something. In contemplating a rose, one’s thoughts are about the rose.”

The ABC model affords a direct connection between language-based concepts (which are able to be expressed either through speech or self-talk), and the associated perceptual attractors. Thus the ABC model provides a solution to the problem of both symbol-grounding (the attachment of meaning to symbols) and intentionality (the attachment of meaning to mental states—essentially, thoughts).

The symbol-grounding problem was described by Harnad (1990) with the questions: “How can the semantic interpretation of a formal system system be made *intrinsic* to the system, rather than just parasitic on the meanings in our heads? How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything other than other meaningless symbols?”[†]

Harnad maintains that symbolic representations must be grounded bottom-up in non-symbolic representations. The two forms of ‘grounding’ used in the ABC model are perceptual attractors and symbolic (linguistic-based) temporal sequences. In the ABC model, feature detectors are used for both sensory inputs which lead to perceptual attractors, and for external auditory (speech) inputs. In a sense, it is impossible to separate these two forms of grounding as they both use the same mechanisms coincidentally and contiguously.

compositionality

According to Bienenstock & Geman (1995), compositionality “... refers to our ability to construct mental representations, hierarchically, in terms of parts and their relations.”

According to some, mental representations have a *combinatorial syntax and semantics*, and as the power of the digital computer arises in part from the fact that it is designed to be an extreme example of this organisation (a programming language operating on data is the prototype of the classical view), the computational view is preferable.

[†]See also Harnad (1992, 1993a, 1993b, 1993c, 1994) and Harnad, Hanson & Lubin (1994).

Fodor goes as far as to suggest that the need for compositionality means that cognitivism is “the only game in town.” For example, Fodor & Pylyshyn (1988) argue against connectionism on the basis that: (1) compositionality and structure sensitivity are necessary to support cognitive processes, and (2) only symbolic representations are capable of doing so. Their reasoning suggests that cognition must have computation and syntactic compositionality as in a digital computer to produce the ‘symbol processing’ of language.

However, as we showed in Chapter 2, the ABC model (and the SOMA and LAPS modules in particular) exhibit compositionality.

language & society

Gleason et al. (1989, page 172) suggests that “language is a social activity and why and how children speak is influenced by social-motivational factors based in each child’s own emotional needs (e.g., attachment, affiliation) and the external pressures of his or her social world.” This comes about because “the infant’s predisposition for social interaction (e.g., the psychological need for attachment and affiliation, an early preference for human faces, early recognition of the human voice) becomes a powerful motivator for the infant to imitate, interact with, and eventually talk to other human beings” (Gleason et al. 1989, page 176).

Language as social behaviour has been studied by many, including Vygotsky ([1962] 1986, Cole et al. 1978) and Piaget ([1923] 1960, Youniss & Damon 1992). The philosophers Gadamer and Heidegger also both argued that our ability to think and to give meaning to language is rooted in our participation in a society and a tradition.

Maturana (1978, page 50), in looking at language from a biological perspective, suggests that linguistic behaviour is behaviour in a *consensual domain*. For example, Maturana (1980, page 41) maintains that:

The linguistic domain as a domain of orienting behavior requires at least two interacting organisms with comparable domains of interactions, so that a cooperative system of consensual interactions may be developed in which the emerging conduct of the two organisms is relevant to both The central feature of human existence is its occurrence in a linguistic cognitive domain. This domain is constitutively social.

Maturana (1978) suggests a *structural coupling* between individuals in a society. Meaning is relative to what is understood through a shared tradition of discourse, and through interactions we share meanings. The mutual coupling is among the language users and is not a coupling of an individual and some external reality.

Children's social propensities lead them to take on their parents' (and hence, society's) values; thus, socialisation is accomplished in large measure through language.

Many believe that words are definitions, that they are (relatively) precise and the same meaning is shared by all in society. But this is clearly not true. Within the ABC model, words are attachments to initially sensorimotor concepts and then to more abstract associations of meaning. As each person undertakes a unique set of experiences in building up these word meanings, each meaning will be different. A general consensus in meaning is achieved through individual comparisons and differentiations with others in society.

Consider the abstract words such as *love* and *beauty*. These are very subjective concepts and each of us may have a different meaning for the words, and these may in turn be different again from any lexicon or dictionary entry. Misunderstanding often arises in dialogue between people because of these very differences in meaning.

language & thinking

Whorf contends that "language is not simply a reporting device for experience but a defining framework for it" (Whorf 1956).

It is clear that the interaction of words and perceptual concepts greatly extends the cognitive abilities of humans. For example, Baker & Cantwell (1982, page 286) maintains that "there is mounting evidence that language plays a role in the development of concepts, thought, play, socialization, self-image, humour, memory, reading, and education."

Baker & Cantwell (1982, page 287) add that "the role of language in the development of the child seems to be in the formation of concepts. Although there is some controversy regarding whether children merely use labels for concepts that they have already learned or whether their using labels actually aids in acquiring concepts, there is a growing body of literature indicating that language does aid in concept formation."

In reviewing the literature on language acquisition and cognitive development

Schlesinger (1977) concluded that there is an interaction between these two factors which contributes to the development of each.

Children not exposed to language or deprived in learning language in some way often show deficiencies in cognitive abilities. Inhelder (1976) observed that children with language problems seem to develop age-appropriate concrete operations of thought but often have deficiencies in figurative symbolism. Reports of cases of socially-deprived or socially-isolated individuals (sometimes referred to as *wolf* or *feral children*) indicate that they appear to be cognitively deficient (for example, see Curtiss 1977). Their lack of language abilities seems to greatly diminish their cognitive abilities.

The interaction between the child's own language capacity and cognitive ability is the basis of most Piagetian views of language development.

Maturana (1978, page 61) puts it that "we literally create the world in which we live by living it. If a distinction is not performed, the entity that this distinction would specify does not exist; when a distinction is performed, the created entity exists in the domain of the distinction only, regardless of how the distinction is performed."

Language interacts with and extends our sensory perceptions. By grafting words to describe objects, properties, events, features and so on, we direct our attention to those feature which differentiate each particular referent. As put by Harnad (1982), "what you cannot *tell apart* you cannot *tell*"—that is, differences must be found in the world in order to differentiate a new concept, a new term.

The interplay of perceptual and linguistic mechanisms within the ABC model allows for a complex and detailed conceptualisation of the world, greatly exceeding that which could be achieved without language. In this sense, language enhances awareness.

language & consciousness

Luria (1982) suggests that natural animal languages are not true languages but only quasi-languages. This is because they do not allow for abstract thinking or analysis, nor are they able to place an object or quality into a system of relationships and associations. Luria maintains that natural animal languages can only express "a condition or what the animal is going through" Luria (1982, page 29).

So, rather than conveying information, the animal 'language' merely serves to infect others of its kind with the same feelings that the animal which is expressing the sign

is going through. This is in keeping with Luria's idea that animal consciousness is constrained to the sensory level.

It is Luria's contention that the structure of consciousness also evolves and parallels the evolution of word meaning so that (Luria 1982, page 53):

During the earliest stages of ontogenesis, consciousness has an affective character. During the next stage, it begins to assume a concrete character. Words, through which the world is reflected, evoke a system of practically actuated connections. It is only at this final stage that consciousness acquires an abstract verbal-logical character, which differs from the earlier stages both in its meaning structure and in psychological processes, although even at this stage the connections that characterise the previous stages are covertly preserved.

For Luria, the two decisive factors which enabled the transition from animal consciousness (sensory) to human consciousness (sensory and rational) are the division of human labour (which provided a need for the exchange of information) and language. Luria (1982, page 27) states:

Language, in the course of social history, became the decisive instrument which helped humans transcend the boundaries of sensory experience, to assign symbols, and to formulate certain generalizations or categories. Thus, if humans had not possessed the capacity for labour and had not had language, they would not have developed abstract "categorical" thinking.

While we agree with Luria in the importance of language in the development of consciousness, we propose an extension to his ideas.

If we were asked to specify the difference between animals and humans as regards language; that is, why are humans the only animals to use language, we might hazard a guess and suggest that it is the mechanism of self-talk. After all, non-human mammals with good statistical learning and computational capacities do not develop language (Gallistel 1990).

While the social division of labour may have provided the impetus, it was the evolution of the extra recurrent linkage between an extension of the speech motor area

(now called Broca's area) and an extension of the auditory input areas (now called Wernicke's area) which enabled self-talk to come into existence.

Language and self-talk then were both able to simultaneously bootstrap themselves on top of the sensorimotor perceptual attractors. Language without self-talk is reactive in the sense of animal languages. Self-talk, though, can be self-generating, and so it can accelerate the rate of learning and conceptualisation of the individual (in association with others in society).

As we discuss later in Section 5.3 where we examine consciousness, a primary component of consciousness is self-talk. If we are unable to report an event to ourselves (through self-talk) then we are generally unaware of it; that is, we are not conscious of it. It seems that the self-talk recurrent loop provides the mechanism not only for internal thinking, but also provides the missing ingredient to enable external language.

5.2 Implicit and Explicit Learning

Massaro (1995) defines implicit learning as "the putative process by which we learn the structure of a domain without intending to do so, and without knowing what it is that we have learned." Lewicki et al. (1987, page 523) add that "subjects are able to acquire specific procedural knowledge (i.e., processing rules) not only without being able to articulate what they have learned, but even without being aware that they have learned anything."

Schacter, McAndrews & Moscovitch (1988, page 243) define implicit knowledge as "knowledge that is expressed in performance without the subjects' phenomenal awareness that they possess it", whereas explicit knowledge is "expressed knowledge that subjects are phenomenally aware that they possess." Schacter et al. use the phrase "failure to gain access to consciousness" when describing situations in which implicit knowledge is expressed in the absence of explicit knowledge.

An obvious example, as discussed in the previous section, concerns the learning of a language. We are able to learn to recognise and to produce utterances that are

‘grammatical correct’, and yet be unable to state what the ‘rules’ of the grammar are. †

Reber (1967) demonstrated that the learning of grammars could be studied empirically in the laboratory through the use of artificial grammars. Subjects who were exposed to strings of letters generated by a finite-state grammar could implicitly predict letters in the sequence even after only a few minutes of exposure to the grammar. The subjects were able to classify new stimuli as either being grammatical or not, but were not able to articulate the rules of the grammar (Reber 1967, 1989, Reber & Allen 1978). ‡

Jiménez & Cleeremans (1994) asked subjects to explicitly predict the next element of a sequence. They concluded that the reaction time task was “tapping genuinely unconscious knowledge—unconscious not just in the sense that subjects did not know that they had the knowledge, but in the sense that the relevant forced-choice tests did not elicit it.”

Recently there has been a resurgence of interest in the topic, and a number of paradigms are now used to test implicit learning, such as artificial grammar learning (Reber 1967, 1989), sequential pattern acquisition (Cleeremans & McClelland 1991, Nissen & Bullemer 1987) and process control (Berry & Broadbent 1984, 1987, 1988). See also Berry (1994) for a review.

Much of the intense debate over implicit learning has concerned the extent to which the learned knowledge is unconscious. Some workers deal with the phenomenon of implicit *learning*, whereas other researchers discuss implicit *memory*. For example, Schacter (1992) defines implicit memory as “an unintentional, nonconscious form of retention that can be contrasted with *explicit* memory, which involves conscious recollection of previous experiences.” Typically, explicit memory is associated with recall and recognition tasks that require intentional retrieval of information from a specific prior study episode, whereas implicit memory is assessed with tasks that do not require conscious recollection of specific episodes. For more than a century, similar distinctions have been made between conscious and unconscious memories.

A number of researchers have examined the use of neural networks in modelling implicit

†Altman, Dienes & Goode (1995) examine implicit learning in language.

‡We examine a representative finite-state grammar used by Reber in Appendix B.

learning. For example, Cleeremans & McClelland (1991) have used a simple recurrent network to study the implicit learning of sequences. Cleeremans (1993) has explored and tested a variety of connectionist models of implicit learning, including the SRN model of Elman (1990).[†]

Berry (1994) provides an excellent review article that discusses the history and overall results obtained in the last quarter of a century. Hirst (1995) also provides a review.

The major point to be made for the current discussion is this: many researchers maintain that at least two different kinds of memories exist—implicit memory, in which perceptual representations are stored and retrieval occurs without awareness, and explicit memory, in which more elaborated representations are stored and retrieval occurs with awareness (Schacter 1987a, 1989b, Tulving & Schacter 1990).

These are the two forms of learning that are found in the ABC model. As discussed in the previous section where we examined language and the ABC model, perceptual

[†]See also Phaf, Mul & Wolters (1994) and Cleeremans (1994). In empirical studies on humans, Cleeremans (1993) found good agreement between the performance of human subjects and the SRN model in learning.

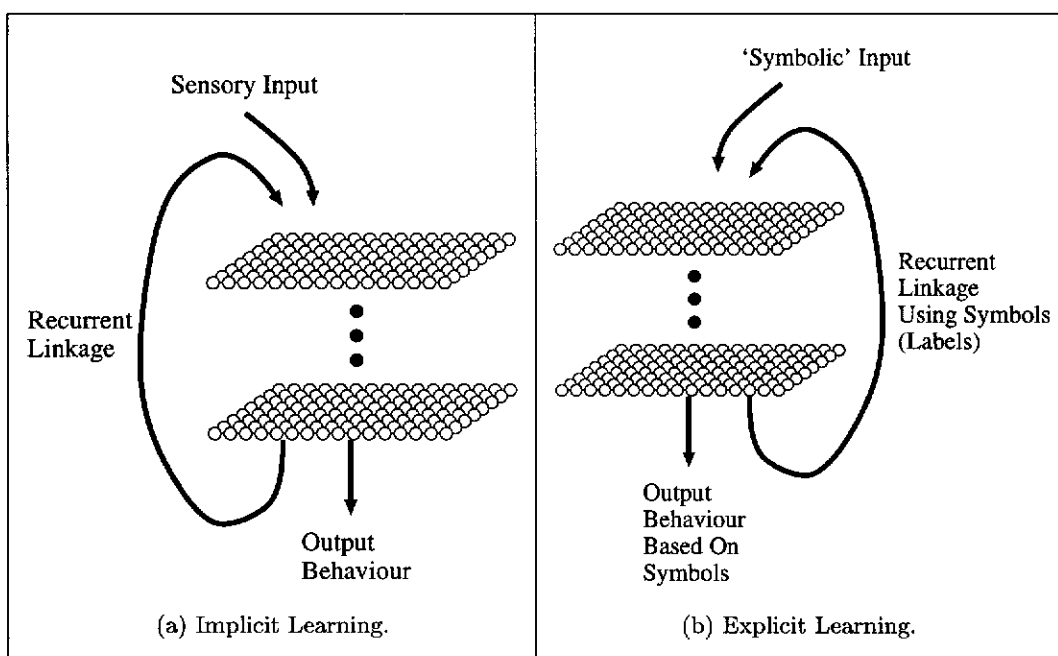


Figure 5.2: Flows for Implicit and Explicit Learning.

learning provides a linkage between sensory inputs and behavioural outputs that does not necessarily involve any component of language. This is shown in Figure 5.2 (a). In this figure, the two SOM sheets are to be taken as representing the full ABC structure separating sensory inputs and behaviours, with the recurrent loop indicating the multiple recurrent loops within the model.

The implicit learning of the perceptual attractors allows for the learning of behavioural skills, such as learning to play tennis through observation or learning the piano 'by ear'. The resulting perceptual attractors and temporal sequences are (at least initially) not connected with any linguistic attractors (or linguistic sequences), and so are not 'reportable'; that is, are not available to linguistic awareness. Linguistic awareness here includes both external speech and internal self-talk.

Explicit learning, on the other hand, involves linguistic attractors and sequences, and is indicated diagrammatically in Figure 5.2 (b). The learning either involves the retention of a new linguistic sequence (for example, learning a new 'fact') that is available for later reporting, or it involves the learning of new combinations of bodily skilled sequences that are already attached to linguistic attractors, and are thus available for later reporting.

5.2.1 The Two-Store Model of Memory

The current dominant model of memory in the cognitivist paradigm is the two-store model introduced by Atkinson & Shiffrin (1968). This model separates memory into a temporary short-term store (STS) and a more permanent long-term form (LTS).

According to Bourne et al. (1979), short-term memory is deemed to be what we are aware of at any given moment. Information may be transferred to long-term memory, and so pass from consciousness (current awareness), yet is normally still able to be remembered (reportable) should the need arise.

Figure 5.3 shows a (somewhat simplistic) diagram of the form of the memory system accepted by most cognitivists. It shows the three divisions originally proposed by Atkinson & Shiffrin (1968). More recent versions (Atkinson & Shiffrin 1971, Shiffrin

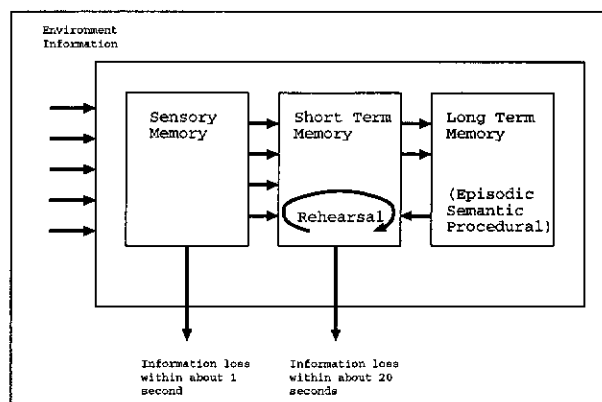


Figure 5.3: Short-Term and Long-Term Memory (Taken from Bourne et al. (1979, page 9)).

1975, 1976) combine the sensory memory and the short-term memory into a single component, and the short-term process is now not viewed as a physiologically separate structure so much as the temporarily activated portion of the long-term store (Raaijmakers 1993).[†]

Alternative theories have been proposed to overcome some perceived difficulties of the two-store model, including the levels-of-processing framework of Craik & Lockhart (1972) and the working-memory model proposed by Baddeley & Hitch (1974). Raaijmakers (1993), in turn, criticises these models in favour of the two-store model.

The point to be made here is that current cognitivist models of memory take no account of implicit memory. All memory, according to these theories, is either currently available to be reported (STS), or may be later retrieved (from the LTS) for explicit reporting. All memory is assumed to be either already in some linguistic form, or is readily converted to a linguistic form for reporting.

5.2.2 Implicit Awareness and Brain Damage

During the 1970s and 1980s, a number of researchers were able to show that various kinds of brain-damaged patients exhibited preserved access to nonconscious or implicit knowledge, despite a profound impairment of conscious or explicit knowledge.

[†]A recent update of the two-store model is the SAM model of Raaijmakers (1993).

Blindsight is perhaps the most well known of these syndromes. Here patients with lesions to striate cortex deny any conscious perception of visual stimuli within their scotoma, yet nonetheless can “guess” the location and other attributes of the target light (Weiskrantz 1973).

There are a number of neurological syndromes which exhibit a separation between implicit and explicit learning. As stated by Marshall & Halligan (1988, page 766), where “patients may show tacit awareness of stimuli that cannot be consciously recalled or identified.” These syndromes concern brain damaged patients and the apparent severing of the ‘normal’ access to conscious awareness, and include:

amnesia

an apparent inability to remember or learn,

blindsight

the residual visual capacity of cortically blind patients to locate events without explicit awareness,

prosopagnosia

the inability to recognise and identify familiar faces,

dyslexia

difficulties with reading and writing,

aphasia

problems with syntactic and semantic processing of language,

hemineglect

the inability to be consciously aware of the left (or right) hemisphere of space,

anosognosia

the loss of ability to recognise or to acknowledge an illness or bodily defect.

All of these cognitive *disfunctions* challenge, to some degree, the commonly held view of the way in which conscious awareness is linked with cognitive processes. They demonstrate an apparent dissociation between our ability to explicitly perform various

cognitive functions, and an implicit access to these same functions (Hirst 1995).[†]

We briefly look at some of these defects in the following sections, and explore how the ABC model may give us a measure of understanding of the deficits.

The evidence discussed in the literature has concentrated on explicit/implicit knowledge in neuropsychologically impaired patients. However, there is evidence that suggests that (at least some) implicit information is also accessed in normal cognitive processes. In various studies, patients and normal controls have shown identical or similar patterns of performance on implicit tests (Schacter et al. 1988). For example, people may respond to a stimulus such as an auditory or visual cue, but claim to be unaware of the stimulus itself. The stimuli include subliminal perception and priming (Tulving & Schacter 1990, Kihlstrom 1987), and galvanic skin responses (Tranel & Damasio 1985).

In the syndromes discussed below, the patients have no *explicit* knowledge of their *implicit* abilities, and perform at or near chance levels if explicit access is required. Schacter et al. (1988, page 264) consider and dismiss various explanations such as *conservative response bias* (in which patients err on the conservative side when reporting explicit knowledge), *disturbance of consciousness or language* (in which the patient may be consciously aware of the information, but is unable to gain access to language production mechanisms), and *different systems for implicit and explicit knowledge* (which considers the possibility that there are two distinct and *dissociable* neural subsystems within each domain; that is, two of everything).

Schacter et al. (1988) highlight two important considerations that must be taken into account in any explanation of these phenomena: first, the *generality* of the dissociation across many different sensory systems and cognitive functions; and second, the *selectivity* of the dissociation—an impairment in one system does not necessarily imply a global disorder of conscious awareness. They “hypothesise that (a) conscious or explicit experiences of perceiving, knowing, and remembering all depend in some way on the functioning of a common mechanism, (b) this mechanism normally accepts input from and interacts with a variety of processors or modules that handle specific types of in-

[†]Analogous phenomena have been observed in normal subjects—see Schacter (1992) for a review. See Prigatano & Schacter (1991) for an overview of contemporary approaches to rehabilitation.

formation, and (c) in various cases of neuropsychological impairment, specific modules are disconnected from the conscious mechanism” (Schacter et al. 1988, page 269).

Other researchers, such as Weiskrantz (1973), support this high-level operator hypothesis. Weiskrantz suggests the concept of a *complementary system*, a specific (executive) conscious mechanism which receives input from other memory and sensory modules. In these models, the explicit/implicit dissociation result when there is a *disconnection* between a specific module and the conscious mechanism, or a damaged module may only be sending degraded outputs to the conscious mechanism that are sufficient for implicit but not for explicit expressions of knowledge.

The view of memory taken in this thesis is somewhat at variance with these current models. In our view, implicit ‘knowledge’ and behaviour is brought about via perceptual learning—the principle mechanism of cognition. Perceptual attractors and temporal sequences are formed from incoming sensory vectors, and result in output behaviours that are not attached to verbal awareness. The subject is not linguistically aware (through the process of self-talk or more commonly linguistic *thinking*) of the connecting mechanisms between sensory inputs and output behaviours. Remember that we include speaking and hence self-talk as an ‘output’ behaviour, even though self-talk is internal to the brain.

Explicit ‘knowledge’, on the other hand, is knowledge that is available to be reported because labels are attached to these perceptual attractors.

Attached labels are not the default (as is the case in the cognitivist model) but must be learned. If, due to brain damage later in life, certain of these learned labels are detached from their associated perceptual attractors, or if the processing of the temporal sequences which lead to the verbalisation of the labels or perceptual attractors is damaged, then the process of verbal reporting is curtailed. We then have no access to the perceptual attractors in verbal terms that are available for placement into sentences for verbal reporting.

As discussed previously, verbal reporting in the ABC model includes both the external medium of speech and the internal medium of self-talk. If we are unable to report an experience to another person via speech, then we are unable to report it to ourselves

via self-talk. We are thus not verbally (or as we discuss later, consciously) aware of the phenomenon that has become detached.

To reiterate, perceptual learning is the mechanism whereby an implicit connection is formed between sensory inputs and resulting behavioural outputs. Perceptual attractors are the mechanism for implicit 'knowledge'.

Verbal attractors are associated with certain perceptual attractors and perceptual temporal sequences, and may also form temporal sequences. Output behaviour from such verbal attractors and sequences is speech and self-talk, which are the mechanism of reporting to others and ourselves respectively. The attached labels are able to be formed into verbal statements (including self-talk) to *describe* or *reference* the specific experience or object, and this is the mechanism of the explicit phenomenon.

In the following sections, we examine a number of implicit/explicit deficits in more detail, and indicate possible dislocations in the perceptual/verbal linkage that might account for the deficit. Obviously this is an initial (and perhaps speculative) attempt at examining these deficits within the framework of the ABC model, and more work is required to give a full account. This is left to future research.

blindsight

Blindsight refers to "visual capacity in a field defect in the absence of acknowledged awareness" (Weiskrantz 1973, page 166). A patient with a scotoma may be able to point to a light source displayed within the scotoma, yet be totally unaware that he is perceiving anything. The patient believes that he is simply 'guessing', but forced-choice methods clearly show the performance is significantly above random. The patient is not explicitly aware of the location of the light, but is asked to 'go along' with the experiment and guess. Decisions are not based "on the subject's own verbal report of his visual experience, which was often lacking" (Weiskrantz 1987, page 77).

Blindsight is found in patients with an ablation to primary visual cortex, resulting in cortical blindness for a part of the visual field.

Various experiments have shown that discrimination of simple shapes (for example, an 'X'-shaped light source versus an 'O'-shaped one, or curved versus straight edged triangular sources), orientations (vertical, horizontal, diagonal), and gratings of

reasonably high spatial frequencies can be achieved in some patients.

The patient does not report a sensation of 'seeing', but "if pressed, he might say in some tests, but by no means in all, that he perhaps had a 'feeling' that a stimulus was approaching or receding, or was 'smooth' (the O) or 'jagged' (the X). But he always stressed that he saw nothing in the sense of 'seeing', that typically he was guessing, and was at a loss for words to describe any conscious perception" (Weiskrantz 1973, page 31).

Weiskrantz (1980) maintained that some form of practice or shaping of the response is required for demonstrating blindsight localisation. Unless patients are initially trained on a relatively simple version of the localisation task, they may at first fail it. Accuracy in blind-field localisation reports increases over a few hundred trials to the point where it is indistinguishable from performance with intact-field presentations (Zihl 1980).

Initial attempts by some researchers to explain these phenomenon in terms of light straying into the unimpaired cortical areas, or the use of spared cortex, or other possible artifacts (Campion, Latto & Smith 1983) have since been dismissed by further experiments which control for these potential issues. Although it would seem impossible to totally discount the null hypothesis, it would seem clear that the evidence confirms that the syndrome is genuine.

Blindsight patients thus exhibit an *explicit* visual deficit, yet also show an *implicit* visual capacity within a scotoma. They are unable to explicitly provide a verbal report on the location of a light displayed in their scotoma, yet are able to implicitly locate the target by 'guessing' its location and pointing to it.

Experiments with monkeys have shown that after a unilateral striate cortex removal, the animals are able to recover much of their sensitivity to light stimuli in the hemianoptic field, and are able to move about freely and grasp objects. Their visual sensitivity is not totally recovered, however, and appears to be of a different *qualitative* nature than their pre-optive vision. They are able to detect and locate objects, but are unable to distinguish the objects by sight alone. For example, a particular monkey "was exquisitely sensitive in locating and picking up small objects, such as tiny specks of paper, but she could not distinguish a bit of paper from a small peanut until she touched them, although she responded to either with a fine sensitivity" (Weiskrantz 1973, page 18).

Following lesions of the superior colliculus, however, monkeys exhibit “impaired orienting responses towards visual stimuli and a reduction in responses to novel targets” (Weiskrantz 1973, page 162). Further, if the lesion also includes additional cortex lying anterior to the striate cortex, then residual vision is still present but is seriously degraded.

The evidence thus shows that areas other than the visual cortex are used in monkey vision, specifically the superior colliculus. In fact, the retina of the monkey projects to at least six anatomically distinct parts of the brain (Covey & Stoerig 1991, page 12).

Other studies have shown that the human visual system is much more dependent upon a functioning striate cortex. Any damage to the occipital lobe can produce much more severe symptoms in humans than in monkeys, sometimes producing absolute defects. †

One theory put forward as early as the late sixties (Schneider 1969) suggested that higher animals have two visual systems. A strong form of the two-visual system argument holds that the striate cortex is responsible for the *identification* of objects, whereas the mid-brain pathways (principally the superior colliculus) are responsible for mediating the *detection* and *localisation* of visual events, and to direct the eyes to the object for identification. These are the so-called *what* and *where* components of the visual system respectively.

The superior colliculus is a phylogenetically older, more primitive visual area than the cortex. With lower animals, such as frogs and fish, the superior colliculus is the *major* brain centre for vision. Cats with ablated striate cortex are able to move about relatively freely, monkeys less so, and ablated striate cortex presents a severe disability to humans.

In humans, over 80% of retinal connections travel to the visual cortex via the lateral geniculate nucleus. Most of the rest travel to the superior colliculus, although

†One possible reason given for this difference is the fact that the *geography* of the monkey and human brains is different. According to (Weiskrantz 1973, page 16): “in man, the striate cortex lies almost entirely buried within a sulcus on the medial surface of the brain, whereas in the monkey much more is exposed on the lateral surface. This means that in man it would be unusual for there to be a striate cortex lesion that did not also involve prestriate and other cortex, whereas in animal research a lesion restricted to striate cortex alone can be placed with some accuracy.”

as with the monkey, there are multiple other pathways to various areas of the brain. The superior colliculus also obtains inputs from the auditory and vestibular systems, integrating these components to provide overall orienting behaviour. †

It appears that in humans, the phylogenetically newer visual cortex has taken over the primary visual function. Thus researchers considered the possibility that blindsight in humans is facilitated by the superior colliculus, and the present findings “fit within the general framework of ‘two visual systems’ in allowing extrastriate routes a capacity for mediation of detection, location and orientation of a visual event, but requiring integrity of striate cortex (together with other more anterior cortical structures, especially the temporal lobes, to which striate cortex projects over several synapses) for its identification” (Weiskrantz 1987, page 90). ‡

In line with the ABC model, we suggest that the two-vision hypothesis is somewhat simplistic. We propose that the phylogenetically older superior colliculus, and the newer striate cortex *both* attend to visually locating items of interest to the creature (the so-called *where* component), and *both* attend to some form of sensory identification (the *what* component). However, it is only in humans that a *linguistic label* is attached to the identification process so that we are *linguistically* ‘aware’ of the particular object in that we are able to name it. Further, the linguistic labelling is a mechanism of the cortex only, so that the ‘what’ component of recognition occurs only within the cortex.

The interesting thing about blindsight according to the ABC model is not that the patients are able to point to objects within their scotoma, or to perform some limited recognition of the shape—presumably some other visual pathway supplies an alternative connection between the sensory inputs and the behavioural action of pointing (for example, see Cowey & Stoerig 1991), but rather that the patients are *unaware* of the target object; that is, the sensory inputs are not connected to any linguistic attractor.

We are able to provide an explanation for blindsight within the ABC model by

†See Appendix Section A.3 for details on the secondary visual pathways.

‡Studies by Stoerig & Cowey (1989) suggest that the previously suspected subcortical pathway via the superior colliculus may need to be revised. They show that blindsight patients are able to discriminate colours, and note that the spectral sensitivity curves indicate colour opponent processes. As chromatic opponency has never been detected in the midbrain visual pathways, this suggests that the remaining retinal projections to the thalamus and from there to the extrastriate cortex may be involved.

first reiterating that there are a number of visual pathways (which evolved at various periods or were adapted from existing mechanisms), but that the neocortex seems to be the only brain component which contains the human language facility. Once again, when we refer to language we also refer to self-talk. The ABC model suggests that visual *recognition* is achieved when sensory inputs excite language attractors, so that we are able, through self-talk (verbal thinking) to tell ourselves the label of the thing being recognised.

Blindsight patients suffer an ablation of the striate cortex (area V1). This means that all of the 'downstream' areas connected to the ablated area will also suffer a defect, including the attachment of a label via associative links. The mechanism of verbal attractors being connecting to visual cortical areas to enable verbal label recognition is thus disabled within the area of the ablation, and so the patient will not be able to verbally (including self-talk) recognise the object. The patient will then not be 'conscious' of the object as he is not able to put a label to it in order to 'think' about it.

However, the extra-striate visual pathways will provide locational information for the object, so that if the patient is asked to 'guess' and point at the object they will be able to do so. The fact that the patient is still able to point to the object suggests that the pointing capability, learned during development, incorporates the cortex *and* the other visual areas in a coordinated whole, so that when the cortex component is disabled the other visual areas can provide adequate pointing capabilities. The requirement for a training period suggests that additional associations need to be made (or enhanced) between what remains of the connections within the cortex to the extra-cortical visual pathways and the 'pointing' behaviour attractor.

The ABC model separates the processes of *visual* recognition—via visual sensory attractors and connected motor behaviours—and *label* recognition—the process of 'retrieving' a previously attached label in order to name an object. It is only through the attachment of a label to the perceptual attractor for an object that we are able to be 'aware' of an object and to report it, to others or to ourselves via self-talk.

Other labelled connections remain in place. For example, the patient is fully (verbally) aware of the act of reaching and is able to report on it; that is, the perceptual attractors associated with the patients proprioceptive (and spared visual) attractors re-

main linked to reportable verbal attractors. The patient's motor output thus appears to be voluntary and the subject is certainly aware of what movements he is making.

hemineglect

Hemineglect (or visuo-spatial neglect) "is characterized by an impaired ability to attend to the egocentric side of space contralateral to the damaged hemisphere" (Schacter et al. 1988). A classic case of hemineglect is described by Sacks (1986, page 73). The patient was a woman who suffered a massive stroke which damaged her right cerebral hemisphere. She appeared to totally lose the concept of 'left', and was unable to attend to anything on her left side. For example, she could only see half of her face, and would apply make-up just to the recognised right side.

She would complain of small meal portions because she could only attend to the right half of her plate. She could not be directed to the left of her plate, as the concept of *left* no longer had any meaning to her. Once shown her problem, she could intellectualise her deficit, but it seemed impossible for her to look or turn left. She was forced to develop a strategy (a learned behaviour) of revolving on her chair to the right to eat her meals, each spin allowing her to eat half of the meal remaining—the gastronomic version of Zeno's paradox.

Marshall & Halligan report the case of another patient with sustained right cerebral

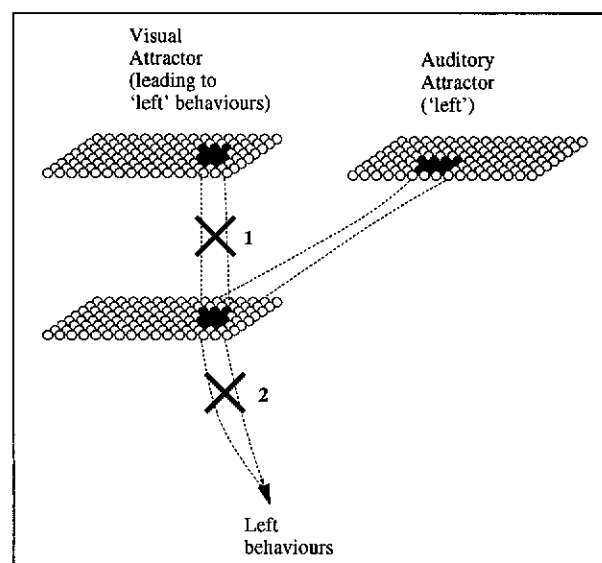


Figure 5.4: Hemineglect.

damage resulting in left-sided neglect. When “presented simultaneously with two line drawings of a house, in one of which the left side was on fire “she judged that the drawings were identical; yet when asked to select which house she would prefer to live in, she reliably chose the house that was not burning” (Marshall & Halligan 1988, page 766). Here again, an explicit recognition is lacking, yet an implicit, covert knowledge is made available to her decision processes. The authors concluded that “the so-called ‘neglected’ stimulus can exert an influence upon cognitive functioning, albeit at some pre-attentional, pre-conscious level” (Marshall & Halligan 1988, page 767).

What we suggest is happening in the case of hemineglect experienced by the Sacks patient, is that the very ‘concept’ of *leftness* has been dislocated from the perceptual attractors and motor behaviours that are normally attached to the ‘left’ concept. The actions associated with turning left and so on have been detached from the verbal attractor for ‘left’ (and associated words which were also attached to the ‘left’ perceptual attractors). The verbal concept of ‘left’ is not attached to any appropriate behaviour because of the cortical damage—to her, ‘left’ produced no behaviour, and so *means nothing*.

The whole left hemisphere of the cortical visual system is not functioning and so the patient is unable to see anything in that hemisphere. The subsequent (downstream) linguistic attractors associated with the concept of ‘leftness’ that are associated with that visual hemisphere are thus no longer attached to any meaningful behaviour.

In case of the Marshall & Halligan patient, there is evidence that the alternate (non-striate cortex) visual systems could still ‘see’ the burning left side of the house, and thus make a determination that the house was unsafe through some affective association, but the attachment of the left hemisphere to the reporting verbal attractors is severed and so the patient is unable to report the differences in the two drawings.

As with the other disfunctional syndromes discussed in this section, the argument is perhaps somewhat sketchy at this stage, but it seems clear that the dislocation is with the linguistic attractors leading to an inability to report. The separation of perceptual and linguistic cognitive mechanisms suggested by this phenomenon is consistent with the structure of the ABC model.

prosopagnosia

Patients who suffer from prosopagnosia have an apparent inability to recognise and identify familiar faces—faces they should readily identify, such as a spouse, children and friends. Prosopagnosia is typically produced by bilateral lesions to occipito-temporal cortical regions, but a significant number of cases result from unilateral right-hemisphere lesions (Schacter et al. 1988). See Damasio (1985) for a review.

This syndrome is typically produced by lesions in the occipito-temporal cortical region, and the deficit may extend to other recognition problems beyond facial recognition. Recent evidence, however, has shown that prosopagnosic patients exhibit an implicit knowledge of the faces they fail to recognise explicitly. Studies measuring skin conductance responses (SCR) indicate significant correspondence between correct matching of names with faces of famous people and family members.

In one experiment, prosopagnosia patients were asked to choose from amongst five names to match a face (one being correct, the other four lures but of similar 'status' or 'occupation'; for example, all five names being famous actors. On this explicit task, the patients could do no better than random. Yet the SCR results showed a strong correlation with the correct name-face pairing.

Recognition using other means such as the sound of a voice, or a characteristic gait, is usually not impaired.

This and other experiments indicate that *implicit* recognition of faces is found in prosopagnosia patients. However, their *explicit* awareness "does not give rise to the phenomenal experience of familiarity reported by neurologically intact individuals" (Schacter et al. 1988, page 253). Some information about the faces is available, but the patient is unable 'to gain conscious access to these stored data' (Bruyer 1983, page 280).

Thus it seems that the face is 'recognised' by the visual system, but again the dislocation is with the label associated with the face; that is, the person's name.

amnesia

Experiments demonstrate that amnesiac patients with severe impairment of explicit memory can exhibit intact implicit memory (Schacter 1987b). "Amnesiac patients suffer from a severe and selective inability to remember recent experiences and to learn

various types of new information, despite preservation of most perceptual, linguistic, and intellectual skills" (Schacter et al. 1988, page 244). The problem usually results from lesions of either the medial temporal or diencephalic regions of the brain.

On tests of explicit memory recall, such as remembering prior events or actions, amnesiacs show extremely poor performance, some even failing to remember experiences just minutes old. However, it has been shown for some time that patients show implicit memory for recent experiences. Studies have shown that amnesiac patients perform relatively well, even normally, on a wide range of implicit memory tests that do not require explicit (or conscious) recall.

For example, in one experiment, patients were asked to perform a simple task of pressing a key beneath a light when the light was activated. An array of light-key pairs was used, with the lights flashing in a repeated pattern. It was shown that the amnesiac patients could respond more quickly to a regular pattern of light activations than to a random ordering, thus indicating that the patient had learned the sequence and could anticipate the next in the series. Yet when asked to explicitly remember the next in the sequence, they were unable to do so.

It has also been shown that amnesiacs can acquire skills gradually and slowly through implicit memory of single episodes, and "can acquire new factual information, even though they do not explicitly remember having learned any facts and claim no familiarity with the information that they do retrieve" (Schacter et al. 1988, page 247).

One group of researchers found that a densely amnesiac patient could learn complex factual skills in using a microcomputer to write and edit programs, yet the patient "did not explicitly remember having learned anything about the computer, and claimed at the beginning of each session that he had never worked on a computer before" (Schacter et al. 1988, page 247).

The skills learned by the patients in this case (operating, programming and interacting with the computer) seem to be tasks that are learned by the perceptual mechanisms of the ABC model. Extensive repetitions of the tasks were required across numerous learning sessions. However, the linking of these behaviours with explicit linguistic attractors and sequences does not occur.

anosognosia

Anosognosia is an unawareness and/or a denial of deficits, such as blindness, or the loss of the use of a limb. The syndrome usually follows hemiplegia, hemianopia, and some other neuropsychological disorders. The patients claim to be entirely unaware of the existence of the deficits that are all too obvious to others.

Researchers such as Weinstein, Cole, Mitchell & Lyerly (1964) and Weinstein & Friedland (1977) have shown, however, that anosognosic patients may possess some implicit knowledge of deficits that they deny explicitly. What seems to be missing is again an explicit linkage between the concept and label for their particular condition. Their world-knowledge and behaviour was learned previously with these components in place (i.e., their limb or other body parts), and the elimination of the contextualisation of their deficit means that it has no linguistic meaning and so is not able to be included in any reporting to others or themselves.

We grow up in intimate contact with our bodies and our limbs, and our being is concerned with this integrated whole. Should some of that intimate knowledge be destroyed, our self-knowledge of our bodies may be deficient.

All of these syndromes show some form of dislocation of the linguistic mechanisms of cognition from the perceptual mechanisms. Although at some level the patients are 'aware' of certain knowledge, their ability to put that knowledge into a linguistic form for reporting is curtailed.

In separating perceptual conceptualisation, and the linguistic component of cognition, the ABC model is able to partially account for these phenomena. Should the previously learned associations between the patients perceptually learned behaviours, and the subsequently learned (and linked) linguistic behaviours be severed in any way, then the sort of deficit observed in the implicit/explicit dislocation patients would be expected.

The primary source of evidence for the patient's lack of explicit or conscious knowledge is their reporting of their subjective experiences. As we discuss later in relation to consciousness, the patient may be aware (conscious) at some level (say at a somatosensory perceptual level) of some phenomenon, but if they are unable to report the phenomenon

in a linguistic form (speech and/or self-talk) then they are not ‘consciously’ aware of it. Self-talk (thinking) is such a large component of our self-understanding, that if it is disturbed in this way then our knowledge of the phenomenon is greatly diminished.

Schacter et al. (1988, page 266) consider and dismiss this possibility that “the patient is in some sense ‘consciously aware’ of the information that [they] have referred to as ‘implicit’, but are unable to express this awareness verbally.” However, their reasons for rejecting this possibility are not appropriate to the ABC model. They suggest that because language is reproduced in other circumstances, then it is unlikely that there is a “disruption or disconnection of language production mechanisms” (Schacter et al. 1988, page 266).

Rather than damage to a language *production* mechanism, we propose that some of the language *linkages* to perceptual attractors are disturbed—more a dislocation of linguistic *data* than damage to the *mechanism*.

Another objection given by Schacter et al. is that the phenomenology of these impairments is different to that of other neuropsychological syndromes such as naming disorders, in which the patient cannot explicitly produce the name of a familiar object in a picture. In the case of these *anomic* patients, they often have explicit access to many other kinds of information about the missing word, and they state they ‘know’ perfectly well what the object is and does.

However, we suggest that this syndrome is as a result of a dislocation within the linkages *between* linguistic attractors (words), rather than a dislocation between perceptual attractors and linguistic attractors. The latter dislocation would seem to be a much more severe and debilitating deficit, as it perturbs a fundamental means employed by humans in making sense of their world, that of self-talk about their learned knowledge of their being and their body.

Schacter et al. (1988, page 269) hypothesise that “conscious or explicit experiences of perceiving, knowing, and remembering all depend in some way on the functioning of a common mechanism”. The common mechanism of the ABC model is the labelling of perceptual attractors (that is, the linking of ‘linguistic’ attractors with perceptual attractors), and the subsequent linking of these linguistic attractors into the temporal

sequences of speech and self-talk.

Every concept which forms part of our understanding of our being must be learned—there are no inbuilt concepts. This means that there are no such things as linguistic simples—basic terms from which all other terms may be derived. All terms are formed by an association of a perceptual attractor and a linguistic attractor, with the perceptual attractor being formed first.

Thus concepts such as *left*, *my-arm*, or *my-wife's-name* have no meaning as isolated linguistic terms but must be associated directly with bodily behaviours. The concept of *left* must be linked with those learned behaviours and perceptual attractors that are found in a normal person which pertain to 'left' behaviours. If that linkage is broken, then these concepts lose their 'meaning'.

5.2.3 Split-Brain Patients and Hemispheric Specialisation

Schacter et al. (1988) suggest that it is instructive to compare the phenomena of implicit/explicit dislocation with some of those that are observed in split-brain (or commissurotomy) patients (Sperry 1966, Gazzaniga 1995a).

For patients whose corpus callosum—the massive fibre bundle connecting the two corti-

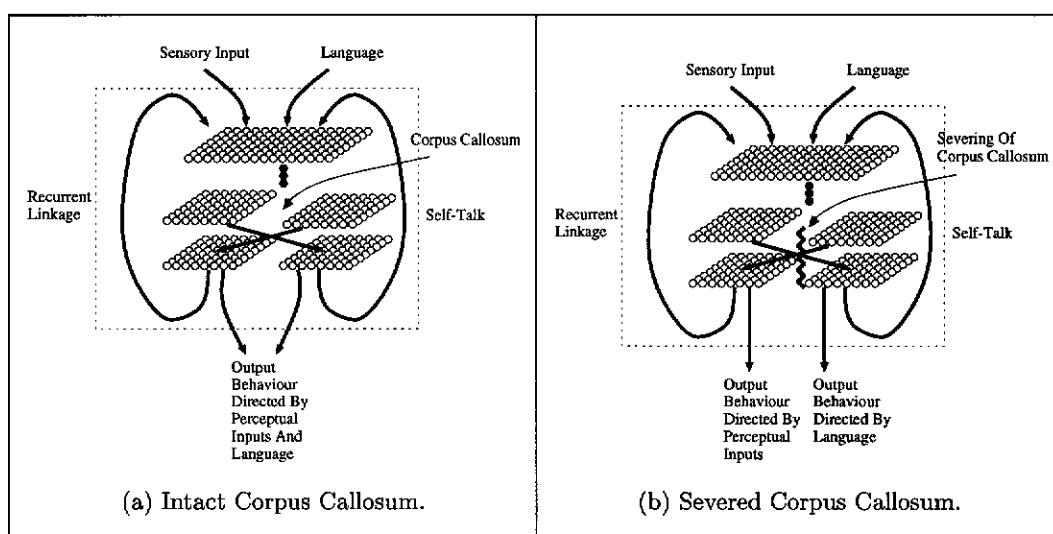


Figure 5.5: Commissurotomy.

cal hemispheres—has been cut, the left-hand side of the brain (for right-handed people) appears to not be aware of the activity in the visual system taking place on the right side, whereas in a normal person it is (Geschwind & Galaburda 1986).

For example, if a stimuli is confined to the left visual field and thus projected to the right hemisphere, the patient may not be able to verbally state what they have seen. However, if they are allowed to use their left hand to select the presented stimulus from a number of alternatives, they are able to do so with a high degree of accuracy (Gazzaniga & LeDoux 1978). Schacter et al. (1988, page 267) suggest that “the right hemisphere possesses extensive conscious awareness but has difficulty organizing a verbal response and thus cannot express its ‘awareness’ through language when it is disconnected from the verbal mechanisms in the left hemisphere.”

Crick & Koch (1991) maintain that this strongly suggests that some of the information associated with *consciousness* can traverse the normal corpus callosum. Further, it also suggests “that such information, with the exception of some emotional states, cannot be transmitted from one side of the cortex to the other by the subcortical pathways that remain intact in this operation.” Others maintain that the experiments suggest the possible existence of independent systems of consciousness in each hemisphere.

The view of this phenomenon consistent with the ABC model is somewhat different. In Figure 5.5 (a), the two major components of cognition that we have described previously, recurrently-linked perceptual attractors and the also recurrently-linked linguistic component of self-talk, are shown in simplified form. Rather than indicate all levels of the model, we instead show only three layers of SOMs. However, the principles remain the same.

In a normal person, as indicated in Figure 5.5 (a), the lower component maps of the model are of necessity split over the two hemispheres. Cross association linkage occurs through the corpus callosum.

Over the developmental period, the self-organisation of these lower maps will force similar components of the preceding input vectors to be located close together on the maps. This will have the tendency to separate the two sets of vectors—the sensory/perceptual attractor sourced vectors and the linguistic/self-talk sourced vectors—onto opposite

sides of the corpus callosum divide. For the majority of right-handed people, the linguistic component is concentrated in the the left-hemisphere, with the perceptual component found in the right-hemisphere. As stated by Lenneberg (1967, page 153), “[in] the absence of pathology, a polarization of function between the right and left takes place during childhood, displacing language entirely to the left and certain other functions predominantly to the right.”

Paterson & Zangwill(1944, 1945) provide compelling psychological evidence confirming Hughlings Jackson’s view (derived from clinical observations) that the right hemisphere *leads* in non-language visual ideational functions.

The fact that the right hemisphere is able to subserve the acquisition of speech in young children who have sustained early damage to the left, further enhances our hypothesis. In this case, the self-organisation of the SOM surfaces for both cognitive components is forced onto the one hemisphere. Zangwill (1978) points out that this phenomenon was in fact predicted by Broca (1861), who regarded both hemispheres as being essentially equipotential at birth.

Lenneberg (1967, page 181) found that “the equipotentiality of the hemispheres systematically diminishes from perfect equipotentiality (from birth to 20 months); to marked signs of reduction of equipotentiality (ages 11 to 14 years); to none for language functions (mid-teens to senium).” The self-organisational nature of the hemispheric specialisation is further supported by the observation by Dax, as early as 1836, that “although sudden destruction of the left hemisphere structures resulted in marked aphasia, slowly developing lesions in the same regions often failed to produce aphasia” (Smith 1978, page 141).

Theories of the specialisation of function of the two hemispheres have had a diverse history. Some have claimed that the left, language-dominated hemisphere is in some way superior and *in-charge* of the lowly right hemisphere. For example, the right hemisphere has variously been described as “nondominant, minor, subordinate, silent, and as the weaker, less clever, dependent brother” of the left hemisphere (Smith 1978, page 136). Some have even suggested that there are really two seats of consciousness within the brain, one associated with each hemisphere.

The “progressive specialization by the left hemisphere in [linguistic tasks], and by the right hemisphere in visual ideational and other nonlanguage cognitive functions” (Smith 1978, page 139) has a natural and simple explanation within the ABC model. It is also able to explain the role of the “right hemisphere to compensate for diminished language functions immediately following damage to left hemisphere mechanisms” (Smith 1971, page 203) as being a process of self-organisation on a SOM surface.

5.2.4 Skill Acquisition, Language and Thinking

Generally, in acquiring a skill—in learning to drive, dance, or pronounce a foreign language, for example—at first we must slowly, awkwardly, and consciously follow the rules. But then there comes a moment when we finally can perform automatically. At this point we do not seem to be simply dropping these same rigid rules into unconsciousness; rather we seem to have picked up the muscular gestalt which gives our behavior a new flexibility and smoothness. Dreyfus (1992, page 249)

In this section we examine the relationship between the mechanisms of motor skill acquisition, language and thinking. We show that the ABC model is able to account for the observation that humans are able to move from linguistic rules to skilled behaviours with practice.

Whereas infants must experience the world through sensory means and build up perceptual attractors before they are able to label these attractors, and then subsequently build up temporal sequences of these linguistic attractors into a language, the process of learning a skill by those already possessing language can move in the reverse direction. That is, we are able to begin with some linguistic (symbolic) rules to direct our behaviour in the initial stages, but then through practice, the subsequent perceptual attractors resulting from the repetition of the tasks will be able to perform the skill in their own right without the need for the linguistic instructions.

The usual starting point for obtaining a skill is the learning and manipulation of certain context-free elements. As pointed out by Dreyfus (1992), “this is the grain of truth for

the cognitivists model.” Thus, for example, a chess beginner will first learn the strict rules of the game, and these rules will invariably be in some linguistic form.

After one starts to obtain an understanding of the domain, however, one sees meaningful aspects, not context-free features. These aspects are invariably *not* in a linguistic form, but rather in the form of some perceptual or motor gestalt behaviours. For example, the advanced beginner chess player might observe some context-dependent characteristics such as an unbalanced pawn structure.

After having seen and played many games, and achieved a great deal of experience, the proficient chess player is able to see situations that tend toward a certain outcome. Certain aspects of the current situation will tend to stand out as significant. At the expert level of a chess grand master, the required response seems to ‘pop into’ their head, and they are able to see immediately what is required. The simple features and rules of the beginning player no longer have a role in such expert performance—other than to perhaps later explain a particular move. †

As other examples, we might be told how to swing a golf club by a teacher, or told how to drive a car. At first the corresponding skill is clumsy as we need to coordinate these language instructions with our muscle behaviours.

However, over time and with practice we appear to develop a ‘gestalt’ behaviour which just relies on the sensory inputs, and does not need the ‘rules’ any more—even though the rules might still be available within our memory (as temporal sequences of linguistic vectors).

Figure 5.6 illustrates the process. An external teacher may provide initial instructions for how to draw back to golf club, how to strike the ball, and so on. These instructions may be memorised so that the inner self-talk recurrent loop is also able to give linguistic instructions to the arm, leg and other muscles.

Over time, however, the proprioceptive and visual inputs involved in playing the shot will form perceptual attractors and temporal sequences of their own, which are able to be honed by further practice. Eventually, the need for the linguistic sequences

†See Dreyfus (1996) for an extended discussion of skill acquisition.

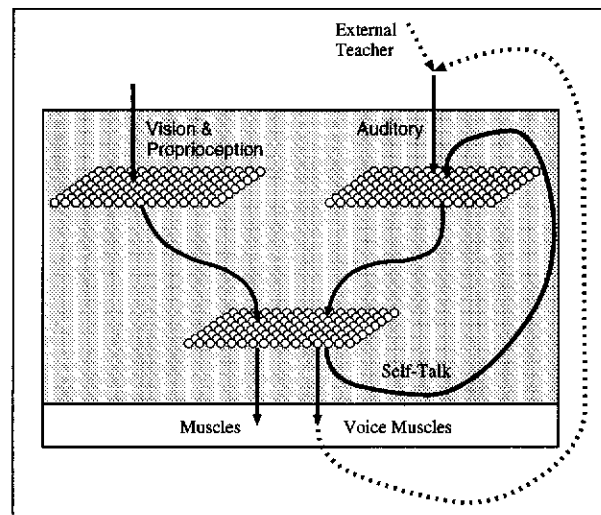


Figure 5.6: Gestalt Behaviour and Rules.

disappears and the motor skill is achieved.

The concept of a trainer (instructor) requires the transfer of some context-free linguistic description (theory) of a motor or perceptual skill to another person—the student. A sequence of linguistic terms is used by the student as the initial mechanism, and in conjunction with already *general* control over their muscles, they learn to perform the task as described. After sufficient practice, the motor skills act in their own right, and the linguistic *instructions* are no longer required.

After the perceptual and motor skill sequences have supplanted the linguistic instructions, resorting back to the linguistic mechanism can alter the skilled behaviour. This is a well known phenomenon with sports people who may think too much about what they need to do rather than just letting it flow. †

5.2.5 Automaticity

The previous discussion was essentially the ABC model's explanation of what is known in the literature as *automaticity*. A number of researchers maintain that people can overcome limitations (in consciousness) by *automatising* what initially required attention (Schneider & Shiffrin 1977, Shiffrin & Schneider 1977, Cheng 1985, Schneider & Shiffrin 1985).

†See Section 5.3.6 for further discussion on this point.

Definitions of automaticity vary (for example, see Kahneman & Treisman 1984), but we will follow that used by Schneider & Shiffrin. According to these researchers (Schneider & Shiffrin 1977, page 1):

Automatic processing is activation of a learned sequence of elements in long-term memory that is initiated by appropriate inputs and then proceeds automatically—without subject control, without stressing the capacity limitations of the system, and without necessarily demanding attention.

Controlled processing is a temporary activation of a sequence of elements that can be set up quickly and easily but requires attention, is capacity-limited (usually serial in nature), and is controlled by the subject.

Automatic sequences may appear in different forms, depending on the context. For example, a red traffic light may elicit a braking response from a car driver, but a slowing or stop walking response from a pedestrian. Automatic processes require an appreciable amount of consistent training to be learned fully, and once learned, are difficult to modify or suppress (Schneider & Shiffrin 1977).

Controlled processes require active attention by the subject, and only one, or possibly two serially interwoven processes may be undertaken at a time. Controlled processes are able to be set up, altered and applied in novel situations with relative ease, and for sequences which have not become automatic.

Automatic processing usually does not interfere with concurrent processing and does not require effort of intentionality (Shiffrin & Schneider 1977, Hasher & Zacks 1979).

Automaticity transforms processes that once occurred consciously and effortfully into ones that occur effortlessly and without awareness or consciousness. They do not require attention, but may attract it if training is appropriate.

An example of the transition from a conscious, controlled process to an automatic one is the beginning pianist. While she may initially consciously read a note and then press the correct key, her later transition to expert will see her expertly, smoothly and without any conscious awareness simply play what is written on the page (Sudnow 19781).

Various scholars differ on the relationship between automatic and effortful processing. For example, Shiffrin & Schneider (1977) suggest that the two differ qualitatively, whereas Posner & Synder (1975) contends that the two only differ quantitatively, with automatic processing simply being a 'sped up' version of effortful processing. Cheng (1985) maintains that the automatic processing reflects changes in the organisational structure of the processing.

Humans are able to expand their capacity by increasing their effort (Kahneman 1973), or by improving their perceptual and attentional skills (Hirst 1975). We are able to see, hear, and sense more or less of the world depending on how we organise and develop our skills. Our repertoire of automatic skills, the amount of effort we exert, and our levels of perceptual and attentional skills are all able to be extended and enhanced through automatisisation.

When we extend our organisational and perceptual skills, as our previously effortful behaviours become automatic, and as we put more effort into a task, then the limits on our conscious experience can change.

The transfer of conscious (linguistically-based) behaviours into automatic (perceptual attractor-based) behaviours is explained within the ABC model as discussed in the previous section.

5.2.6 Knowing-How and Knowing-That

An old philosophical problem is whether our understanding of the world is *knowing-how* or *knowing-that*. The cognitivist version of this discussion, the so-called distinction between procedural and declarative memory, was initially introduced into the lexicon via artificial intelligence (for example, see Winograd 1975).

Procedural memory is supposed to involve the representation of knowledge about the world as procedures, and subsequently the development of skills through extensive practice. Declarative memory, on the other hand, is said to involve the representation of the *outcome* of a processing operation that is available to conscious recall (Schaacter 1989a, page 702), the representation of the knowledge being facts and assertions.

Declarative knowledge might be termed ‘knowing-that’, while procedural knowledge is ‘knowing-how’. But for the ABC model, there is no distinction between these two supposed types of memory; that is, to even state a piece of information (who is the pope?) requires motor control, albeit of a short duration.

Within the ABC model, these two memory types are of the same type—just on a different time scale. The ‘fact’ that the *Challenger* shuttle exploded is a linkage between existing sequences; the ‘fact’ that my new colleague is called Simon is learning a new sequence; the skill of learning to play tennis is learning temporal sequences, but in a more complex and integrated (whole-body) manner.

The learning of a sequence such as a name involves the sequencing of (voice) muscles (with the subtle extension of self-talk), just as tennis involves the learning of muscle sequences.

While the two supposed memory types both require temporal sequences for their execution, knowing-that is primarily associated with the linguistic component of cognition, whereas knowing-how is usually associated with sensorimotor behaviours. Knowing-that may be the immediate reporting of a linguistic sequence—knowing-how is usually a motor sequence.

5.2.7 Perceptual Skills and Embodiment

A number of major philosophical figures of the current century have looked at skilled behaviour. As opposed to the rationalist tradition of Anglo-American analytical philosophy, a number of European phenomenologists have challenged the representational viewpoint. One of the areas they considered was that of the acquisition and use of skilled behaviour, and one of the major figures, Merleau-Ponty, held that we can also acquire the skill of perception (Merleau-Ponty 1962).

Merleau-Ponty provides the example of learning to feel silk. We learn to hold and feel silk in a certain way, and have certain expectations about its touch. Merleau-Ponty maintains that it is the body which enables the interconnection of skillful action and perception (Dreyfus 1996).

It is easier to become aware of the body's role in taste, hearing and touch. For example, an ability to discriminate subtle differences in wines is able to be achieved after considerable hours of practice. While previously we may not have been conscious of the differences in taste, we are able to increase our ability to identify and discriminate the tastes and smells associated with each wine. In the process, we change our conscious experience of wine tasting.

Seeing, too, is a skill that has to be learned. The skilled photographer or artist will see the world in a different way to the untrained eye.

The ABC model is an embodied model. It must learn its perceptual skills and motor skills *relative* to an existing body. Further, because of the self-organisation on the SOM surfaces, it also exhibits the property of improved performance with practice. If more input vectors relate to a particular domain (the domain of that being practised), then more of the nodes on the appropriate SOMS will be allocated to that domain, hence enabling finer discrimination and better control.

The ABC model will simply learn the weights appropriate to a particular body at a particular time. Because of its dynamic self-organisation of the weights, the model will also be able to accommodate gradual changes in the body over time, the process of 'growing up'.

Cognitivists, in making the assumption that human behaviour must be formalisable in terms of a heuristic program for a digital computer, are compelled to effectively exclude the body in developing any theory of intelligent behaviour. The body is only added on as some indirect attachment in robotic systems.

The cognitivist view follows on from the traditional view of thinkers, from Plato to Descartes, who have relegated the body to a non-existent role in cognition, maintaining that the body gets in the way of intelligence and reason. The ABC model, in line with Merleau-Ponty, takes the opposite view—that the role of the body is indispensable for cognition.

5.2.8 The Use of Tools

How do humans use tools? How are they able to become so skillful in using tools so that they seem almost to become an extension of their bodies?

The philosophers Heidegger, Merleau-Ponty and Polanyi have studied the use of tools, and each suggests that our use of tools is *not* an extension of our explicit knowledge of the tool itself as an object.

The example often given is that of a blind man with a cane. While he may know the characteristics and features of the cane itself by running his hand over its surface, nevertheless his use of the cane will in no way rely on any conscious knowledge on his part of these features. That is, his explicit knowledge of the weight, hardness and texture of the cane is irrelevant to his successful use of the cane as an aid to moving about in the world. He is not consciously aware of its position in physical space, its features, or the varying pressure that it exerts in the palm of his hand. He is able to bring the cane into contact with an object in physical space without needing to be aware of the physical location of the probe. Rather, the cane has become an extension of his sense of touch.

Polanyi (1962, page 59) states:

While we rely on a tool or a probe, these are not handled as external objects ... they remain on our side ... forming part of ourselves, the operating persons. We pour ourselves out into them and assimilate them as parts of our existence. We accept them existentially by dwelling in them.

Another extremely interesting tool employed by blind people is a box, worn on the persons back, which can deliver a series of taps to certain positions of the back in response to active visual prompts obtained from an attached camera. While only supplying broad, overall information of the scene in front, the device is nevertheless useful in helping the blind person move around in the world. The question is, how does it work?

The ABC model suggests a mechanism of skill learning through automaticity. At first

the patient needs to consciously (verbally) interpret the taps; for example, 'if a tap is high and right, then there is some object in that location', but later the patient is able to receive learned perceptual attractor information directly and unconsciously. He is able to interpret each tap (or combination of taps) in exactly the same way that we feel sensations from our somatosensory system emanating from our hands. Once these perceptual attractors are formed, the patient does not need to use any verbal interpretations, and this has the phenomenology of an 'unconscious' attention to the device. Furthermore, the impression is that the sensations are coming from the camera location, and not from the back.

Similarly for the blind person with the cane. It is as if he were born with an extended arm, and the somatosensory inputs that he receives from the cane are interpreted not as tactile sensations to the palm of his hand, but as implicit tactile inputs from the body of the cane. Again, while the blind person will have to initially learn to use the cane through trial and error, and interpret the tactile sensations received at his palm explicitly, the process will soon become automatic as new perceptual attractors are formed.

Merleau-Ponty (1962, page 106) suggests that:

The whole operation takes place in the domain of the phenomenal; it does not run through the objective world, and only the spectator, who lends his objective representation of the living body to the active subject, can believe that ... the hand moves in objective space.

As stated by Dreyfus (1992, page 253):

... Merleau-Ponty admits that this ability seems "magical" from the point of view of science, so we should not be surprised to find that rather than have no explanation of what people are able to do, the computer scientist embraces the assumption that people are unconsciously running with incredible speed through the enormous calculation which would be involved in programming a computer to perform a similar task. However implausible, this view gains persuasiveness from the absence of an alternative account.

5.3 Consciousness

It is not our intention within this thesis to offer the definitive word on consciousness. This is a task that is clearly beyond our capabilities. However, the ABC model does offer some alternate views to consciousness that may offer a fresh understanding of some of its complexity. Time and space pressures prevent a more expansive analysis and discussion, but at least we hope to lay out the essence of a future, more detailed approach to an understanding of consciousness. First, we begin with a brief history of the study of consciousness †

For Descartes, all mental states were of necessity conscious, and the notion of an unconscious mental state would have been close to a contradiction in terms. As he stated in his reply to the ‘Fourth Set of Objections’:

As to the fact that there can be nothing in the mind, in so far as it is a thinking thing, of which it is not aware, this seems to me to be self-evident. For there is nothing that we can understand to be in the mind, regarded in this way, that is not a thought or dependent on a thought. If it were not a thought nor dependent on a thought it would not belong to the mind *qua* thinking thing; and we cannot have any thought of which we are not aware at the very moment it is in us.

The psychological work of Wundt and James in the late 19th century followed on from the tradition of Descartes in that they tried to examine the causes of behaviour by introspection, and developed psychological theories on the basis of introspective evidence.

The seminal theories of Freud on the unconscious established the important idea that much of the brain’s activities are not open to conscious awareness, and that there may be such things as unconscious beliefs and desires. Freud recognised that accessibility to consciousness was not an essential component in the explanation of behaviour, and that a conscious quality was not constitutive of all behaviour. ‡

†This history is based, in part, on Chalmers (1996).

‡The Freudian concept of the unconscious is one of dynamic conflicts between deeply meaningful

At around the same time that Freud was opening up the possibility of unconscious cognitive behaviour, the behaviourist movement in psychology had thoroughly rejected the introspectionist tradition. They developed a new outlook for psychology that required a thoroughly objective analysis of behaviour. The behaviourist movement, in the main, left no room for a consideration of consciousness. The behaviourists differed in their particular theoretical positions, with some allowing for the existence of consciousness, but even these researchers suggested that consciousness was irrelevant to psychological explanation. Some behaviourists denied the existence of consciousness altogether. Many behaviourists even denied the existence of any kind of mental state, maintaining that internal states were methodologically irrelevant in the explanation of behaviour. They claimed that an explanation of behaviour could be obtained entirely in external behavioural terms.

The move from behaviourism to computational cognitive science brought back a role for internal (unconscious) mental states, but the role of consciousness, and the division between conscious and unconscious states was not well defined. Most cognitivists simply ignored it.

5.3.1 Definitions of Consciousness

Before we go on to describe the ABC model in relation to consciousness, it is appropriate to spend some time defining what we mean by consciousness. This is somewhat difficult as many people mean different things when they talk about consciousness. Further, as Farber & Churchland (1995) suggest, it is perhaps best that we do not attempt a *precise* definition at this stage of our understanding. Be that as it may, we still need to take a representative look at the views of a number of researchers in order to focus our discussion.

Hirst (1995) suggests that consciousness refers to “the awareness people have of mental objects, be they percepts, images, or feelings.” He suggests that “to be *aware* entails being able to report this awareness. People are not merely aware of a percept, image, or feeling; they are also aware that they are experiencing it. They know that *they* are forces creating at times a bubbling cauldron of psychic pressure (Kihlstrom 1987).

seeing, imagining, and feeling. Any act of consciousness involves self-awareness as well as awareness of the world, mental images, or feelings” (Hirst 1995, page 1307).

Tulving (1985) distinguished among three types of consciousness: *anoetic* (non-knowing), which entails simple awareness of external stimuli; *noetic* (knowing), which involves symbolic representations of the world; and *autonoetic* (self-knowing), which involves awareness of the self and personal experience extended in time.

Farthing (1973) make a distinction between primary consciousness—the simple perceptual awareness of external and internal stimuli—and reflective consciousness—“thoughts about one’s own conscious experiences *per se*.”

Farber & Churchland (1995) suggest three broad categories of conscious phenomena:

conscious awareness:

That is, *someone* is consciously aware of *something*. This form of consciousness may relate to sensory awareness, or it may be a more generalised awareness of fatigue, dizziness, anxiousness, comfort, hunger, and so on. The awareness may be of a temporal duration or of a spatial layout. We are aware of ourselves as being a thing that changes somewhat but endures through time.

The awareness may also be a form of meta-cognitive awareness—being aware that we have knowledge (such as “I am aware that I know the name of my paternal grandfather”), or being aware of what one is now thinking (the train of thoughts that led up to the present one), or even being aware that one is aware.

Conscious awareness may involve some recall (linguistic or of mental imagery) to indicate an awareness of events that occurred earlier.

higher facilities:

Here the implication is one of agency and control. Agency is not thought to be a component of ‘unconscious’ (but otherwise animate) entities such as insects, zombies and robots. Furthermore, agency would appear to be necessary for free will. A conscious creature cannot simply be an aware but passive receptor of information—they must be able to perform a repertoire of *facilities* including *attention* (in order to influence what it will and will not become aware of), *reasoning* (a type of information processing that

operates at a higher level of abstraction—such as a knowledge of algebra—as opposed to, say, a conditioned aversion such as cringing when a bell sounds), *self-control* (with consciousness acting as an arbiter for internal disputes, or the imposition of moral beliefs or reasoning over base, physical impulses).

As Farber & Churchland (1995) state, this raises the question: “are we the captains of our own ship?” One possibility is that free will and agency is an illusion. Gazzaniga (1985) (and to some extent Dennett (1991)) suggest that when we experience ourselves as having made a conscious decision, what we are in fact getting is a “story” reconstructed from the (external and internal) evidence. †

As stated by Farber & Churchland (1995):

... many of the apparent characteristics of conscious mental processes are radically unlike those of unconscious processes, or for that matter of neural processes in general: Conscious thought is comparatively slow, serial, and abstract; it deals with only a few objects at a time; its contents are readily translated into a communicable form (i.e., language); ‡ and its storage and processing limits can be overcome by the use of external objects such as books, calculators, maps, and word-processing programs.

states of consciousness:

Consciousness is often used to refer to a *state*, which can be present in varying degrees, and which in some sense embodies what is going on in a person’s mind.

Before proceeding, we must reject the notion that consciousness is a single, holistic phenomenon that is either present or not—some special property that we cannot hope to explain. The underlying assumption used in this thesis is that consciousness is explainable without the need to go beyond the physical confines of the body and brain.

†We discuss this further in Section 5.3.3 where we examine the introspection and verbal reports experiments of Nisbett & Wilson (1977). These experiments provide empirical support for this view, and suggest that the “introspective” stories that one produces in order to provide reasons for our behaviours are heavily influenced by external context.

‡Some, such as Fodor (1976), suggest that the basic structure of thought is linguistic.

5.3.2 Limits to Conscious Experiences

Hirst (1995) discusses the data limitations to conscious experience. For example, humans and other animals are restricted by certain limits on their sensory organs—we are unable to see wavelengths above a certain frequency, or to hear sounds below a certain decibel level. Different animals experience different sensory limits, and even have completely different senses, such as sonar for bats and dolphins.

We are also limited in the number of things that we are able to attend to at the same time. For example, James ([1890] 1950, page 409) suggested that the number of attended items is “not easily more than one, unless the processes are very habitual; but then two, or even three.”

The limitations extend to the combination of senses that are able to be attended to. For example, Wickens (1980, 1984) has shown that it is easier to attend to one auditory and one visual message than either two auditory messages or two visual messages. †

We seem to be aware of many of the results of perceptual and memory processes. However, as we indicate in the following section, these appear *not* to be the result of deliberate, logical processes but rather contextual rationalisation. On the other hand, we appear to have only limited access to the processes that produce this awareness. For example, if one is asked the question, “What is the first name of your wife (or husband, son or daughter)?”, we are usually able to produce an answer immediately—you *know* the name, which pops out of memory with no awareness of how the result is obtained.

If, however, we are asked the question “How many windows are there in your house?”, we need to perform a *conscious* process to imagine the house and its rooms, and count the windows. This much slower cognitive process involves the sequential scanning of some form of cognitive map, and is a conscious cognitive process. We appear to be aware of the process involved. We can vary the strategy and perhaps start from a different location, and so we may even be conscious of one’s conscious control of the process.

†Through the process of automatisisation, people are, to some extent, able to overcome some limitations in attention (Schneider & Shiffrin 1977, Shiffrin & Schneider 1977, Cheng 1985). See the discussion of automatisisation in Section 5.2.5.

5.3.3 Introspection and Verbal Reports

How aware are we of our 'mental reasoning', of our beliefs and desires. When asked to explain the reasoning that we used on a particular choice, evaluation or judgement, how accurately can we relate the factors making up our decision?

It appears that we have very limited introspective access to the origins of our 'higher order' cognitive processes. In an extensive review, Nisbett & Wilson (1977) have compiled an extensive discussion on the limits to our introspective capabilities. They provide strong evidence that humans have little or no direct introspective access to so-called higher order cognitive functions.

By manipulating experimental and real-life situations, Nisbett & Wilson and other researchers have shown that subjects are sometimes (a) unaware of the existence of a stimulus that importantly influenced a response, (b) unaware of the existence of the response, and (c) unaware that the stimulus has affected the response.

For example, in an experiment conducted by Nisbett & Schachter (1966), subjects were requested to undergo a series of electric shocks of increasing intensity. Before the electric shocks, the patients were given a placebo and told that the symptoms of the pill would be heart palpitations, breathing irregularities and so on. These symptoms are in fact those that normally accompany the experience of electric shocks, but this was not revealed to the subjects. The subjects were subsequently able to tolerate four times the amperage than a control group who were not given placebos.

The subjects given the placebo seemingly attributed their arousal symptoms to the pill rather than the shock, and so were able to tolerate more. However, when questioned at a debriefing session after the experiment about their increased tolerance, most subjects failed to attribute their increased tolerance to the pill, positing instead other extraneous reasons (such as shocks previous experienced in hobbies or employment that were thought to build up a tolerance). Even when told of the experiment, the placebo, and the hypothesis of the study (that they would incorrectly attribute their symptoms to the pill rather than the shock), "subjects typically said that the hypothesis was very interesting and that many people probably would go through the process that the ex-

perimenter described, but so far as they could tell, they themselves had not" (Nisbett & Wilson 1977, page 237).

The three major conclusions reached by Nisbett & Wilson (1977, page 233) are described in more detail:

- People often cannot report accurately on the effects of particular stimuli on higher order, inference-based responses. Indeed, sometimes they cannot report on the existence of critical stimuli, sometimes cannot report on the existence of their responses, and sometimes cannot even report that an inferential process of any kind has occurred. The accuracy of subjective reports is so poor as to suggest that any introspective access that may exist is not sufficient to produce generally correct or reliable reports.
- When reporting on the effects of stimuli, people may not interrogate a memory of the cognitive processes that operated on the stimuli; instead, they may base their reports on implicit a priori theories about the causal connection between stimuli and response. If the stimulus psychologically implies the response in some way ... or seems "representative" of the sorts of stimuli that influence the response in question ... , the stimulus is reported to have influenced the response. If the stimulus does not seem to be a plausible cause of the response, it is reported to be non-influential.
- Subjective reports about higher mental processes are sometimes correct, but even the instances of correct reports are not due to direct introspective awareness. Instead, they are due to the incidentally correct employment of a priori causal theories.

Nisbett & Wilson propose that when people attempt to report on their cognitive processes, that is, on the processes mediating the effects of a stimulus on a response, they do not do so on the basis of any true introspection. Rather their reports are based on a priori, implicit causal theories, or judgements about the extent to which a particular stimulus is a plausible cause of a given response.

Accurate reports only occur when influential stimuli are salient and are plausible causes of the response they produce. When influential stimuli are not salient or are not plausible causes, then inaccurate reports result. Subjects resort in the first instance to contextually relevant, culturally supplied explanations for their behaviour. Failing that, they then seek an explanation that “may be adduced as psychologically implying the behavior” (Nisbett & Wilson 1977, page 249).

Nisbett & Wilson suggest that people do not consult a memory of the mediating process when asked to explain how a particular stimulus influenced a particular response, but rather apply or generate a causal theories about the effects of that type of stimulus on that type of response. The causal theories may have any of several origins:

- the culture or subculture may have explicit rules stating the relationship between a particular stimulus and a particular response (“I came to a stop because the red light started to change.”)
- the culture or subculture may supply implicit theories about causal relations—the presence of a particular stimulus may psychologically imply a particular response (“Jim gave flowers to Amy; that’s why she’s acting pleased as punch today.”)
- an individual may hold a particular causal theory on the basis of empirical observation of covariance between stimuli of the general type and responses of the general type (“I’m grouchy today. I’m always grouchy when I don’t break 100 in golf.”)
- causal hypotheses may be generated by linking even novel stimuli and novel responses, provided that the stimulus is connotatively similar to the response

An example given by Nisbett & Wilson illustrates this point. If we ask another person why he enjoyed a particular party and he responds with “I liked the people at the party,” we may be extremely dubious as to whether he has reached this conclusion as a result of anything that might be called introspection. We are justified in suspecting that he has instead asked himself *Why People Enjoy Parties* and has come up with the altogether plausible hypothesis that in general people will like parties if they like the

people at the parties. Given that he did not meet any of his enemies at the party, he will maintain that people-liking was the basis for his party-liking.

5.3.4 Conscious, Nonconscious and Unconscious

Before we look for explanations of consciousness, we should first ascertain the relevance and veracity of our current views on consciousness and the unconscious. †

As stated by Farber & Churchland (1995), “when people have just been hit on the head and are now lying immobile, it is usually safe to infer that they are *unconscious*—that they have no awareness, are not engaged in reasoning or self-control, do not feel any emotions, and in short are enjoying no more of a mental life than do plants and insects.”

This is one form of the lack of consciousness, and it reflects a disruption or breakdown in the normal state of awareness experience by people. Other examples of this disrupted state of consciousness include drowsiness, intoxication, sleepiness and fainting. ‡

However, this is not what is usually meant when we talk of *the unconscious*. What is usually meant is that even when we are awake and alert, certain activities occur in the brain that we are not aware of.

Our conscious experience of decision making and control is now being seen as a poor reflection of the way things actually get done in the brain. Many of the processes of cognition are more or less inaccessible to our conscious awareness (Gazzaniga 1985,

†Some have suggested that certain phylogenetically older regions of the brain, such as the centre-cephalic system located at the core of the brain stem and thalamus, represent the site of consciousness because any damage to these areas causes the animal to lapse into a permanent state of unconsciousness. But this view is too simplistic, and is analogous to suggesting that the site of power in a car engine is in the on/off switch. The centrecephalic system may be necessary, but is clearly not sufficient.

‡As an aside, Churchland (1995) proposes that the recurrent pathways from the intralaminar thalamic nucleus (thalamus) can account for both the disappearance of consciousness in deep sleep and its muted or disjointed reappearance in REM sleep. He proposes that the recurrent processing through the intralaminar nucleus explains the unified character of consciousness—all processing must go through the bottleneck of the intralaminar nucleus, and this imposes unification on the processing.

This view supports the claim made in Section 3.2.7 which suggested that the thalamus provided overall synchronisation and timing for the vector flows within the cortex, as required by the ABC model.

Lycan 1987, Dennett 1991).

We have known for some time that we are not conscious of certain bodily processes, such as the dilation of the pupils. Other behaviours may be brought under some temporary degree of conscious control, such as breathing. These activities are usually handled automatically and unconsciously by the brain, yet we know (and “feel”) that we can potentially take these functions over “at will”. For example, there is anecdotal evidence of Indian holy men being able to consciously put their heart into fibulation.

However, the extent of unconscious processing in the brain goes much deeper than these activities. The so-called higher level processes of planning, deciding, thinking and so on, which many have previously maintained are exclusively conscious activities, have been shown to proceed outside of our conscious awareness.

As put by Mandler (1975, page 241), “the analysis of situations and appraisal of the environment . . . goes on mainly at the nonconscious level.” Mandler (1975, page 245) also suggests that “there are many systems that cannot be brought into consciousness, and probably most systems that analyze the environment in the first place have that characteristic. In most of these cases, only the products of cognitive and mental activities are available to consciousness.”

Miller (1962, page 56) adds that “it is the result of *thinking*, not the process of thinking, that appears spontaneously in consciousness.” This view is shared by others, including Neisser (1967, page 301) who suggests that “the constructive processes [of encoding perceptual sensations] themselves never appear in consciousness, their products do”, and Nisbett & Wilson (1977, page 232) who add that “recent research has made it increasingly clear that there is almost no conscious awareness of perceptual and memorial processes.”

Greenwald (1992, page 778) notes that in the past, certain “academic psychologists have sometimes gone beyond empirical skepticism to suggest that the concept of *unconscious cognition* has no place in psychology”, but goes on to dismiss this now unsupportable view of cognition.

Some psychologists attribute only shallow processing to the unconscious, suggesting

that information impinging on the sensorium can be processed at a meaningful level only if it emerges into consciousness. Others assign a deep level of analysis to the unconscious (Marcel 1983a, Marcel 1983b, Greenwald 1992). For example, Marcel (1983a) claims that subliminally presented stimuli support the claim for deep unconscious processing. Hirst (1995) provides details of this subliminal priming effect. Other results from brain-damaged patients also suggest support for the claim of deep unconscious processing.

Greenwald (1992, page 778) discusses various forms of unconscious cognitive activity, including subliminal activation, selective attention, unconscious (implicit) learning, self-deception and memory repression.

Farber & Churchland (1995) suggest that highly skilled and practised actions in everyday life, such as driving an automobile, provide examples of unconscious behaviour. These actions, which are performed without apparent mental effort and often without the ability to remember what one has done, are described by some as being done unconsciously.

The most general meaning of unconscious is *unaware of*, and Farber & Churchland (1995) maintain that there are two senses of ‘unaware of’ that appear in cognitive psychological research and theory:

outside of attention

one is unaware of stimuli that impinge on the receptors but fall outside of the metaphorical spotlight of selective attention,

lack or failure of introspection

one is unconscious or unaware of the occurrence, causes, or other attributes of attended objects, events, or actions when one cannot report those properties.

It is this latter sense that is of interest in relation to the ABC model. † In Section 5.3.10 we look at how this separation of behaviour and the reporting of that behaviour is

†We briefly touched upon the concept of attention in Chapter 4, suggesting that (at least for saccades) the process of attention can be explained by learned scanpaths rather than attention within an iconic world view.

achieved within the model, but first we need to examine just how conscious we really are of our surrounding world and of our cognitive processes.

While the distinction between consciousness and the unconscious is usually associated with Freud, there is little doubt that many psychologists now regard unconscious processing as an important, if not primary, component of cognition.

5.3.5 Conscious Awareness of Voluntary Acts

Some intriguing results regarding conscious awareness have been obtained by Libet (1985). He observed that the awareness of the urge to initiate a voluntary act, as reported by the subject, appears about 350 msec *after* the act develops unconsciously, as indicated by recordings of cortical activity. Thus, although from the first person perspective (experienced by the subject) the urge to initiate the voluntary action precedes the action, the third-person observations of the experimenter show that the action is actually in progress before the urge is registered in the subject's consciousness. † ‡ §

From the experimenter's (third-person) perspective, the readiness to act is signalled by a readiness potential. As this precedes the reported conscious urge, preparedness to act must be preconscious. However, from the subject's (first-person) perspective, the preparedness to act is manifested not as a readiness potential, but in the form of a conscious volition or urge. Hence, from a first-person perspective conscious volition

†Note that this is *not* a different temporal ordering of the *urge* and the *muscle initiation* of the action, but an apparent initiation of the process from an unconscious mechanism rather than via direct conscious control. There are no time reversals.

‡Many of Libet's writings are collected together in Libet (1993b).

§One of Libet's suggestions to explain this phenomenon of consciousness was to suggest that the purpose of conscious control was not to initiate an action, but to *veto* it. That is, although voluntary actions may be unconsciously predetermined, they can still be consciously vetoed during the buffer time. As Libet (1985) suggests, "conscious volitional control may operate not to initiate the volitional process but to select and control it, either by permitting or triggering the final motor outcome of the unconsciously initiated process or by vetoing the progression to actual motor activation.

However, this would leave open the question of how this conscious veto is initiated—it would appear that such vetoes would likewise first have been unconsciously predetermined in exactly the same way that actions were. See Libet (1993a, page 390) for a discussion on this point.

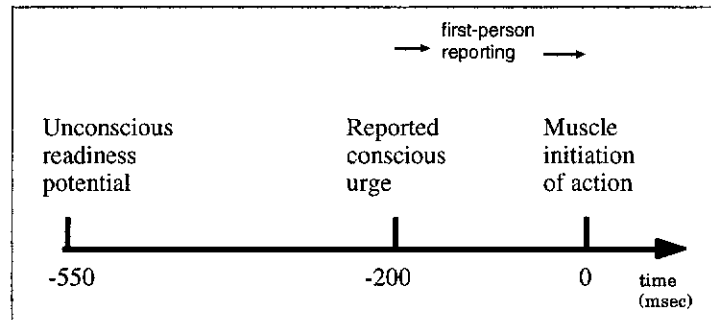


Figure 5.7: Reporting of Conscious Activity by Libet.

appears causal, whereas from a third-person perspective it does not.

For example, the reaction time to a tactile stimulus may be as little as 100 msecs, but awareness of that stimulus does not arise until at least 200 msecs after it projects onto the cortical surface. So from the perspective of the experimenter, the subject's awareness appears to arise after his response. For the subject, on the other hand, awareness of the stimulus occurs before his response (Velmans 1993, page 413).

A description of a typical experiment used by Libet is instructive. The action might be a simple finger or wrist movement, but the subject was instructed to "allow each act to arise 'spontaneously,' without deliberately planning or paying attention to the 'prospect' of acting in advance" (Libet 1985, page 530). The subject was allowed complete freedom over their decision to act at any time.

However, the subjects were asked to pay close introspective attention to the instant of the onset of the urge, desire, or decision to perform each such act and to the correlated spatial position of a revolving spot on a clock face (indicating 'clock time'). Thus the action was designed to be a conscious, endogenously-willed motor action.

Libet's experiments have been questioned on technical grounds, but there are a number of other experiments which raise similar fundamental issues which have not been questioned (Dennett & Kinsbourne 1993). These include timing differences in actual and subjective reporting of somatosensory inputs—the cutaneous "rabbit" (Geldard & Sherrick 1972, Geldard & Sherrick 1983, Geldard & Sherrick 1986).

Another experiment performed by Libet, Wright Jnr., Feinstein & Pearl (1979) indi-

cated that if a subject's left cortex—in an area corresponding to the right hand—was stimulated *before* his left hand was stimulated, then the subjective report reversed the expected sequence—that is, the patient reported first left, then right hand tingle experiences.

In a fascinating experiment, Grey Walter (1963) asked subjects to view a series of slides from a carousel projector. The patients were fitted with electrodes in their motor cortex. They were asked to advance to the next slide at a time of their choice—a 'free' decision. The subjects were unaware, however, that the projector controller button was actually a dummy, and the slide advance mechanism was actually initiated by an amplified signal from the implanted electrode. The subjects reported that it seemed as if the slide projector was anticipating their decisions—just as they were 'about to' push the button, the projector would advance the slide.

In these and other examples, an apparent dislocation in the timing of events questions the *prima facie* plausible thesis that our conscious acts are the *cause* of events in our nervous systems that control our bodily acts.

Churchland et al. (1994) are of the opinion that:

All cortical visual areas, from the lowest to the highest, have numerous projections to lower brain centres, including motor-relevant areas such as the striatum, superior colliculus, and cerebellum. The anatomy is consistent with the idea that motor assembly can begin even before sensory signals reach the highest levels. Especially for skilled actions performed in a familiar context, such as reading aloud, shooting a basket, and hunting prey, this seems reasonable.

This is exactly the situation with the ABC model. Conscious awareness requires us to *tell ourselves* (via self-talk) what is going on by traversing the conceptual/cortical recurrent loops of the language system. However, the actual muscle action may be set in train by a perceptual attractor which may in turn have been set off by some sensory input. Thus the conscious awareness of the action (via self-talk) may lag some time behind the direct behaviours of learned motor sequences. Because the ABC model has

placed language and self-talk (and hence a major component of conscious awareness) into a subordinate role to perceptual behaviour, it is no longer reliant upon verbal (conscious) control to initiate actions.

Libet (1993a) suggests some implications of these results:

- Our subjective experiences elicited by sensory signals are delayed by up to 0.5 sec. Libet suggests that the actual delays are subjectively ‘corrected for’ by antedating the experience back in time.
- All of our cognitive events may begin unconsciously.
- Unconscious cognitive functions may be mediated by relatively brief durations of neural activity—even as little as 100 msec or less. Conscious events, on the other hand, proceed relatively slowly.
- Quick behavioural responses, such as avoiding obstacles in walking or running, or responding to signals while driving a car, are (at least initially) unconscious.

5.3.6 Sports Psychology

Unconscious reactions are essential and predominant in most sports, such as tennis, football and boxing. In these activities “it is commonly accepted that the immediate fast responses to sensory signals are made unconsciously and that the intrusion of conscious awareness fatally delays the responses” (Libet 1993a, page 388).

Durations of cerebral activity too brief for awareness can nevertheless subserve meaningful sensory signal detection and subsequent behavioural responses. This has been demonstrated by a number of experimenters, including Libet, Pearle, Morledge, Gleason, Hosobuchi & Barbaro (1991). Further, the reaction time to a sensory signal is the same no matter whether the subject is aware of the signal or not (Taylor & McCloskey 1990).

Another experimental finding of relevance was made by Jensen (1990). When subjects were asked to slow their reaction times deliberately, a procedure that presumably re-

quires conscious awareness of the signal, their reaction times suddenly increased by more than 300-400 msec over that of their normal behaviour.

Ornstein (1991, page 139) recalls a story of Ayrton Senna, the then world-champion racing-car driver, in which Senna created exceptionally fast lap times in practice by allowing his driving to proceed without conscious control. Senna drove in an 'unsupervised' and 'uninhibited' manner, lapping an unheard-of two seconds faster than any other driver. Senna slowed down and returned to the pits somewhat shaken when he realised that his driving was *out of control*.

But for exceptionally trained sports-people, this is the exact situation they hope for—where they are *in the groove*, or *flying from the seat of their pants*—where conscious thoughts do not interfere with their highly trained behaviours.

This is also true for 'slower' sports such as golf, and many a potential champion has 'choked' on the thought of winning when their conscious thoughts intrude and interrupt their natural, unstressed game. Telling yourself how to play a shot in a championship would seem a recipe for disaster. Such self-conscious thoughts about 'correct' muscle control when under stress will alter the player's muscle behaviour—invariably for the worst, and with possibly a positive feedback as to the level of stress, the problem may get worse as the game proceeds.

The ABC model is able to account for this phenomenon. The sport-person's natural game is as a result of finely 'tuned' temporal sequences of perceptual attractors, built up over many hours of practice. While they may have used a coach's instructions in initially developing their skills, once it becomes well practised the skill is no longer directed by verbal instructions, either from a coach or through self-talk. There is no requirement or possibility of linguistic interaction, and the skill is a direct behavioural response to the appropriate sensory inputs.

If, however, a player resorts to trying to consciously control his skill, then the timing and coordination will be altered, (as shown by Jensen (1990)). This will invariably result in a decrease in performance as the conscious control will intersperse linguistic 'muscle instructions' into the sensorimotor mechanisms.

5.3.7 Problem Solving

There is a striking similarity in the description of the processes of problem solving given by creative people—artists, writers, mathematicians, scientists and philosophers. Invariably, these creative people describe an unconscious process in which they see themselves almost universally as bystanders. The process of problem-solving is almost completely hidden from conscious view. According to Ghiselin (1952, page 15), its “production by a process of purely conscious calculation seems never to occur.”

The overwhelming characteristics of significant creative problem solving are (a) the influential stimuli are usually completely obscure—the individual has no idea what factors prompted the solution; and (b) even the fact that a process is taking place is sometimes unknown to the individual prior to the point that a solution appears in consciousness. In many cases, the individual has stopped actively thinking about the problem to move on to other things, or has even been asleep, when the solution appears almost as if by magic.

Maier (1931) has shown that the processes involved in mundane problem-solving in everyday life do not differ, in the degree of conscious participation, from the problem-solving of creative geniuses. An example of Maier’s elegant work involved him presenting individual subjects with a difficult problem to solve, such as how to join together two ropes hanging from the ceiling but at a distance apart so that the subject could not reach both at once. Various implements were scattered about the room, and the subjects were allowed to use these in order to obtain a solution. Most subjects were unable to solve this problem until Maier introduced a perceptual ‘hint’ unknown to the subject—he surreptitiously started to swing another rope from which a weight was attached.

Almost immediately, the subjects were able to solve the problem by tying a weight to one rope, setting it in motion, and then pulling the second rope close enough so that they could grasp the weighted rope as it swung close by. When questioned about their reasoning in obtaining a solution, the subjects invariably came up with reasons such as “it just dawned on me.” Most made no reference to Maier’s hint in subsequent probing into their reasoning.

According to the ABC model, problem solving is essentially the application of previously learned sequences in new situations. We do not apply logical operators to input data to obtain a logical solution. Rather, at each moment we seek to apply previously learned behaviours to the current situation. †

Creative problem solving, on the other hand, seems to involve the linking of a previously learned behaviour to the current situation in a novel and more appropriate manner. At this stage we do not have a proposed mechanism for this within the ABC model.

5.3.8 Expert Knowledge

Polanyi (1962) and Gross (1974) have argued persuasively that “we can know more than we tell,” by which it is meant that people can perform skilled activities without being able to describe what they are doing and can make fine discriminations without being able to articulate their basis.

This phenomenon is well known in the case of expert system development as discussed in Section 3.3.5. While the experts are able to perform their tasks at a high level of skill (using sensory inputs), they are unable to describe their reasoning.

The ABC model explains this by suggesting that the expert’s skill is produced by sensorimotor skilled behaviours. However, they may not have appropriate linguistic attractors or sequences linked to these perceptual attractors, thus rendering them unable to explain their skillful behaviours.

5.3.9 Other Models of Consciousness

The study of consciousness is the new flavour of the month. Whereas once it was discussed only after a few drinks and in the company of friends, consciousness is now well established as a valid subject for scientific scrutiny. As such, a number of theories on the origins and purpose of consciousness have been proposed.

It is not appropriate that we spend much time on an appraisal of these other theories.

†See also the discussion of planning and *situatedness* in Section 5.5.1.

Chapter 3 of the recent book by Chalmers (1996) has a brief description of a number of these theories. Most are in the cognitivist paradigm, and so we believe that these will suffer from the difficulties outlined at other places within this thesis.

For example, Johnson-Laird (1988) suggests that the brain is a complex hierarchy of largely parallel processors, which is controlled by a separate operating system at the top of the hierarchy. The conscious mind somehow corresponds to this operating system, and the mechanisms of consciousness provide results for some of the computations the brain makes, while keeping hidden the details of how they were computed. However, as suggested by Crick & Koch (1991), “if there is an operating system of this type it is not easy at this moment to see any particular brain area in which it is located.”

Jackendoff (1987) provides an alternate view in which consciousness is not associated with the highest levels in the hierarchy but with the intermediate levels. Jackendoff arrives at his theory by considering the language system, the visual system and his postulated ‘music system’.

Other models have been proposed by Baars (1988), Chalmers (1996), Dennett (1991), Edelman (1989b), Lycan (1987), and Ornstein (1991).

5.3.10 ABC and Consciousness

As we discussed in a previous section, there are many aspects to consciousness. In this section we examine a number of these, and make comment on how the ABC model provides a possible explanation for each. The conclusions drawn are, of course, rather speculative—a necessary condition in any discussion of consciousness at this stage. However, the ABC model does seem to provide a number of explanations of consciousness which are quite promising. However, further experimentation is required before more definitive answers will be forthcoming.

states of awareness

Sleep, exhaustion, illness and the taking of drugs are some of the ways in which our experience of the world is altered—our awareness is diminished, or in the case of some drugs, even giving the appearance of being enhanced.

The ABC model suggests that brain is an integrated device that receives inputs from the outside environment as well as internal inputs of various kinds. The process of accepting this information proceeds at some rate, which we suggest is determined by phylogenetically older parts of the brain than the cortex. There is a form of natural ‘clocking rate’—a natural rhythm.

Should the input from the outside environment be altered in any way (removed completely as in sleep, or perhaps slowed down due to illness or drugs), then our awareness of the world would be affected. Our normal expectations of our interactions with the world would not be forthcoming.

Through self-talk, we would be able to report this experience to ourselves, and so be aware of it.

Other drugs might enhance or disrupt the firing of certain components of our vector sensory inputs. For example, the visual sensory vector elements could be increased relative to the other sensory inputs, taking our sensory inputs out of the range of previous experiences, and thus distorting and appearing to enhance our visual sensory perceptions.

sensory inputs

A primary component of our consciousness is our awareness of our sensory inputs. These include sensory inputs from the environment (visual, auditory, somatosensory and so on), as well as numerous readings from internal sources (such as proprioceptive, hunger, fatigue, pain, etc.).

Primary consciousness is the sensory awareness of something—that is, to see it, or to hear it, and so on. The ABC model maintains that these sensations are experienced directly and not through some mediating ‘representations’ as with the cognitivist model. The sensory inputs are connected directly (and recurrently) with motor behaviours via soft-reflexes. This enables an intimate ‘feel’ and ‘awareness’ of our muscles, of our body, and of our usual surroundings.

As we have discussed elsewhere, one of our motor behaviours is self-talk. Thus we are able to ‘feel’ and ‘experience’ the sensory inputs with direct reporting of the experience to ourselves through self-talk.

being able to report

As we discussed in Section 5.2, a great deal of cognition is implicit; that is, not able to be reported. Implicit learning is not experienced as conscious. For example, blindsight patients only feel some vague sensation of an object within their scotoma even after some training. They initially are totally unaware of the target light, and will only point to it when asked to guess.

It is proposed, then, that *self-reporting* via self-talk is a primary phenomenon of consciousness. After all, what is primarily our sense of consciousness other than telling ourselves via self-talk that we are aware of something? Knowing-that is being able to confirm to ourselves that we are aware of something. Knowing-that is having some knowledge available in linguistic form.

And as we show elsewhere, our reporting is usually not an analytic appraisal of our being through introspection, but rather the regurgitation of previously learned ‘social theories’ to fit in with the current context—in other words, our verbal behaviours are context driven as are our other motor behaviours.

Consciousness without reporting would possibly be somewhat like the experience of the blindsight patient. He is able to react (point) to objects, but he is overwhelmingly unaware of them. He is unable to describe and report on them. Of course, humans may have come to rely on reporting to such an extent, and it plays such a major role in our conscious awareness of our being, that the other components of consciousness such as sensory awareness may have been diminished.

Secondary consciousness is thus ‘thinking’ of something—reporting to yourself. This includes a knowledge of the terms and concepts relating to consciousness itself—being aware that you are being aware, and so on.

anticipation

Consciousness has a component of familiarity and of expectation. We are at ease with our surroundings and are intimately aware of them. We are able to anticipate to some extent the events of our world. Our awareness of the world does not appear to be something that needs to be reinforced or re-discovered at each instant, but rather there is a sense of temporal continuity.

The recurrency of the ABC model provides a way to make the processing of inputs supplied later to be sensitive to what happened in the processing of earlier inputs. That is, the recurrent linkages provide one of the features of consciousness, that of a kind of short-term memory. It supplies a form of expectation—a sense of what is likely to occur next.

As the network continuously collects information and self-organises within a relatively stable environment, we are, as it were, at one with our world. The world is familiar to us, and our interactions with it form a contiguous and continuous sequence. Further, because of the self-organising nature of the model, similar inputs will produce similar outputs, thus further enhancing our sense of continuity with our surroundings.

context

The impression that we ‘feel at ease’ in a particular context is due to the fact that in one sense we are re-living it. Our behaviours within a context were learned within that context, and then re-applied when that context confronts us again. Our behaviour thus may seem to be very *appropriate* and we may feel that we are using ‘common sense’ in our behaviours.

recall

Being able to recall past events and previously learned information is an important survival necessity. The recurrent loops at multiple levels of the ABC model ensure that we are able to recall these aspects of the past. The sensory recurrent loops enable us to remember past images, sounds, touches, pains, smells, fears and so on. The recurrency within the linguistic loops enable self-talk (thinking).

The recall of past events affords a sense of continuity in our awareness. The realisation that our current experiences concur with those of the past enhances our awareness.

self-knowledge

The concept of the self as a separate entity is a learned concept. This may be achieved through observation (for example, the awareness of a self-image seen in a mirror is only found in higher primates), or more usually through the use of language within a society. Our knowledge of ourselves and our place in the world is greatly enhanced through the availability of language.

As we discussed in Section 5.1.6 in relation to language, and in Section 5.2.5 in relation to automaticity, the use of language enables us to bifurcate the world into fine-grained categories and concepts, and to learn extended motor skills that would be difficult, if not impossible to achieve through observation in the absence of language. These processes greatly enhance our level of awareness of our surroundings.

unity

We are aware of a wholeness in our being—a personal unity. Our behaviours appear to be coordinated and complete.

The ABC model, in associating sensory inputs, and in concatenating various input vectors into a contiguous whole in some extended space, incorporates mechanisms which sustain this sense of unity.

As well, the self-organised physicality of the brain as described by the model means that it is integrated into its world—our being is our history of interactions within the world. We fit into our world because we have self-organised to do just that. We are not a separate, isolated entity collecting data, but a constantly changing and self-organising being that relates intimately to its surroundings.

5.3.11 Internal ‘Feels’ and Qualia

A number of philosophers, such as Chalmers (1996), contend that the aspects of consciousness that we have discussed in the previous sections—explanations of cognitive *abilities* and *functions*—are the *easy* problems concerning consciousness. They maintain that the really hard problem of consciousness is that of *qualia*.

Chalmers suggests that the current work on consciousness may be divided into two categories—the ‘easy’ problems such as “how does the brain process environmental stimulation?” or “how does it integrate information?” or even “how do we produce reports on internal states?” While these are important issues, he contends that the major issue is “why all this processing is accompanied by an experienced inner life?” Chalmers suggests that he finds himself “absorbed in an orange experience, and *something is going on*”—the experience of orange.

O'Rourke (1993) describes qualia as private, intrinsic properties (pains, colour) as compared to intentional, contentful mental states (beliefs, desires). These 'ineffable' states demand a first-person perspective and remain the sticking point of most theories of consciousness. It seems "impossible to analyse the subjective character of these phenomena ... in objective physical terms which are comprehensible to any rational individual independently of his particular sensory faculties" (Honderich 1995, page 736). Examples of qualia are the way coffee smells, the feeling of a tooth-ache, the experience of the colour red, the sound of a bird's song.

How is it that the workings of a physical device can generate *awareness*, as opposed to just *responsiveness*—how could a device have experiences rather than merely behaviour? How do we account for these internal 'feels'.

Chalmers suggests that we have great conceptual and intuitive difficulties in equating the subjective phenomenology of conscious experience with the workings of a physical device—the brain. This of course leads many (including Chalmers) to seek a dualist explanation for consciousness.

Much philosophical heat has been generated out of the fact that we only experience these feels in our own case, and hence it seems only marginally justified in assuming that others feel the same 'feels' that we do. That is, we can only account for our own subjective inner feelings and are not able to prove that others have the same feelings.

But this is surely correct—we do indeed only have access to our own feelings, conceptualisations and self-talk.

Humans within a society are able, however, to use language to subjectively compare experiences. We all have roughly the same structures of brain and body, the same physiology, and so we might expect to have similar experiences.

Qualia are not fixed entities. There is no such absolute thing as the experience of pain. For example, the phenomena of hypnosis in reducing or eliminating pain suggests that it also has some subjective component (Hilgard & Hilgard 1983).

In terms of the ABC model, most qualia are perceptual attractors, or in other words, perceptual concepts. The feel of red is the perceptual sensation one experiences with

the observation of redness.

As we all learn perceptual attractors individually, the verbal terms associated with our experiences may differ. The perceptual attractor of red for one person will not be the same as that of another. All attractors in each individuals are idiosyncratic, and so share this quality with the idiosyncratic nature of body shapes or hair colours. In this sense, then, a totally objective science of human conceptualisation not possible—our brains are irreducibly subjective.

In this vein, James ([1890] 1950, page 226) suggested that “the universal conscious fact is not ‘feelings exist’ and ‘thoughts exist’ but ‘I think’ and ‘I feel’.”

While this subjectivity is unavoidable, solipsism is not. It is only through the shared experiences of society that we acquire the words and the concepts that describe our subjective feelings. Thus, as Wittgenstein (1963) pointed out, we are unable to acquire a *private language*, as language itself is essentially social.

5.4 Common Sense and Everyday Cognition

In this section we examine the concept of common sense and our everyday thinking.¹⁴ As stated in O’Rourke (1993) “. . . ‘folk psychology’ is our shared, common sense theory of human behaviour that successfully explains, e.g., that we carry umbrellas because of our belief in their efficacy in keeping us dry.”

The current philosophical understanding of common sense in Western society is based on the metaphysical revolution of Galileo and Descartes, which in turn grew out of a tradition going back to Plato and Aristotle. This ontology defines the ‘rationalist orientation,’ and incorporates two separate domains of phenomena: the *objective* world of physical reality, and the *subjective* mental world of the individual with their private thoughts and feelings. The understanding is based on several assumptions which are taken for granted:

1. We are inhabitants of a ‘real world’ made up of objects bearing properties. Our actions take place in that world.

2. There are ‘objective facts’ about the world that do not depend on the interpretation (or even the presence) of any person.
3. Perception is a process by which facts about the world are (sometimes inaccurately) registered in our thoughts and feelings.
4. Thoughts and intentions about action can somehow cause physical (hence real-world) motion of our bodies.

Dreyfus (1992) posits three major assumptions—the psychological, epistemological and ontological—that are held (in error) by cognitivists. † These assumptions imply that humans must be a *device* which calculates according to rules on data which take the form of atomic facts, and they arise as a result of some powerful beliefs that are held within Western society.

The underlying belief is the Platonic view that all reasoning can be reduced to explicit rules, and the world reduced to atomic facts, such that the rules can be applied to the facts without the risk of objective interpretation. This belief is now supported by the invention of the digital computer, a general-purpose information-processing device. The computer performs computations on data which are reduceable to atomic elements logically independent of one another, and it calculates according to explicit rules. As put by Dreyfus (1992, page 231):

The impetus gained by the mutual reinforcement of two thousand years of tradition and its product, the most powerful device ever invented by man, is simply too great to be arrested, deflected, or even fully understood.

A number of philosophers have challenged this Platonic view, and among the pioneers in this formulation of a new approach to cognition are Heidegger, Wittgenstein and more recently, Merleau-Ponty and Gadamer.

Heidegger, in his quest for an understanding of *Being*, argues that the separation of the subjective and the objective views of the world denies the possibility of a more

†We examine these assumptions in more detail in Section D.4.

fundamental unity—*being-in-the-world (Dasein)*. The Western philosophical tradition has it that I (the subject) perceive something else (the object). To Heidegger, the simple objective stance in which reality concerns the ‘objective physical world’, and the simple subjective stance in which personal thoughts and feelings constitute the primary reality, are both invalid and must be rejected. One cannot exist without the other. The objective, external world and the subjective, inner world do not exist independently.

The ABC model suggests that we begin life with a conceptual *tabla rasa*,[†]

and through interactions with the world modify our physical being in a process of self-organisation. We are not a separate being collecting new data, but a constantly changing physical device in dynamic contact with our surroundings. We are a physically different person today as compared to our being of yesterday, and the person who wakes up tomorrow will be similar to, but not the same as, the person who woke up this morning.

We are what our interactions with the world have self-organised us to be, and our interpretations of the world are determined by our self-organised being. We are a uniquely ‘wired’ whole that is in intimate contact with the environment. We will act upon the environment and change it, and at the same time the environment will change us.

The cognitivist model suggests that our being remains constant except for the addition of new knowledge about the world. The metaphor for ABC is more akin to the reshaping of the Earth’s surface with plate tectonics, erosion, volcanic activity and so on. The surface is constantly changing and at each moment reflects a new reality. In the same way our brain is constantly undergoing a process of physical change as new attractors are formed and old attractors are modified following ongoing interactions with the environment and our own being through self-talk.

The process is one of continuity. A general purpose computer can run a word-processing package at one moment and a database package the next. But our being only changes in contiguous and continuous increments. We must constantly start from the base of

[†]By *tabla rasa* we mean that the infant is born without built-in *concepts*, but of course, in line with the proposed ABC model, there must be some inherited neural *structure*.

our current being.

Our current being includes our current concepts, our current opinions, our current prejudices, our current understanding and interpretation of the world as a whole. As Gadamer (1976, page 9) notes:

It is not so much our judgements as it is our prejudices that constitute our being . . . the historicity of our existence entails that prejudices, in the literal sense of the word, constitute the initial directedness of our whole ability to experience. Prejudices are biases of our openness to the world. They are simply conditions whereby we experience something—whereby what we encounter says something to us.

As well, conceptualisations of the world are determined by our interactions with it. We see the world in idiosyncratic ways, determined predominantly by our society, but also by our individual experiences.

In their discussion of Heidegger's philosophy, Winograd & Flores (1986, page 32) present a number of points that reflect Heidegger's views. We reproduce and extend these points here, and show how the ABC model supports and provides a mechanism for Heidegger's observations.

explicit beliefs

Heidegger maintains that we cannot make explicit all of our implicit beliefs and assumptions. His reasoning is that there is no neutral viewpoint from which we can appraise our beliefs—we must always operate within the framework of our belief structures.

There are two ways in which we can think about our understanding of our beliefs: as a function of how they relate to the beliefs of others from other cultures and backgrounds; and our (internal) understanding of our beliefs.

Heidegger is correct when he suggests that it is difficult to step outside of our belief structures (our neural attractors, sequences and associations) and try to consider alternate understandings of the world. We are shaped by our previous interactions with the environment and our culture. The ABC model suggests that we are 'soft-wired'

to think in ways determined by our history. To take on the concepts of others we will need to form neural attractors at least similar to those that form the other's view.

An alternate 'view' is an alternate behaviour. To consider another's view is to take on another (perhaps conflicting) behaviour, and to assimilate that behaviour with the essence of our being as determined by our previous history.

The second aspect of making explicit our beliefs and assumptions was perhaps not known to Heidegger. As we discuss in other sections of this chapter, we ourselves are often completely unaware as to why we make certain decisions, or why we perform certain actions. Explicit awareness of our beliefs demands that they be expressed in linguistic form, either as speech or self-talk. But as we show elsewhere, much, if not most of our behaviour is performed without any linguistic component—behaviours using perceptual attractors and sequences.

practical understanding

Heidegger maintains that practical understanding is more fundamental than detached theoretical understanding. This goes against the traditional view of Western philosophy that a detached theoretical point of view is superior to an involved practical viewpoint.

Heidegger insists that we have primary access to the world through practical involvement, and that we are always acting un-reflectively. He terms this *ready-to-hand*—our skilled, seemingly effortless behaviours in concert with our surroundings. In particular, we are all highly skilled in our everyday behaviours.

This is again in accord with the ABC model. We learn to behave within the world primarily through perceptual attractors, the direct connections between our sensations of the world and our behaviours. Working, playing, driving a car, drying the dishes—these are all behaviours that require much skill and coordination of muscles. And the coordination is with external objects that exhibit various dynamical behaviours themselves—the motion of a car, the push of the wind and the pull of gravity as we jog, the path of the tennis ball.

We do not calculate our behaviours in the manner of a computer, observing an instantaneous image of the tennis ball, solving a differential equation or two, then moving our arms and legs to the desired position to return the volley. Our brains are universal approximators able to learn complicated sequences of motor actions (given

sufficient practice). Motor skill is the fundamental ability that determines all of our behaviours, and as we discussed previously, motor skill even forms the basis of our thinking.

Theories of the behaviour of objects in the world, even our so-called common sense theories of our every-day behaviours, are secondary to our skilled motor behaviours. As we show elsewhere, we develop language and thinking as an adjunct to our perceptual behaviours, and language (including self-talk) takes only a partial initiating and correcting role in our motor behaviours, rather than being the basic mechanism proposed by cognitivism.

As we observe regularities in the world, we may put those regularities into linguistic form—for example, the abduction “red at night, shepherd’s delight, red in morning, shepherd’s warning”. The more sophisticated observations of science are expressed in the more sophisticated language of mathematics, and are called theories. But theories are observations made by a developed brain, not the mechanism of cognition.

Heidegger suggested that while detached contemplation may be illuminating, it also obscures the phenomena themselves by isolating and categorising them. Once we have a theory of some phenomenon we will tend to use this theory without further consideration or contemplation. This we saw in Section 5.3.3 where we examined the inability of subjects to provide verbal reports of their behaviours through introspection. Rather than use detached contemplation, we seem to rely instead on some socially derived theory that appears most appropriate.

Cognitivism and Western analytic philosophy, in their current position on cognition, language and thought, give primacy to detached contemplation. But while detached contemplation may allow us increasingly sophisticated and intricate world theories, it is not the principle mechanism of cognition.

Heidegger puts detached contemplation into the context of cognition as *praxis*—a concern-full acting in the world—and does not diminish its importance. But he is more concerned with the issues of context and situated action, issues which he terms *thrownness*. Thrownness is the observation that we are ‘thrown’ into situations and need to respond in context sensitive ways. Winograd & Flores (1986, page 34) give the example of chairing a meeting, and having to behave in the dynamic situations that ensue:

- you cannot avoid acting—even doing nothing is making a choice;
- you cannot step back and reflect on your actions—you often have to respond immediately with no time to analyze things explicitly in order to choose the best course of action;
- the effects of actions cannot be predicted—people may object, or intervene, and you must often ‘flow with the situation’;
- you do not have a stable representation of the situation—this may become clearer after the event, but an appreciation develops as events unfold;
- every representation is an interpretation—each person may have a different interpretation of events, and you have no way to determine right or wrong, or to evaluate the different values or emotions that drive people;
- language is action—language does not ‘state facts’ but creates a situation.

Heidegger was the first to suggest that our ordinary everyday life is more like a dynamic meeting which requires situated actions, than like the detached, contemplative and logical analysis of rules suggested by cognitivism. Our interactions with other people and with the inanimate world are much too dynamic and context dependent, and we are constantly put into a situation of thrownness. †

The ABC model is very much context driven. The concatenation of vectors ensures that diverse sources of information from within and outside our beings are taken into account in determining our behaviour. Our actions are based on previously learned behaviours that were learned in similar situations. Further, the model suggests that many of our behaviours proceed directly from sensory inputs to motor outputs without any need for a linguistic component.

To reiterate, practice is primary, and theory is derivative.

representations

Heidegger rejects *mental representations* and maintains that we do not relate to things primarily through having representations of them.

The traditional view of Western society has been that in order to perceive and relate to things, we must have some representation in our brains that corresponds to

†This is also the view held by researchers into *situated action*—see, for example, Suchman (1987). We briefly discuss the views of situated action in Section 5.5.1.

our knowledge of them. Obviously our brains must change in some way in order to recognise a new object, and so some form of 'representation' must be incorporated in the brain. But the traditional view is of an 'atomic' form of representation, invariably in linguistic terms.

But as Heidegger points out, the status of this representation is called into question if, instead of thinking of our interactions with the world as detached contemplation, we instead focus on behaviour. Heidegger's famous example is that of driving in a nail with a hammer. According to Heidegger, when we are engaged in nail-hammering-behaviour (as opposed to thinking about a hammer), we do not need to make use of any explicit representation of the hammer. Our ability to perform this activity does not come about from specific knowledge of a *hammer*, but from a learned familiarity with *hammering*.

In terms of the ABC model, we learn a nail-hammering-behaviour based on perceptual attractors—the sight of the nail, the somatosensory inputs of the hands on the nail and holding the hammer, the proprioceptive inputs of the body and arm positions and motions, even the sound of the hammer hitting the nail cleanly. This is a motor skill that can be learned without any 'knowledge' (that is, linguistically-based statements or theories) about the hammer itself.

We may have been instructed in the proper use and functions of a hammer at school, but once we are skilled at hammering we do not need to utilise these linguistic descriptions, but instead rely on our motor skills. It is only when our skill breaks down, such as the nails keep getting bent to one side, that we need to examine the hammer as a separate object (perhaps the hammer is old and mis-shapen and so we are unable to strike the nail in the required manner).

In drawing the distinction between perceptual motor skills on the one hand, and the linguistic (symbolic) mode of behaviour on the other, the ABC model provides a strong rejection of the explicit mental representations approach of current cognitive psychology, linguistics, artificial intelligence and cognitive science.

Explicit representations are human inventions which exist within society. While they may indeed be remembered by the brain, they in no way form the primary mechanism of cognition.

meaning is social

Heidegger maintained that our understanding of the world is fundamentally social and cannot be reduced to the meaning-giving activity of individual subjects.

The cognitivist (rationalist) view of cognition centres around the individual. For example, in linguistics we study language by examining the behaviour and characteristics of an individual language learner or language user. We study reasoning by examining the deductive processes of an individual. But this is an inappropriate starting point according to Heidegger. After all, these abilities were learned within a society and from social interactions. Social activity is the ultimate foundation of intelligibility, even of our very sense of existence

According to Heidegger, a person is not primarily an individual subject or ego, but rather a manifestation of *Dasein* (essentially our *being* which can be *made*, or perhaps better, *formed*) within a space of possibilities, situated within a world and within a tradition.

Within the ABC model, the perceptual attractors and linguistic attractors of each individual are determined by the individual's interactions with the world, and in particular, the society in which the individual is raised. The self-organisation of the individual's neural attractors comes about within the structures, language usage, prejudices and myths of that society.

Further, because of their similar social interactions, schooling, training and experiences, the brains of the individuals in a society will tend to be 'wired similarly'. There will be an overall similarity in world-views, even though individual differences will exist because of the essentially individualistic upbringing that we each experience.

breaking-down and readiness-to-hand

To Heidegger, objects and properties are not inherent in the world, but only come about in the event of a *breaking-down*, at which time they become *present-at-hand*. Consider again the example of the person using a hammer to drive in a nail. Heidegger suggests that to the person doing the hammering, the hammer as such does not exist. It is a part of the background of *readiness-to-hand* that the hammerer takes for granted without recourse to any explicit recognition or identification of the hammer as an object. It forms a part of the hammerer's world, but is not present any more than are the tendons

of the hammerer's arm or the coat that he wears.

Looking at this from the point of view of the ABC model, the hammerer is involved in a perceptual skill without recourse to the linguistic (or self-talk) component of cognition. Recognition of the object as a separate entity would require conscious (linguistic) awareness of the hammer, and this is not required for the action of hammering.

In learning the skill of hammering, the neuronal weights appropriate to the action of hammering are modified into new perceptual attractors for the nail-hammering-behaviour. The appropriate sequences of muscle actions, the muscle strengths required and so on are learned in association with the use of the hammer. The hammer essentially becomes an addition to the hammering person's body in the behaviour appropriate to hammering.

We only become conscious of the hammer itself as a separate entity during the hammering behaviour if there is some kind of breaking down or *unreadiness-to-hand*. Its 'hammerness' emerges if it breaks or slips from our grasp or mars the wood. It also comes to mind when we can't find it.

In this case, the direct perceptual skilled behaviour is interrupted by a need to refer to the hammer via our linguistic (or self-talk) abilities. The hammer is then seen in a different light, as a separate object rather than an integral and yet invisible component to skilled behaviour.

thrownness

Heidegger concluded that our acts always happen within thrownness. They do not result from a process (conscious or non-conscious) of representing, planning, and reasoning, but always occur within a dynamic world that frequently demands instant behaviour.

Although we engage in conscious reflection and systematic thought at various times, these are secondary to the pre-reflective experience of being thrown into a situation. We are not able to fully disengage ourselves from the situation and observe the world as a detached observer, but are always engaged acting within a situation.

The ABC model is potentially able to behave in such a dynamic environment as it is contextually driven, and a hardware version should be extremely fast, but this capability needs to be shown through further experimentation.

5.4.1 Biology of Learning

The ABC model suggests that organisms which engage in activities that are triggered by changes in their environment, learn from these interactions. The interactions will change the structure of the organism, and hence its future behaviour. Cognition must be viewed not as an activity in some mental realm, but as a pattern of activity that is relevant to the functioning of the person or organism in its world.

Our 'common sense' theory of knowledge suggests that what we know is a representation of the external world. In this cognitivist view, there are two parts to the brain—the representations and the part that manipulates them. But this view must be questioned.

The work of Maturana is important in this regard. For example, Maturana (1980, page 45) examines cognition and learning from a biological viewpoint, and concludes that:

Learning is not a process of accumulation of representations of the environment; it is a continuous process of transformation of behavior through continuous change in the capacity of the nervous system to synthesise it. Recall does not depend on the indefinite retention of a structural invariant that represents an entity (an idea, image, or symbol), but on a functional ability of the system to create, when certain recurrent conditions are given, a behavior that satisfies the recurrent demands or that the observer would class as a re-enacting of a previous one.

In the ABC model, the brain is 'moulded' by learning. In this regard, our behaviours are like learned soft-reflexes, with a continuum between instinctive behaviours (innate hard-reflexes) and learned behaviours (soft-reflexes). Maturana (1978, page 45) supports this view:

If ... the observer wants to discriminate between learned and instinctive behavior, he or she will discover that in their actual realization, both modes of behavior are equally determined in the present by the structures of the nervous system and organism, and that, in this respect, they are indeed indistinguishable. The distinction between learned and instinctive behavior

lies exclusively in the history of establishment of the structures responsible for them.

Maturana (1978, page 39) makes the point that learning is statistical, and that the prior history of the organism determines its brain structure. The statistics are not collected as a side issue, and then used to calculate a behaviour, but rather go to determine the neural connections continuously—the embodiment of the connections are formed by the history:

... history becomes embodied both in the structure of the living system and the structure of the medium even though both systems necessarily, as structure determined systems, always operate in the present through locally determined processes History is necessary to explain how a given system or phenomenon came to be, but it does not participate in the explanation of the operation of the system or phenomenon in the present.

The statistics are collected incrementally, and the resultant brain structures mean that the creature is ready to enact those statically moulded behaviours at any time, without the need for calculations.

5.4.2 The Notion of Theories

Dreyfus & Dreyfus (1988, page 318) suggest that Western philosophy:

... has from the start systematically ignored and distorted the everyday context of human activity. The branch of the philosophical tradition that descends from Socrates through Plato, Descartes, Leibniz, and Kant to conventional AI takes it for granted, in addition, that understanding a domain consists in having a *theory* of that domain. A theory formulates the relationships among objective, *context-free* elements (simples, primitives, features, attributes, factors, data points, cues, etc.) in terms of abstract principles (covering laws, rules, programs, etc.).

The apparent success of the role of theories in the ‘natural sciences’ compounded this view that the world is not only orderly, but that any orderly domain must be composed of context-free elements with some form of abstract relations between those elements. This was assumed to account for the order of that domain and for man’s ability to act intelligently within it. However, a number of philosophers, including Leibniz and most subsequent theorists, extended this rationalist account to all forms of intelligent activity, including everyday behaviour. For example, Leibniz (1951, page 48) wrote:

The most important observations and turns of skill in all sorts of trades and professions are as yet unwritten. This fact is proved by experience when passing from theory to practice we desire to accomplish something. Of course, we can also write up this practice, since it is at bottom just another theory more complex and particular.

The cognitivist approach follows this tradition. Based on the assumption of a transfer of methods that have been developed by philosophers and that are successful in the ‘natural sciences’, the cognitivist contends that the domain of everyday cognition must be formalisable. This then demands that in order to understand cognition we must find the context-free elements and principles, and then base a formal, symbolic representation on this theoretical analysis.

We state elsewhere that theories in the natural sciences (for example, physics) are the result of forming *approximations* to actual behaviours by fitting formal (mathematical) systems to the data. For example, Newton’s laws of motion are an approximation to the actual behaviour, and this approximation was improved upon by the relativistic theories of Einstein. These theories will no doubt be superseded by a better mathematical ‘fit’ of some future theory. Here, the theory is not the essence but an approximate description. Subsequent work will find other ‘mechanisms’—for example, the gravitational ‘action at a distance’ mechanism replaced by the exchange of virtual particles. The description may proceed ever deeper into a fractal world without ever necessarily coming to any end-point or final resolution. A new, ‘better’ description may be just around the next experimental corner. In other words, even in physics, the formal system is a description, not the essence of the phenomenon.

Wittgenstein and Heidegger both then call into question the very tradition on which the symbolic information-processing model is based. Both were concerned with the importance of everyday behaviours, and both held that one could not have a *theory* of the everyday world.

The cognitivist model requires the existence of a set of primitives from which all other knowledge can be generated by some form of combination of these primitives. Western philosophy and AI chose essentially linguistic terms (symbols) as their primitives.

After writing *Tractatus* (Wittgenstein 1960b), the apotheosis of the cognitivist tradition, Wittgenstein spent many years trying to find the atomic facts and basic objects required for his reductionist theory. He came to the conclusion that the rationalistic philosophical position was untenable, and in *Philosophical Investigations* (Wittgenstein 1963) attacked his own previous writings and views.

Wittgenstein argued that the analysis of everyday situations into facts and rules is only meaningful in some context and for some purpose. In trying to find the ultimate context-free, purpose-free elements (as is required in order to provide a computer program with the necessary primitive symbols), we find that we are in effect trying to free aspects of our experience of just the pragmatic organisation which makes it possible to use them intelligently in coping with everyday problems.

Wittgenstein criticised the logical atomism of the *Tractatus* (Wittgenstein 1963, page 21):

What lies behind the idea that names really signify simples'?—Socrates says in the *Theaetetus*: 'If I make no mistake, I have heard some people say this: there is no definition of the primary elements—so to speak—out of which we and everything else are composed . . . But just as what consists of these primary elements is itself complex, so the names of the elements become descriptive language by being compounded together.' Both Russell's 'individuals' and my 'objects' (*Tractatus Logico-Philosophicus*) were such primary elements. But what are the simple constituent parts of which reality is composed? . . . It makes no sense at all to speak absolutely of the 'simple parts of a chair'.

Wittgenstein rejected linguistic terms as the primitives (or simples) for the so-called necessary and sufficient conditions required for rule-like descriptions of behaviours and objects.

Heidegger held similar views following his extensive phenomenological investigation into the everyday world and everyday objects. He concurred with Wittgenstein that the everyday world could not be represented by a set of context-free elements.

Heidegger (1978) concluded that there were other ways of ‘encountering’ things than relating to them as objects defined by a set of predicates. His analysis of the use of tools (such as the use of a hammer) suggested that our behaviours use perceptual and motor skills in a particular context. The skill and context of our everyday behaviours need not be represented as a set of facts. The context and our everyday ways of skillful coping in it are not something we consciously think about, but are rather an integral part of our being.

In terms of the ABC model, our behaviours are learned temporal sequences which result from our previous interactions with our environment. The skills in performing these everyday activities are soft-wired into our brains, so that given a certain context we will be able to ‘select’ an appropriate behaviour. Our skilled behaviours are in the weights of the neural connections, not in linguistic rules, and *conscious* consideration (thinking) in linguistic terms is not necessary for their execution.

The ABC model proposes that a theory of the everyday world (as rationalist philosophers have always held) is not the mechanism of cognition. Common sense is rather a combination of skills, practices, and discriminations which have no representational content to be explicated in terms of elements and rules.

Husserl was another philosopher who examined the rationalist proposal. He was forced to conclude after 25 years of unsuccessful attempts to solve the problem of putting common sense into a theoretical framework that the task was ‘tremendously complicated’ (Husserl 1970, page 291). Indeed, the task is still unsolved after some two thousand years of searching—from Socrates through Leibniz to early Wittgenstein. The ABC model provides an alternative paradigm.

5.5 Comparison with Other Models and Related Work

There is not much point in comparing the ABC model to other work that uses the cognitivist paradigm. The two are incompatible, and this thesis proposes that the cognitivist model of cognition is incorrect and hence irrelevant. However, there are a number of other research areas that have had a strong influence on the development of the ABC model and so we briefly discuss these in the following sections.

5.5.1 Situated Cognition and Situated Action

A large body of work has been concentrated around the need for the inclusion of context dependent and dynamic behaviours in any discussion of cognition. These studies suggest that the *situation* determines our behaviour and not the cognitivist notion of ‘planning’.

For example, Suchman (1987) contends that our actions are influenced by the situation in which we find ourselves, and that we must behave in ways that take account of the immediate circumstances in order to satisfy any overall goal that we may have.

Suchman, an anthropologist, gives an example case of a canoeist attempting to traverse some white water. While the canoeist may decide on an overall strategy for moving from where she currently finds herself to a point beyond the rapids, once she starts into the white water the dynamics of the situation demand that she rely more on her skill and assessment of the immediate situation, and relegate the plan to just overall guidance. The plan may even have to be abandoned if the canoeist deviates too far from it because of intervening circumstances.

Suchman maintains that *plans*—potential actions that are constructed ahead of time, are less significant than *situated actions*—decisions made on the spot in response to the current context. Plans may play a part in determining what is happening and what to do next, but they are not the primary factor in the overall orderly nature of action.

The researchers who comprise this group are diverse and in many cases not entirely in agreement. However, they all hold the view that, in some way, the current cognitivist

approach is inadequate for an understanding of cognition. Clancey (1993) provides a view of Situated Cognition (SC) based on neuropsychological research, which “rejects the hypothesis that neurological structures and processes are similar in kind to the symbols we create and use in our everyday lives” (Clancey 1993, page 87).

Slezak (1993, page 92) also makes this point: “the symbolic approach conflates ‘first-person’ representations in our environment (e.g., utterances and drawings) with ‘third-person’ representations (e.g., mappings a neurobiologist finds between sensory surfaces and neural structures)”.

Clancey (1993) quotes five central claims of situated action (SA) which we reproduce in part as an endnote as they exactly support the ABC model. ¹⁵

Other works which are regarded as being within the situated cognition umbrella are in education (Greeno 1989), ecological psychology (Kugler & Turvey 1987, Shaw, Turvey & Mace 1982, Warren & Whang 1987), the ethnographic study of activity (Cole 1990, Hutchins 1990, Hutchins 1991, Laboratory of Comparative Human Cognition 1983, Lave 1988, Lave & Wenger 1991) and philosophical situation theory (Barwise 1989, Barwise & Etchemendy 1987, Barwise & Perry 1983, Devlin 1991). A series of articles on everyday cognition is found in Rogoff & Lave (1984).

A more detailed discussion of the issues relating to the situated cognition paradigm can be found in a number of articles, including Verschure (1993), Slezak (1993), and a series of discussions published in the *Cognitive Science* journal (Norman 1993, Vera & Simon 1993a, Greeno & Moore 1993, Agre 1993, Suchman 1993, Vera & Simon 1993b, Clancey 1993, Vera & Simon 1993c).

5.5.2 Neural Darwinism

The ABC model probably owes its greatest initial debt to the Neural Darwinism model of Edelman and his coworkers (Edelman 1989a, Edelman & Finkel 1984, Edelman 1978, Edelman 1985). Edelman has developed an extensive theory of brain function, based in part on evolutionary processes in brain development, and his work has opened up many new areas for research.

The theory also strongly relies on the re-entrant linking of topographic maps, in combination with feedback from the environment following action. The theoretical aspects have been developed into the 'Darwin' series of real and simulated robotic systems, and the theory provides a basis for an explanation of the learning mechanism of animals.

The ABC model differs in many respects from the Neural Darwinism model, especially as regards to language and self-talk. As such, we consider that the ABC model goes much further than Neural Darwinism in providing a complete model of cognition.

The ABC model also suggests that a major component of learning is the self-organisation of SOM surfaces, whereas the Neural Darwinism model implies a greater role for evolutionary factors. Other relevant papers are Edelman (1989a), Reeke Jr. & Sporns (1993), Reeke Jr., Sporns & Edelman (1990), Reeke Jr. & Sporns (1990), Reeke Jr. & Sporns (1991), Sporns, Tononi & Edelman (1991), Sporns, Gally, Reeke Jr. & Edelman (1989), Reeke Jr. & Edelman (1988), Edelman et al. (1992), Reeke Jr. & Edelman (1984), Edelman & Reeke Jr. (1982) and Smoliar (1989).

Chapter 6

Summary and Ongoing Research

The principal body of evidence for the symbol-system hypothesis ... is negative evidence: the absence of specific competing hypotheses as to how intelligent activity might be accomplished whether by man or by machine.

Newell & Simon (1981, page 50).

We trust that we have shown in this thesis that this is no longer the case, and that we have shown that the ABC model offers an alternative and more appropriate model of cognition.

The achievements of this thesis can be summarised as follows:

- We have shown why the current primary paradigm of cognition, the cognitivist approach, is inadequate to explain human and animal cognition.
- We have also pointed out the failings of the somewhat simplistic treatment of cognition afforded by current connectionist models.
- In Chapter 2 we introduced a powerful new model for the learning and production of temporal sequences.
- In Chapter 3 we introduced a full model of cognition and conceptual development, including the development of language and self-talk.

- In Chapter 4 we took a fresh look at the visual system and introduced a new model of vision consistent with the ABC model.
- In Chapter 5 we examined some further implications of the ABC model, developing and extending our hypothesis for the role of language and self-talk in cognition. As well, we provided explanations for implicit and explicit learning, skill development and automatisisation using the model. We also examined consciousness in relation to the ABC model.

The ABC model provides a radical departure from previous models of cognition, giving an integrated, single-architecture model that overcomes many of the problems associated with previous models. It is an integrating model in many respects, indicating consistent mechanisms for innate and learned behaviours, and showing a common mechanism for language, self-talk and motor skills. The model provides a simpler, more cohesive view of cognition.

6.1 Future Research

The ABC model opens up a number of new and exciting research areas. Some of these are discussed briefly here.

theoretical development

A number of areas were mentioned within the text as being set aside for future research. As well, the ABC model opens up various other areas in theoretical cognitive science that have not been discussed in this thesis;

- categorisation, including the notion of ‘basic-level’ categories (Rosch & Mervis 1975, Rosch 1988),
- colour categorisations (Lakoff 1986),
- the notion of relativity of language and conceptualisation as introduced by Sapir ([1929] 1949) and Whorf (1956),
- child development in the spirit of Vygotsky & Luria (1993),

- second language acquisition, creoles and pidgin languages (Deuchar 1987, Aitchison 1987, White 1991),
- agency and free will.

temporal sequence learning

The LAPS model of temporal learning is to be studied in much more detail, with the intention of developing a general tool for temporal sequence learning.

parallel hardware

As discussed in the text of the thesis, current models of cognition are severely hindered by their need to use serial binary computer hardware. This means that for all but the simplest of cases, their performance is simply too slow to enable real-time decision making to occur.

One of the major areas for future research is the development of the ABC model in hardware. Our belief is that current parallel models (the SIMD and MIMD architectures) are really just multiple-serial devices with severe problems of co-ordination and synchronisation.

Truly massive parallelism requires millions (or even billions) of operations to be happening at the same time, with only local (and not global) constraints. It will only be when these tools are available that full-scale cognitive modelling will become feasible.

The ABC model will be developed as a hardware device.

machine learning system

The ABC model lends itself to a general purpose machine learning tool. The system will be able to associate inputs of various kinds in a manner that avoids questions of metric, and will learn temporal sequences of these associations. As well, the system will be able to be trained to emit certain behaviours (such as a language descriptions) when certain events occur.

6.2 Conclusion

Smith (1978, page 136) notes that in 1543, Vesalius, regarded as the founder of modern neuroanatomy, wrote (Vesalius 1543 [1901], page 255):

But how the brain performs its functions in imagination, in reasoning, in thinking, and in memory . . . I can form no opinion whatever. Neither do I think that anything more will be found out by anatomy or the methods of those theologians who deny to brute animals all power of reasoning and indeed all facilities belonging to what we call the chief soul.

Some three hundred years later, Goltz (1888, page 130) concluded that “the research into the functions of the various parts of the brain is a very old question. Many have tried and many have erred.”

We hope that the ABC model adds something to a better understanding of the brain.

Appendix A

Neuroanatomy, Neurophysiology and Neuropsychology

A.1 Introduction

Some contend that neuroscience is a borderline discipline for cognitive science—see, for example, Gardner (1985, page 7)—but the view taken in this thesis is that it is of primary importance. It is pointless developing a science of cognition without bedding that theory firmly on the empirical bedrock of neuroscience. This appendix examines some of the known structure and behaviour of the brain in order to provide a biological basis for the ABC model.

It is only in the past 25 years that a coherent understanding of the cortical architecture has been achieved. The overall structures and connections of the various sensory systems of the cortex—the visual, auditory, and somatosensory systems, as well as the association areas, frontal lobes and motor areas—have been recorded and mapped, and a general framework is beginning to emerge. This overall cortical structure is the topic of the current appendix. †

The Brodmann (1909) mapping of the broad cortical regions, was extracted on the basis of cellular organisation (cytoarchitectonics). These numerical designations are shown

†We restrict our attention to the phylogenetically more recent neocortex.

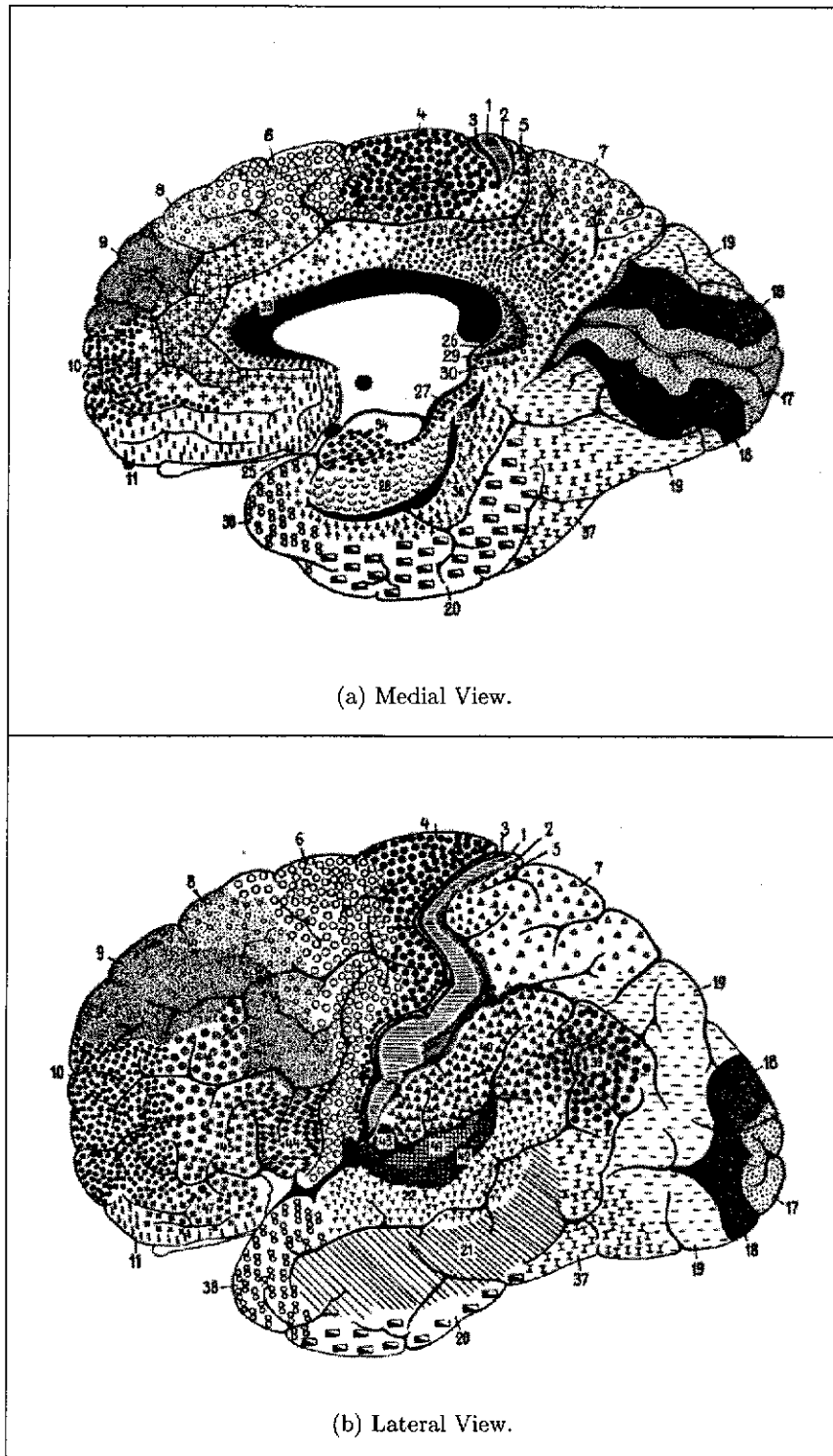


Figure A.1: Brodmann's Mapping of Cortical Regions.

in Figures A.1 (a) and (b) and will be the major designations used in this appendix. [†]

Damasio (1989) suggests that, broadly speaking, the human cerebral cortex may be divided into four distinct regions:

- *early and intermediate sensory cortices*, including the primary visual, auditory, and somatosensory regions (fields 17, 41/42, and 3/1/2) and the surrounding association cortices (fields 18/19, 7, 39, 22, 40, 5),
- *temporal “integrative” cortices* (fields 37, 36, 35, 20, 21, 38, 28) and including neocortical as well as limbic and paralimbic areas,
- *frontal “integrative” cortices* (fields 44, 45, 46, 9, 10, 11, 12, 13, 25) which includes prefrontal neocortices as well as limbic,
- *motor and premotor cortices* (including fields 4, 6, and 8).

We examine the primary visual regions more fully in Chapter 4 where we look at the ABC model in relation to the visual system, and also in Appendix C where we discuss the simulation of the ABC model.

In this appendix we look at some of the recent neurobiological findings, in particular the evidence for topologically organised mappings (SOMs), the connections and functions of the association areas, and the role of the prefrontal cortex in temporal integration.

A.2 Evolution of the Neocortex in Mammals

In most species of mammal, the structure of the neocortex is extremely similar—it consists of six layers, has only a small number of cell types, one of which, the pyramidal cell, accounts for over half of all cells, and its pattern of connectivity, both locally and globally within the cortex, is very uniform, as are its subcortical connections. Except

[†]The anatomical terms of directions and relationships are for the human brain: anterior or rostral (*front or face viewpoint*), superior or dorsal (*top*), posterior or caudal (*back*), inferior or ventral (*underside*), medial (*midline*) and lateral (*outside*).

for being simplified in some orders, this structure has not changed in the evolution of mammals.

The same overall structure accounts for the major expansion of 'association' cortex in the primate order. Even language, thought by some to be the most remarkable achievement of the brain, is found in areas of the brain which have the identical structure.

This suggests to some that "this structure embodies a basic 'computational' module so versatile that it can be hooked together in ever larger configurations and still function, with ever increasing subtlety, to both analyse sensory input and organise motor actions" (Mumford 1991). This is exactly the case for the ABC proposal.

The fact that this cortical structure is present in much simpler animals suggests that its structural complexity may have been overestimated by most current theorists, and that it may not be 'the most complicated structure in the universe' as stated in some elementary texts. More general constructs, such as those suggested by the ABC model, may be more appropriate.

A.3 Thalamus

All cortical input, except for the olfactory sense, is transmitted to the cortex via the thalamus. The thalamus is made up of some fifty nuclei positioned at the top of the brain stem. It is composed of two halves, one in the middle of each cerebral hemisphere.

Each area of the cortex is reciprocally connected in a dense, continuous fashion with a particular nucleus in the thalamus. For example, the primary visual area VI (area 17) is reciprocally connected to the lateral geniculate nucleus (LGN) in the thalamus. Another example is the primary motor area (area 4) which is reciprocally connected to the posterior ventral lateral nucleus (VLp).

For sensory signals, the thalamus is a 'relay station', directing the signals to the primary sensory areas of the cortex. For example, the nucleus VLp nucleus transmits motor signals from the cerebellum to the motor cortex. However, other nuclei collect their inputs from subcortical structures, such as the superior colliculus and amygdala. In

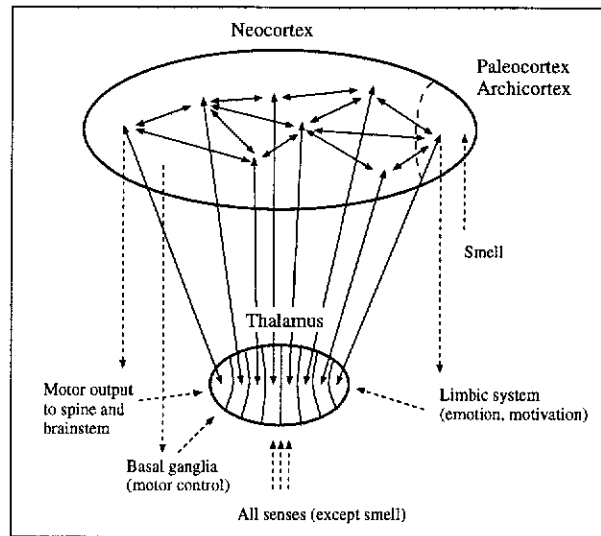


Figure A.2: Schematic of Cortico-thalamic Connections. (Taken from Mumford (1991)).

general terms, each area of the mammalian cortex receives input, via the thalamus, from the sub-cortical structure which performed a similar cognitive function in more primitive animals.

For example, in primitive animals, the integration of visual, auditory and tactile information is carried out in the superior colliculus (tectum). In mammals, the superior colliculus projects to the pulvinar complex in the thalamus, and then to the association areas of the occipital, parietal and temporal lobes which carry out the same functions.

In primates, the visual input to the cortex via the collicular-pulvinar path plays a smaller and smaller role compared to the phylogenetically later pathway from the retina to the LGN to the primary visual cortex. An estimate of the importance of this secondary visual pathway can be obtained by considering the degree of blindness exhibited by animals in which VI has been destroyed. These animals must then rely wholly on the secondary pathway. The impairment of cats is much less than for monkeys, and humans lose all (conscious) sight and exhibit a phenomenon known as *blindsight*.[†]

As well as those nuclei which have specific regional connections to the cortex, some nuclei in the thalamus have no specific connections but have diffuse (non-specific) con-

[†]We discuss blindsight briefly in Section 5.2.2.

nections. These neurons synapse over large portions of the cortex (Jones 1985), and seem to play some regulatory role. The ABC model suggests that these neurons could play a role in value-based (emotional and motivational) gain control. If some particular external or internal condition needs future avoidance (such as the painful experience caused by touching a naked flame) or future strengthening (such as successful food-gathering actions), then a higher learning rate is required. For other more mundane events, the normal learning and self-organisation (via statistics) will suffice.

Interposed between the thalamus and the cortex is a thin layer of cells on the surface of the thalamus known as the reticular complex (RE thalamus). All pathways between the cortex and the thalamus pass through this region. Afferent pathways from both the cortex and the thalamus excite RE cells, which in turn send inhibitory axons both to each other and back to the thalamus to the area of the origin of the pathway. The RE neurons are inhibitory, and Steriade, Domich & Oakson (1986) have shown that peaks of activity in a part of the RE thalamus occur at the same time as peaks of activity in the corresponding nuclei of the thalamus itself. The exact function of the RE thalamus is unknown (Steriade & Llinas 1988, Crick 1984, Sherman & Koch 1986), but we propose that it has a synchronisation role in that it allows for the insertion of appropriate vectors into system, and provides some form of 'clocking' mechanism. Both are required by the ABC model, at least per module.

One hypothesis of the function of the RE was given by Crick (1984) (but rejected by him as 'a not very exciting conclusion'). This was to *even out* the activity of the thalamus.

“... it is at least possible that the effect of the ... neurons of the reticular complex on the thalamic relay cells is to produce a brief burst of firing in response to incoming excitations, followed by a more prolonged inhibition.”
(Crick 1984, page 4588)

This view is consistent with ours, and potentially very exciting. However, more work on this will need to be left to future research.

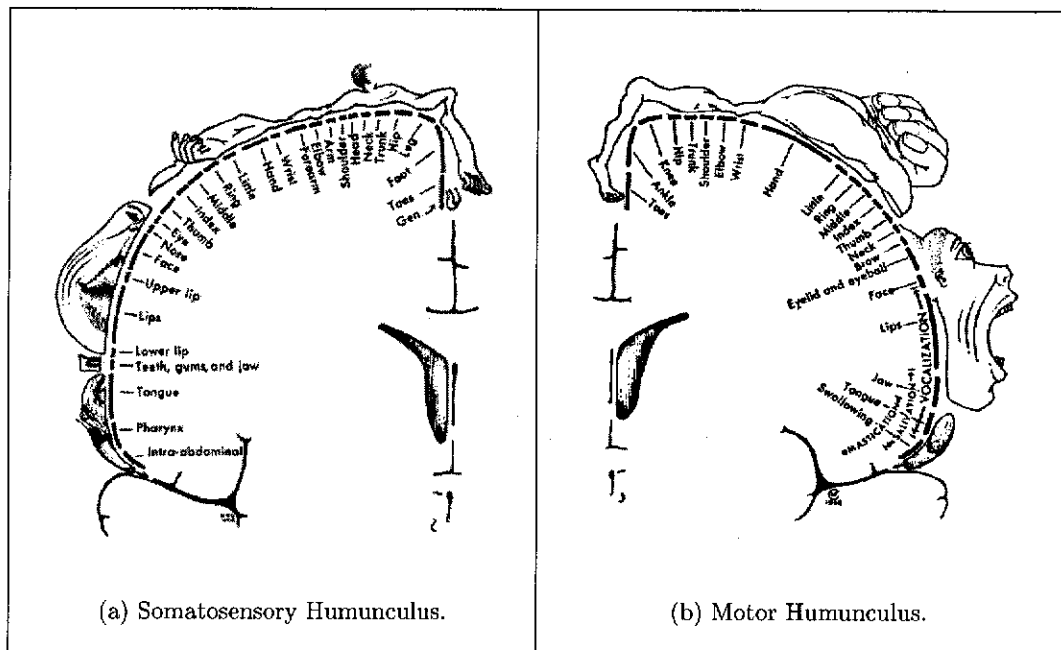


Figure A.3: Topological Mappings.

A.4 Topological Organisation of the Cortex

It is widely accepted from neurobiological research that there exist multiple topological map structures in the cortex and other brain structures such as the cerebellum and certain nuclei. These include the ocular dominance columns (Hubel & Wiesel 1977), orientation columns (Hubel, Wiesel & Stryker 1978), frequency columns (Knudsen, DuLac & Esterly 1987), colour maps (Zeki 1980), somatosensory patches (Kaas 1983), tonotopic or auditory-frequency maps (Tunturi 1950, Tunturi 1952, Reale & Imig 1980), and echo delay and Doppler shift maps in bats (Suga & O'Neill 1979).

The magnification factor of these maps depends on the behavioural importance of particular signals. For example, recordings from the foveal part of the retina, and of the fingertips and lips are greatly magnified in proportion to those of other parts. The somatosensory humunculus, an indication of the relative magnification of the bodily parts as sensory inputs, is shown in Figure A.3 (a). Figure A.3 (b) shows the motor humunculus which gives a picture of the relative amount of motor cortex devoted to efferent connections to the various body parts.

As well as sensory inputs, more conceptual components of the brain also appear to

be topologically mapped. Ritter & Kohonen (1989) suggest that evidence for a fine-structured mapping of conceptualisation comes from an indirect source—cases of selective deficits as a result of strokes or brain injuries (Warrington & McCarthy 1987). Here the deficits are more of a semantic nature, and include the inability to use concrete (impaired) versus abstract (spared) words (Warrington 1975), a deficit in the use of inanimate versus animate words (Warrington & McCarthy 1983, McCarthy & Warrington 1988), living objects and food (impaired) versus inanimate (spared) (Warrington & Shallice 1984), and even a selective impairment of sub-categories such as indoor objects (Yamadori & Albert 1973), body parts (McKenna & Warrington 1978), and fruits and vegetables (Hart, Berndt & Caramazza 1985). See Goodglass et al. (1986) and Caramazza (1988) for review articles on categorical impairment.

Topological relationships for the association areas are discussed in Section A.5.

As well as the input sensory areas and association areas, topological relationships have also been established in the frontal lobes. For example, a pattern of topological relationships has been established that applies to all the thalamic projections to the prefrontal cortex Fuster (1985, page 152). Further, analysis of the known cortico-cortical connections of the prefrontal cortex indicates a degree of topological order, especially in the primate. A mutual correspondence is seen between parts of the prefrontal cortex and other cortical areas (Fuster 1985, page 154). Different prefrontal areas receive projections from different parasensory areas (Jones & Powell 1970, Chavis & Pandya 1976, Jacobson & Trajanowski 1977), and groups of projections from distant and widely separated cortical sources converge on neighbouring or overlapping areas of the prefrontal cortex. A topographical distribution in the prefrontal cortex is also evidenced by the finding that units of like properties tend to cluster together (Fuster 1973, Fuster, Bauer & Jervey 1982), and the fact that a semblance of columnar distribution has been observed (Fuster et al. 1982).

From this wealth of evidence, it seems clear that topological ordering is a major component of the processes of the brain. The next question to ask is whether this structure is innate or the result of self-organisation.

A.4.1 Plasticity and Self-Organisation

Cortical map structures are constructed in a use-dependent manner. Extensive studies of neuronal plasticity in the cat visual cortex are reported by Kasamatsu (1983). Many concern the formation of abnormal ocular dominance columns in cats and monkeys reared under extraordinary visual conditions (Hubel, Wiesel & LeVay 1977, LeVay, Stryker & Shatz 1978, LeVay, Wiesel & Hubel 1980, Stryker & Harris 1986). These results indicate a critical period of development during which the neurons of the visual cortex exhibit considerable plasticity. For example, a monocular lid suture of only three days is sufficient, if done at a particular time in early post-natal life, to force virtually all cortical cells to change their preferred ocularity to the non-deprived eye (Hubel & Wiesel 1970, Olson & Freeman 1975). In the normal visual cortex, more than 80% of cells are driven through both eyes (Hubel & Wiesel 1962). Restoration of visual functions may be achieved, but may vary from considerable recovery to little reversal of the deficit (Kasamatsu 1983, page 15).

The formation of a regular pattern of orientation-selective neurons during development and maturation of the striate cortex suggests self-organisation of the sensory map (Götz 1988, page 214). The self-organisation of the nervous system and the consequential formation of cortical maps has been studied quite extensively (von der Malsburg 1973, Swindale 1980, Kohonen 1982, Miller, Keller & Stryker 1989).

Cortical plasticity due to self-organisation is most evident in the somatosensory system where a number of experiments have shown dramatic reorganisation of afferent receptive fields following modifications to the animals external sensors (Florence & Kaas 1995, Kaas et al. 1983). For example, Merzenich & Kaas (1982) (see also Kaas et al. 1983) severed some of the nerves in the hand of a monkey. The somatosensory cortical map of the hand reorganised over the following few months so that the deactivated sections became reactivated and associated with other undamaged sections of the hand. The overall topographical organisation was retained. Other similar results were obtained following the removal of a digit in raccoons (Welker & Seidenstein 1959, Turnbull & Rasmusson 1991) in which the cortex formerly activated by the digit gradually became responsive to the adjoining digits.

Other results of reorganisation of the somatosensory cortex following the removal of inputs include the lesioning of the follicles of rat whiskers (Kandel & Schwartz 1983, Kossut, Hand, Greenberg & Hand 1988, Levin & Dunn-Meynell 1991), and the surgical separation of fused fingers in humans (Mogilner, Grossman, Ribory, Joliet, Volkmann, Rapaport, Bensley & Llinás 1993). See also Merzenich et al. (1984), Kaas (1995) and Gilbert (1995).

Similar results are observed with retinal ablation (Chino et al. 1992). For example, Gilbert (1995, page 79) describes an experiment in which a lesion is made to the photoreceptor layer of the retina, removing visual input from the cortical region representing that retinal area, as shown in Figure A.4. Over a period of two months, the topography of the cortex was reorganised with a decreased representation of the ablated retinal area and an increased representation of the area surrounding the lesion.

This is entirely consistent with self-organisation. The re-mapping is simply a process of self-organisation as new 'statistics' enter the system.

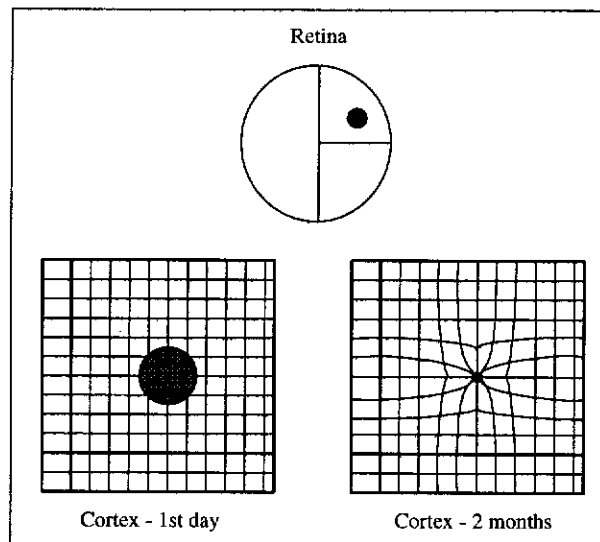


Figure A.4: Cortical Plasticity.

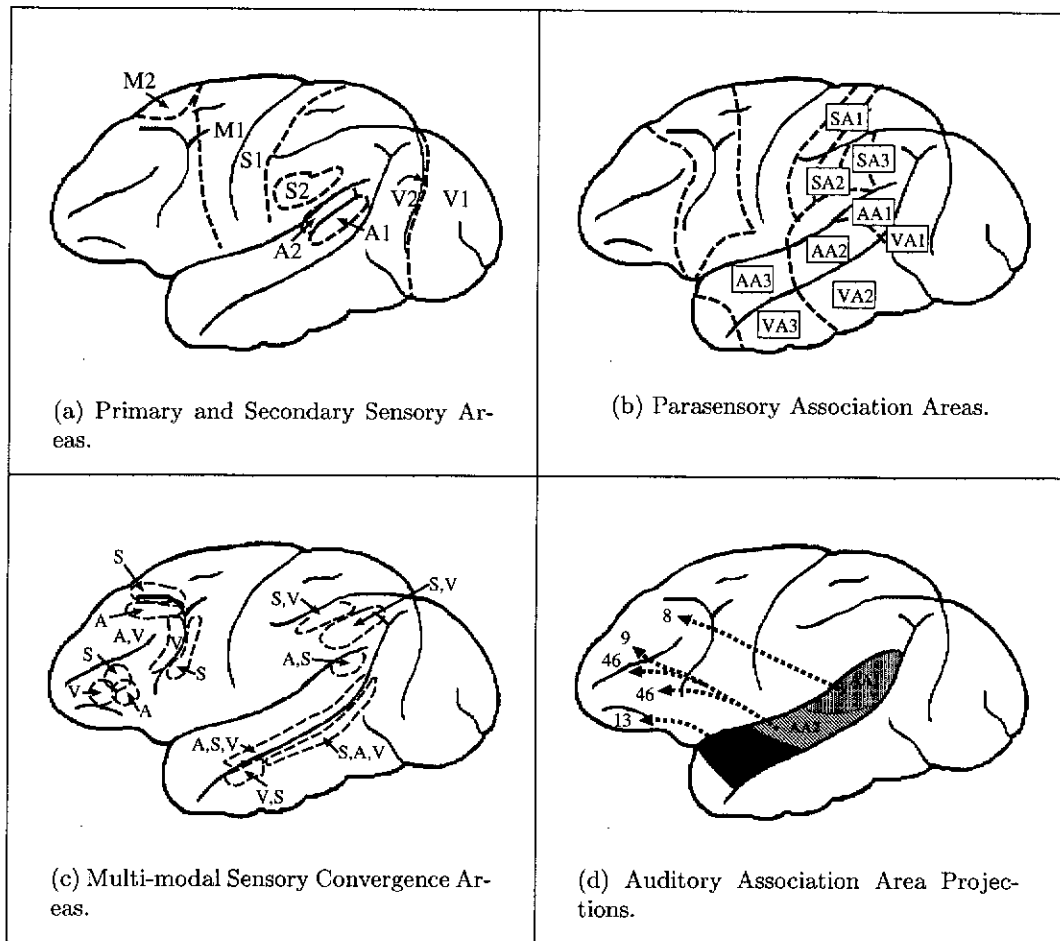


Figure A.5: Areas of Cortex of Rhesus Monkey.

A.5 Association Cortices

The most extensive cortical areas in the brain of primates is comprised of cortical association areas. † In evolutionary terms, these areas represent the focus of greatest differentiation between the brains of primate and non-primate species, and have long been viewed as crucial to higher cognitive and behavioural functions. Pandya & Yeterian (1985) discuss the connections and architecture of the rhesus monkey, which is basically similar to the human brain structure and has well developed cognitive and behavioural capacities, but which has been extensively studied.

†The experimental details discussed in this section draw heavily on an excellent review article on the cortical association areas by Pandya & Yeterian (1985).

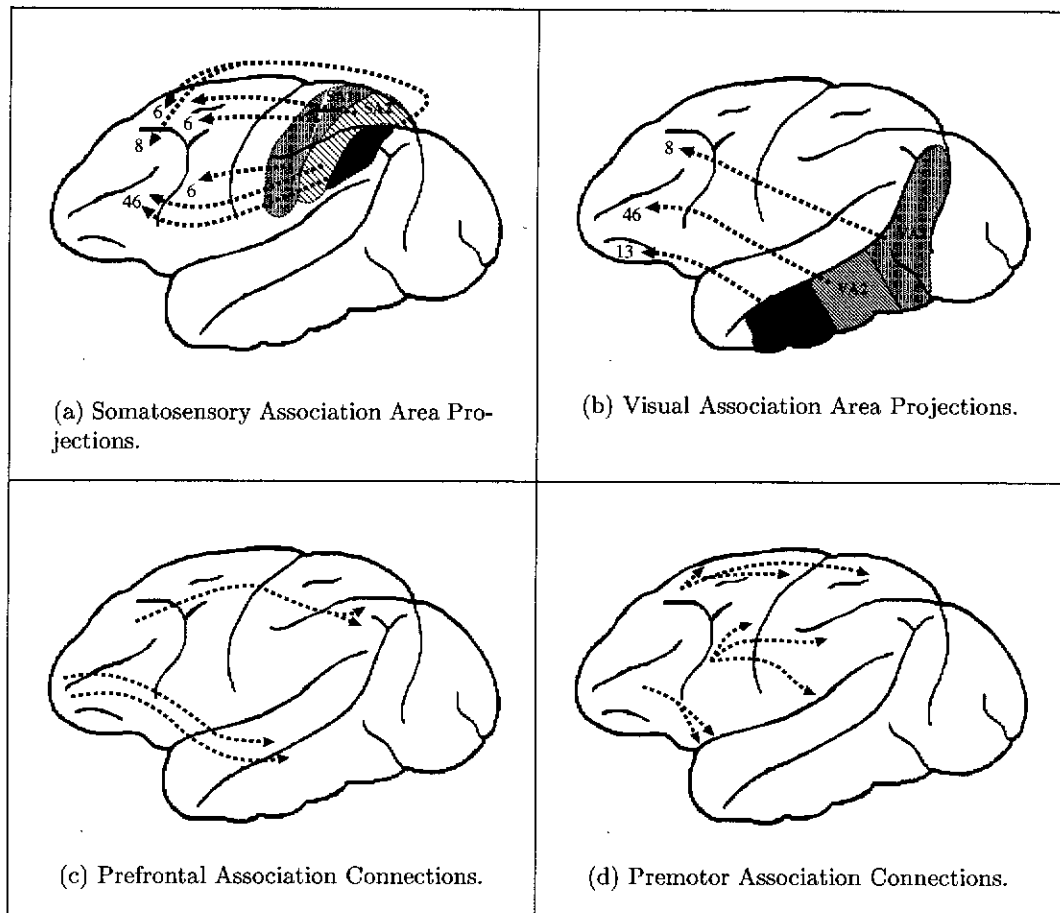


Figure A.6: Areas of Cortex of Rhesus Monkey.

The primary and secondary sensory areas [†] for the rhesus monkey are shown in Figure A.5 (a). The adjacent cortical regions which are downstream from these primary sensory regions are termed first-order parasensory areas, with those beyond the first-order parasensory areas being designated as second-order or third-order association areas (Jones & Powell 1970, Chavis & Pandya 1976). Figure A.5 (b) shows the association areas.

At the junctions of the modality-specific regions in parietal, occipital, and temporal cortices are other regions termed multi-modal areas (Jones & Powell 1970, Seltzer & Pandya 1976, Seltzer & Pandya 1978, Seltzer & Pandya 1980). These are shown in

[†]We are only concerned with the post-Rolandic sensory areas. Post-Rolandic refers to areas posterior to the Rolandic fissure, (i.e., the central sulcus). The post-Rolandic sensory areas are the somatosensory, auditory, and visual areas.

Figure A.5 (c).

Within a given modality, related regions are interconnected in a stepwise manner, beginning with the primary sensory area, and then progressing sequentially through the parasensory and multimodal areas, and finally projecting to frontal association and paralimbic regions (Pandya & Kuypers 1969, Jones & Powell 1970, Van Hoesen, Pandya & Butters 1972, Pandya & Seltzer 1982a).

These various areas are connected via long association connections, while the study of the intrinsic connections of cortical areas within one specific sensory modality (Kuypers, Szwachbart, Mishkin & Rosvold 1965, Jones & Powell 1969, Jones & Wise 1977, Tigges, Tigges & Perachio 1977, Vogt & Pandya 1978, Zeki 1978, Rockland & Pandya 1979, Rockland & Pandya 1981, Fitzpatrick & Imig 1980, Pandya & Seltzer 1982b, Kaas 1983, Van Essen & Maunsell 1983, Galaburda & Pandya 1983) show common organisational features relating to their origin and laminar terminations. †

All three post-Rolandic sensory systems have similar association areas. These are the first-, second- and third-order parasensory association areas of the somatosensory, visual and auditory systems, and each is discussed briefly in the following sections.

A.5.1 Auditory Association Areas

In both human and non-human primates, the cortical auditory system is found in the superior temporal region (STR). The primary auditory area is flanked by the auditory association area (AI) and by the second auditory area (AII).

Behavioural studies of the primary auditory area have suggested that it is involved in elementary auditory processing such as the analysis of frequency and amplitude (Walzl & Woolsey 1943, Merzenich & Brugge 1973), whereas the auditory association area (designated as area 22), is considered to be involved in more integrative functions such as auditory pattern recognition and sound localisation (Weiskrantz & Mishkin 1958,

† *Intrinsic* refers to the connections between adjacent architectonic areas within each lobe of the cerebral cortex, whereas *long association* connections refer to those which project to cortical areas that are distant from the originating region, i.e., connections between nonadjacent areas within a given lobe and those interrelating the various lobes of the cerebral cortex.

Leinonen, Hyvärinen & Sovijarvi 1980, Hyvärinen 1982). The primary and secondary auditory areas are shown in Figure A.5 (a), while the auditory associations are shown in Figure A.5 (b).

Studies of the intrinsic connections suggest that the entire superior temporal region is involved in the cortical auditory system (Fitzpatrick & Imig 1980, Galaburda & Pandya 1983). The common pattern of cell origin and termination for the intrinsic connections within the cortical layers is shown in Figure A.7.

On the basis of the long association connections, the auditory association areas are divided into three major sectors—the first, second, and third auditory association areas, AA1, AA2, and AA3 respectively (Chavis & Pandya 1976). Each of these regions has a distinct pattern of long association projections to the frontal lobe, parietotemporal region, and the paralimbic area. AA1 projects mainly to area 8, AA2 primarily to area 46, 9 and 10, and AA3 to areas 12, 13, 25 and 32, as shown in Figure A.5 (d). Most of the connections between the auditory association areas and the frontal, parietotemporal, and paralimbic regions are in fact known to be reciprocal.

A.5.2 Somatosensory Association Areas

The primary and secondary sensory areas, SI and SII, are located in the parietal lobe and are involved in basic processing of somatic sensation, for example, touch, heat and cold, texture (Randolph & Semmes 1974). The somatosensory association areas occupy most of the posterior parietal cortex (areas 5 and 7) and are thought to be involved in more complex and integrative somatosensory functions (Duffy & Burchfield 1971, Mountcastle, Lynch, Georgopoulos, Sakata & Acuna 1975, Sakata 1975, Robinson &

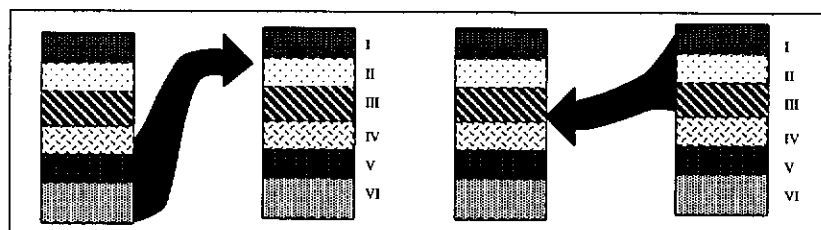


Figure A.7: Auditory Intrinsic Connections.

Goldberg 1978, Lynch 1980, Hyvärinen 1982). The somatosensory association areas are shown in Figure A.6 (a).

As with the auditory system, the somatosensory association areas may be grouped into three sectors (SA1, SA2, and SA3) on the basis of their long connections (Chavis & Pandya 1976). Again, the long association connections of these sectors are directed toward frontal, parietotemporal, and paralimbic regions

Area SI is connected with the motor cortex, MI, as well as the supplementary motor cortex, MII, whereas the association areas of the parietal lobe are connected with the premotor, supplementary motor, and prefrontal regions (Pandya & Kuypers 1969, Jones & Powell 1969, Jones & Powell 1970, Chavis & Pandya 1976, Barbas & Mesulam 1981, Petrides & Pandya 1983). The association area SA1 projects to rostral area 4, to premotor cortex area 6, and to the supplementary motor cortex. The second somatosensory association region, SA2, projects to area 6 as well as MII, and to area 46. Association region SA3 projects to area 6 and area 8. Unlike SA1 and SA2, area SA3 lacks a projection to MII.

On the basis of their cortical connectivity, the association areas appear to be involved in somewhat different somatosensory processes. The SA1 areas appear to be more directly related to somatosensory and motor systems by virtue of their connections with SI, SII and MII, as well as area 6. The SA2 areas do not have direct connections with SI, but are related to SII and MII, as well as the premotor region. Finally, the SA3 areas have a more extensive connectivity to other cortical association regions and appear to be the least directly connected with primary and secondary somatosensory areas (SI and SII). The SA3 areas project to the prefrontal as well as the premotor cortex, in addition to the paralimbic regions.

In summary then, areas SA1, SA2 and SA3 comprise a sequence of projection areas originating in SI, and ultimately leading to the premotor, supplementary motor and prefrontal regions (Pandya & Yeterian 1985). †

†Pandya & Yeterian actually propose two sub-divisions of this scheme which although similar, project to slightly different regions.

A.5.3 Visual Association Areas

The cortical visual system is located in the occipital lobe and the inferotemporal region. The primary visual area (VI—also known as striate cortex and designated as area 17), is involved in the initial visual processing within the cortex (Hubel & Wiesel 1968). The visual association areas include areas 18 and 19, as well as 20 and 21, and these association areas have been shown to be involved in further processing of visual information (Mishkin 1972, Gross, Desimone, Fleming & Gattass 1981, Ungerleider & Mishkin 1982, Mishkin, Ungerleider & Macko 1983).

Numerous studies have indicated a sequential flow of connections within the visually related areas of the occipital and inferotemporal cortices (Kuypers et al. 1965, Tigges, Spatz & Tigges 1973, Seltzer & Pandya 1976, Zeki 1978, Rockland & Pandya 1979, Rockland & Pandya 1981, Tigges, Tigges, Anschel, Cross, Letbetter & McBride 1981, Ungerleider & Mishkin 1982, Van Essen & Maunsell 1983, Rosene & Pandya 1983). The visual association areas surround the primary visual cortex (area 17), and receive topographically organised connections from the striate cortex. The primary visual area then projects to the surrounding prestriate area 18 in a topographic manner, which in turn connects to area 19 in a topographic manner.

Then, in turn, area 19 is connected with area 20 of the inferotemporal cortex, area 20 projects to rostral inferotemporal area 21, and area 21 sends projections to the proisocortex of the temporal pole. The visual areas indicate a systematic stepwise pattern of reciprocal connectivity in the same way as is found in the auditory and somatosensory systems.

Again, in a similar manner to the auditory and somatosensory association areas, the visual association regions may be divided into three distinct sectors on the basis of long cortical connections (Chavis & Pandya 1976). The first visual association area, VA1, consists of areas 18 and 19, the second, VA2, is comprised of area 20 (TE3), while the third consists of area 21 (TE2 and TE1). This is shown in Figure A.8 (a).

As shown in Figure A.6 (b), these visual association areas also send distinct projections to the frontal lobe, the parietotemporal regions, and to paralimbic areas, in a similar

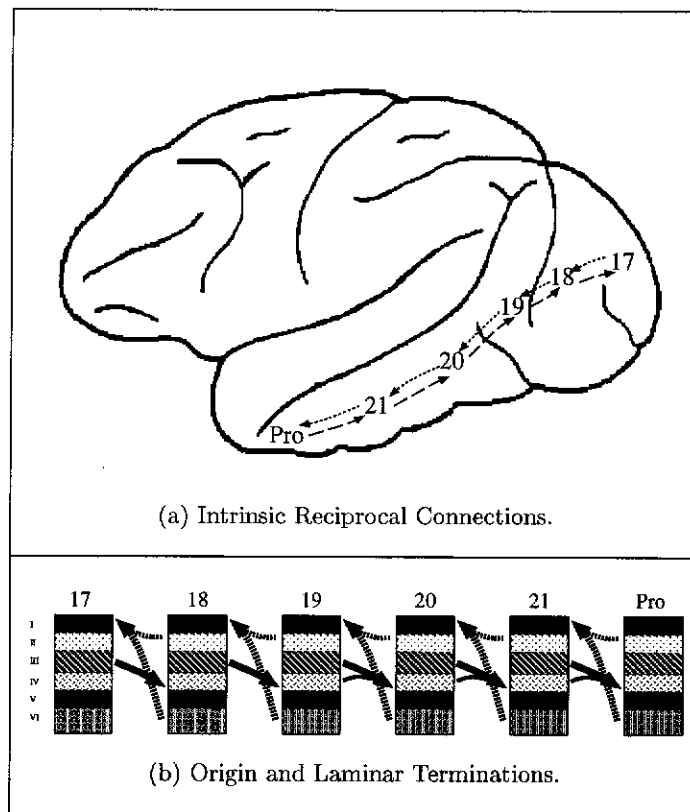


Figure A.8: Visual Intrinsic Connections.

manner to the auditory and somatosensory association areas. Area VA1 projects to the premotor area 8, VA2 projects to rostral area 8 and area 46, while the frontal area connection of VA3 is the same as VA2, but in addition it connects to area 11 on the orbital surface. Again the visual association areas VA1, VA2, and VA3 indicate a sequential distribution of projections to the frontal lobe (Kuypers et al. 1965, Chavis & Pandya 1976, Barbas & Mesulam 1981, Van Hoesen 1982, Rosene & Pandya 1983), each stage being related to frontal, parietotemporal, and paralimbic regions.

A.5.4 Frontal Association Areas

The frontal lobe cortex has traditionally been divided into three broad regions as shown in Figure A.9. The motor cortex (area 4 or MI), is implicated in somatic motor activity. †

†The supplementary motor area on the medial surface (MII), while thought to be part of the premotor region, is also involved in somatic (voluntary) motor activity.

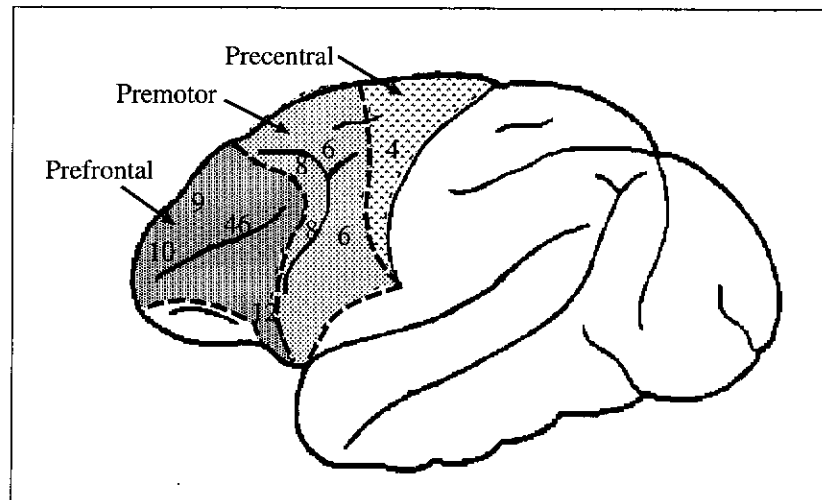


Figure A.9: Architectonic Subdivisions of the Frontal Areas.

The two other regions are the premotor region (areas 6 and 8) and the prefrontal cortex (areas 9, 46, 10 and 12 on the lateral surface, areas 13, 14, and 11 on the orbital surface, and areas 9, 25, 32 and 14 on the medial surface). The premotor and prefrontal regions are thought to be involved in complex integrative functions and are considered as frontal association areas (Woolsey 1958, Rosvold & Mishkin 1961, Mishkin 1964, Nauta 1971, Butters, Pandya, Stein & Rosen 1972, Petrides & Iversen 1976, Damasio 1979, Fuster 1980, Van Hoesen, Vogt, Pandya & McKenna 1980, Wiesendanger 1981, Brinkman & Porter 1983).

The premotor regions (including areas 6 and 8) are connected with the first-order sensory association regions via long association connections. Area 6 is topographically connected to SA1, and the premotor area 8 projects to AA1. There are limited connections from area 8 to the visual association areas (Jones & Powell 1970, Pandya & Vignolo 1971, Künzle & Akert 1977, Deacon, Rosenberg, Eckert & Shank 1982). The connections are shown in Figure A.6 (d).

The long projections of the prefrontal cortex are shown in Figure A.6 (c). Areas 46 and 9 connect to SA2 and SA3 in a topographic manner, areas 46 and 10 connect to AA2 (and adjacent STS), areas 46 and 10 project to VA2 (and adjacent STS), and areas 12, 13 and 25 connect to AA3 and VA3.

The overall connectivity suggests that the connections between association areas are reciprocal in nature. The long association connections of the prefrontal regions appear to be directed to those sensory association areas from which the frontal region receives afferent connections. The premotor region projects to the first-order association areas, the lateral prefrontal cortex connects with the second-order sensory association areas, and the orbital and medial prefrontal regions are related to third-order sensory association areas.

A.5.5 Multimodal Association Areas

The sensory association regions so far described are strictly modality-specific. At the junction of these unimodal association areas are other specialised cortical regions which receive input from more than one sensory modality. This extends to the frontal lobe and the paralimbic regions, which also contain regions which receive connections from more than one modality.

A.5.5.1 Post-Rolandic Multimodal Areas

Figure A.10 (a) shows the multimodal area POa, which receives input from SA1 (head, face and neck representations) as well as strong connections from the first-order visual association area VA1 (representing mainly the peripheral visual field) (Kuypers et al. 1965, Rockland & Pandya 1981, Ungerleider & Mishkin 1982, Van Essen & Maunsell 1983). It also receives projections from an area known to contain cortical vestibular representation (shown as VS on Figure A.10 (a)) (Fredrickson, Figge, Scheid & Kornhuber 1966, Büttner & Lang 1979). Given this relationship with somatosensory, visual, and vestibular areas, it is considered that area POa is a multimodal region involved in the integration of sensory information about head position with peripheral visual input. The POa area is also reciprocally connected with the frontal eye field region (Pandya, Dye & Butters 1971, Mesulam, Hoesen, Pandya & Geschwind 1977, Barbas & Mesulam 1981).

The multimodal area of the caudal IPL (area PG and area Opt) is shown in Fig-

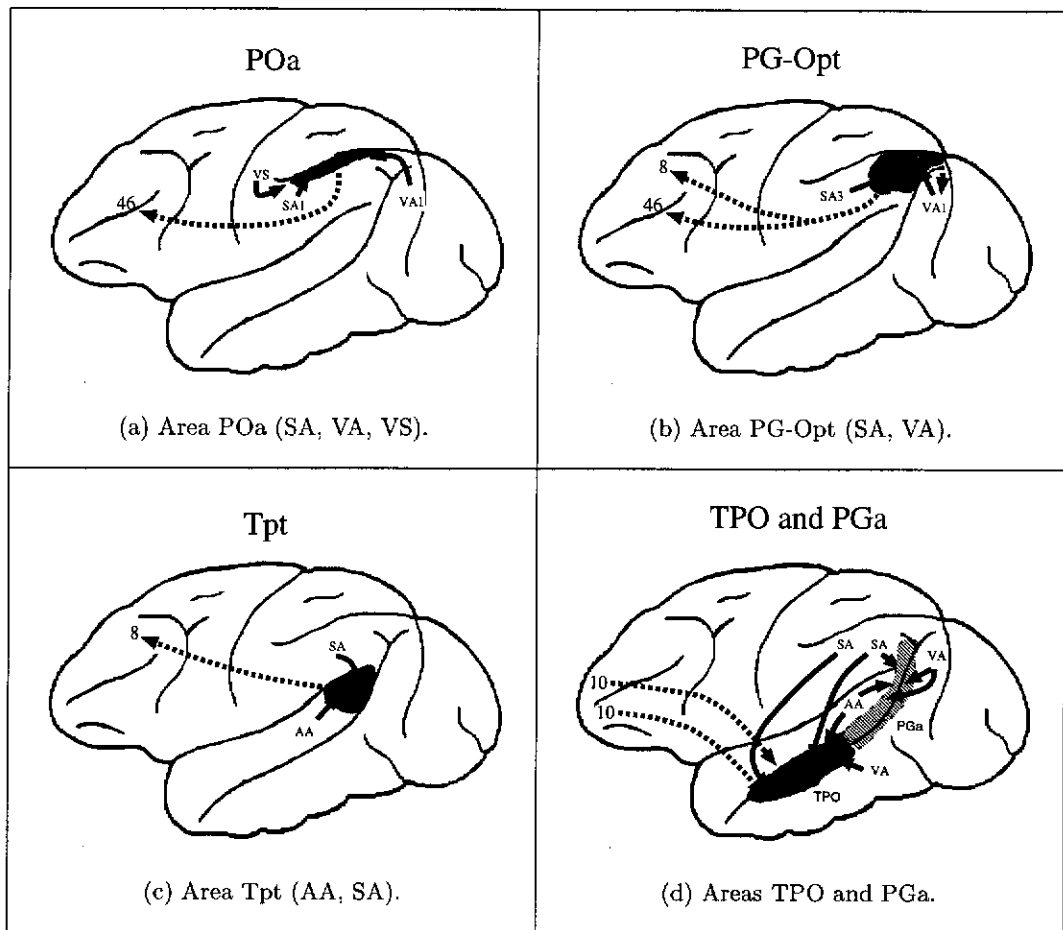


Figure A.10: Multimodal Areas and Connections.

ure A.10 (b). This area receives connections from both somatosensory association area SA3 and from the visual association area VA1 (Kuypers et al. 1965, Pandya & Seltzer 1982b, Ungerleider & Mishkin 1982), as well as paralimbic input (area 23 and the parahippocampal region). It is thus considered that this is another multimodal region which allows for the integration of higher order somatosensory, visual and limbic information (Mesulam 1983). By virtue of its connections, the caudal IPL region is presumed to be involved in visuospatial attention and orientation.

The area shown in Figure A.10 (c), area Tpt, is found at the parietotemporal junction, and it receives inputs from both the somatosensory association areas as well as the auditory association region (Pandya & Kuypers 1969, Jones & Powell 1970). It then projects strongly to the dorsal premotor region of the frontal lobe (Jones & Powell 1970, Chavis & Pandya 1976). This connectivity suggests a role in the integration of auditory

inputs with somatosensory information regarding face, head and neck, suggesting an audio-spatial attention and orientation role in orienting the head toward significant auditory sounds. †

Subregions of bimodal and trimodal convergence of sensory input are found in the cortex of the STS (Seltzer & Pandya 1978). This area, designated TPO, receives input from the auditory association areas AA1, AA2, AA3, somatosensory association areas SA3, and visual association areas VA2, VA3. This is shown in Figure A.10 (d). An adjacent area called PGa receives inputs from the sensory association areas, but is related predominantly to the somatosensory modality. Each sensory association area has a unimodal projection zone in the STS surrounding the multimodal areas TPO and PGa.

A.5.5.2 Frontal Lobe Multimodal Areas

Convergence of sensory inputs has also been shown to occur in the frontal lobe (Pandya & Kuypers 1969, Jones & Powell 1970, Chavis & Pandya 1976). The premotor region has a site of sensory convergence from the first-order sensory association areas—areas AA1 and SA1 send projections in a partly overlapping manner. Similarly, areas VA1 and SA1 have a partially overlapping projection. Finally, a trimodal convergence of all first-order sensory association areas is found in the central portion of the arcuate sulcus.

In the premotor region, single-unit recordings indicate polysensory characteristic neurons (Bignall & Imbert 1969, Nelson & Bignall 1973, Vaadia, Jr, Heinz & Benson 1983). As well, lesion-behavioural studies confirm the importance of this same region in visual, auditory, and somatosensory tasks, and its importance in behaviour which requires the integration of input from more than one modality (Welch & Stuteville 1958, Petrides & Iversen 1976, Van Hoesen et al. 1980). The ventral prefrontal cortex also has a zone of bimodal and trimodal convergence.

Unlike premotor multimodal areas, the prefrontal region receives input from second-

† Another auditory-somatosensory multimodal region may be involved in the orientation of the trunk towards sound sources.

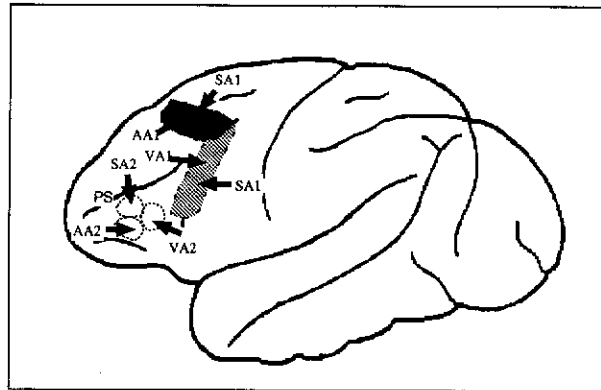


Figure A.11: Premotor and Prefrontal Multimodal Areas.

area sensory association areas. Thus ventral area 46 receives input from AA2, VA2 and SA2 in an overlapping fashion, as is shown in Figure A.11. This convergence zone is also implicated in behaviour contingent upon the synthesis of information from more than one modality (Passingham 1972, Passingham & Ettlinger 1972).

A.5.6 Paralimbic Association Areas

The limbic system, also known as the visceral or emotional brain, is concerned with underlying emotional expression. This includes the basic behaviours necessary for the preservation of the organism, such as feeding, fight and flight, as well as behaviours which are vital to the continuance of the species, such as mating, reproduction, and care of the young. In addition, it appears to play a central role in memory processing and in relating the organism to its environment, both in the intermediate sense and over time. The limbic system is much older phylogenetically than the cortex.

The limbic system receives continuous samples of all incoming sensory information, while its output, directly or indirectly, affects all endocrine, visceral motor, and somatic motor effectors. In Section 3.2.7 we briefly considered the possible role of the limbic system in ‘biasing’ the learning of the cortical regions according to the emotional context—value-based learning.

The paralimbic association areas are situated between the sensory association areas and the limbic regions. They have direct connections with limbic structures. All

of the cortical input to the paralimbic association areas is derived from the second- and third-order sensory association areas. They are also reciprocally connected to the prefrontal and orbitofrontal cortices, but have no direct connection to either parasensory association areas or primary sensorimotor regions.

The strong connectivity between paralimbic and limbic structures implies that they also have a role in motivational and emotional behaviour (e.g., Mesulam (1983)).

A.5.7 Summary of Sensory Association Areas

Pandya & Yeterian (1985) suggest that there are a number of general principles common to all three sensory modalities in regard to their long association connections.

The parasensory association areas in each modality can be divided into three major sectors. The first-order association areas send projections predominantly to the periauricular (premotor) regions, whereas second-order areas are most strongly connected with prearcuate (prefrontal) regions. Finally, the third-order visual and auditory association areas project preferentially to the orbitofrontal region, while the third-order somatosensory association areas project to the rostral prefrontal region.

Pandya & Yeterian (1985) suggest that speculation on the functional role of the association areas may be achieved, based on the organisational pattern of connections as well as behavioural data.

The first-order association areas project to the premotor region of the frontal lobe which in turn has reciprocal connections with the motor cortex. The evidence suggests that this circuitry is involved with the use of sensory information to influence motor behaviour. Lesions of this area produce a deficit in responding to contralateral sensory stimulation (Lilly 1958, Welch & Stuteville 1958, Fuster 1980, Wiesendanger 1981).

The overlapping nature of the cortical sensory input to premotor cortex implies a role for this area in behaviour which is contingent upon the ability to integrate information from different sensory modalities. Behavioural studies show that lesions in this area produce deficits in learning a motor task only when it involves making an intermodal association and not when it depends simply upon a unimodal sensory discrimination

(Petrides & Iversen 1976, Van Hoesen et al. 1980, Petrides 1982).

A somewhat different functional role is suggested for the prefrontal and orbitofrontal cortices. These receive projections from the second- and third-order parasensory areas as well as paralimbic regions, and are less directly connected with the motor cortex. This connectivity suggests that the prefrontal region allows the organism to include both complex sensory data and motivational and emotional factors in subsequent motor actions.

Frontal lobe damage in humans indicates alterations in affect, one of which is an apparent state of indifference (Hacaen 1964, Nauta 1971, Blumer & Benson 1975, Damasio 1979). Many of these patients tend to act impulsively and seem unable to carry out organised behaviour which requires a shift in strategy. These deficits may arise in part from an inability to relate behaviour to internal states resulting in a tendency to perseverate in an original strategy despite its being inappropriate in a different context. We examine the frontal lobes more carefully in the following section.

The differential post-Rolandic connectivity of the premotor and prefrontal regions suggests that these areas make different contributions to the generation of motor behaviour in the context of stimulus input. The premotor cortex has a close relationship with the motor cortex, and receives connections from first-order sensory association areas. This implies a more direct role for this region in the execution of a motor act. On the other hand, the prefrontal cortex receives more highly 'processed' sensory and limbic input. It is also much less directly related to the motor cortex, suggesting that it may serve in decision-making and sequencing processes in which external (parasensory) and internal (limbic) factors are integrated (Nauta 1971).

In summary, Pandya & Yeterian (1985, page 54) state:

"As sensory input arrives at the cerebral cortex, it first passes through successive stages of intra-modality elaboration, allowing progressively more complex discrimination of stimulus features. Subsequently, by a series of further interconnections, this already elaborated complex information is conveyed to multimodal areas for inter-sensory integration. In turn, the in-

formation is relayed to paralimbic and limbic structures for investment with emotional tone and for memory consolidation. Finally, it is the frontal lobe regions where both sensory and limbic influences are integrated in preparation for the organism to execute behaviour appropriate to environmental and internal factors.”

The view presented in this thesis is similar:

- vectors provided by external filters and internal recorders,
- linked temporal sequencing from recurrent loops on to self-organising maps,
- intra-modal and inter-modal associations,
- limbic areas provide emotional context and ‘gain-control’,
- frontal lobe areas provide hierarchical, complex, temporal repertoire based on recurrent linkages, as well as the external and internal vectors,
- motor cortex sequences provide dynamic behaviours,
- feedback from the environment provides reinforcement for learning.

In all, the neuroanatomical data presented in this section presents a picture that appears to be consistent with the ABC model.

A.6 Prefrontal Cortex

The results cited in this appendix draw heavily from a review article by Fuster (1985). Here we look more closely at the prefrontal cortex and especially its role in temporal integration. The prefrontal cortex is the associative cortex of the frontal lobe, and in the human, accounts for nearly one-third of the neocortex.

The prefrontal cortex has traditionally been thought of as playing a role in higher integrative functions. Its location, however, suggests a primarily motor role, and it

contains areas that are at least engaged in two forms of motility: eye movements and speech.

Further, Fuster (1985, page 152) suggest that “[in] the light of current knowledge, . . . it is appropriate to postulate a critical role of the prefrontal cortex in certain complex forms of motor behaviour or in certain aspects of motor control.” Fuster proposes that the principal and most characteristic function of the prefrontal cortex is the temporal organisation of behaviour. This is entirely consistent with the ABC model, and in this section we examine the evidence given by Fuster in support of this claim, and how this fits in with the ABC model.

The connectivity of the prefrontal cortex is high, with many internal and external streams converging and interacting. As well, it is the origin of a wide variety of effector mechanisms. The conclusion reached is that the prefrontal cortex is the substrate for neural integration at a very high level.

The prefrontal cortex receives projections from the lower brain stem and the limbic system (Nauta 1972), suggesting that the medial thalamus provides the prefrontal area with input related to the internal state and motivations of the organism. It also receives direct projections from the hypothalamus.

As stated in the previous section, the prefrontal cortex is connected to the associative areas (in occipital, temporal, and parietal regions), and not directly to the primary sensory areas (Jones & Powell 1970). It thus receives indirect sensory input from the visual, auditory, and somatosensory modalities. Some of the prefrontal input is then most likely the result of elaborate integrations in the three major sensory association domains. The prefrontal cortex may be viewed as a polysensory region, and therefore a region for cross-modal association. There is support for this view from electro-physiological evidence (Bignall & Imbert 1969, Mohler, Goldberg & Wurtz 1973, Schechter & Murphy 1975, Tanabe, Yarita, Iino, Ooshima & Takagi 1975, Benevento & Fallon 1975).

The evidence for a role in motor behaviour for the prefrontal cortex is plentiful. Almost all of the cortical and subcortical structures which send projections to the prefrontal cortex receive reciprocal connections from it. The exception is a group of subcortical nuclei

involved in motor functions, (e.g., the basal ganglia) which only receive efferent connections from the prefrontal cortex (Nauta 1964, Webster 1965, Kemp & Powell 1970). Other nuclei of the hypothalamus and the lower brain stem which participate in various aspects of movement also receive direct projections from the prefrontal cortex (De Vito & Smith 1964, Leichnetz & Astruc 1977). Prefrontal area 8, the frontal eye fields, sends projections to subcortical structures implicated in eye movements (Künzle, Akert & Wurtz 1976, Leichnetz, Spencer, Hardy & Astruc 1981).[†] Furthermore, stimulation of the prefrontal cortex has been shown to induce a variety of inhibitory phenomena in the motor and autonomic spheres (Fuster 1981).

However, the influence of the prefrontal cortex on the primary motor cortex is not direct, but is exerted, like those of the primary sensory cortex upon the prefrontal cortex, through intermediate regions including the basal ganglia, the lateral thalamus, and especially the premotor cortex (Fuster 1981).

Fuster contends that the connectivity of the prefrontal cortex is fully consistent with the view that the prefrontal cortex plays a role in the temporal organisation of behaviour. Its wide variety of afferent and efferent connections, including associative cortices and limbic structures, suggest a unique role in coordinated behavioural processing.

Single-cell observations made in the prefrontal cortex of performing monkeys support this proposition (Fuster 1985, page 158–159):

- a large proportion of cells react with increasing firing frequency to sensory stimuli that serve the animal as cues to the performance of its task (Fuster 1973, Niki 1974, Fuster et al. 1982) and appear to distinguish visual, proprioceptive, or kinaesthetic features,
- reactions of some prefrontal cells to visual stimuli (Kubota, Iwamoto & Suzuki 1974, Sakai 1974, Suzuki & Azuma 1977) and auditory stimuli (Kubota, Tonoike & Mikami 1980) is dependent on the behavioural implications of the stimuli (i.e., whether they call for a motor response),

[†]Mild stimulation of area 8 elicits movements of the eyes and the head (Robinson & Fuchs 1969, Wagman & Mehler 1972).

- some cells are attuned to the spatial relationship of stimuli and response, and, for example, are able to differentiate left from right, up from down,
- units are found which show changes of discharge in relation to various aspects of reinforcement following a task,
- cells are found whose activity is altered in a sustained manner during the delay period, the time interposed between a given event and the appropriate response of the animal.

Other evidence for the postulated role in the temporal organisation of behaviour comes from neuropsychology. Deficits resulting from prefrontal lesions in animals typically concern the performance of delay-tasks—behavioural tasks which require the execution of a motor act in accord with events in the recent past. These tests require the animal to remember sensory items (usually visual) for a prescribed periods of time. Lesioned animals have extreme difficulty in performing these delay tasks correctly.

Different prefrontal areas have been found critical for different delay tasks, suggesting specialisation of the areas with regard to the cues initiating the behaviour.

Fuster suggests that the delay-task deficit is fundamentally one of short-term memory, with an impairment in the animal's ability to connect events *across time*. Their ability to integrate information from temporally separate events in order to execute a motor action is seriously impaired. Lesioned animals also have difficulty in ordering items in a sequence—a deficit of the temporal organisation of behaviour (Pinto-Hamuy & Linck 1965).

Normal animals exhibit a suppression of internal and external interference which might disrupt the behavioural sequence at hand. Lesioned animals, on the other hand, appear to be abnormally subjected to distraction from extraneous sensory stimuli as well as from obsolete internal tendencies or memory traces (Mishkin 1964). For example, in go/no go tasks in which a stimulus requires a given action while another stimulus requires no action at all, the animal is unable to suppress the inappropriate response. Further, in reversal tasks in which a reward associated stimuli and a neutral stimuli (which are presented simultaneously) are reversed, the animal is unable to suppress the

old association.

Prefrontal lesions in humans indicate difficulties in organising behaviour, such as a lack of spontaneity, poor concentration, and the absence of proper planning. One early explanation advanced by Brickner (1934) suggested that the prefrontal syndrome was due to a deficit of integrative synthesis of thinking and behaviour. †

It has often been surmised that the prefrontal cortex is involved in goal-directed synthesis of behavioural structures, and especially in the impairment of language. Humans with prefrontal lesions, (as with monkeys), show a deficit in the performance of self-ordering tasks—tasks that require the serial organisation of items in time (Milner 1971, Petrides & Milner 1982). These tasks impose a considerable demand on organising strategies and short term memory.

Frontal-lobe lesions in humans lead to impairment of speech and language, which according to Lashley (1951) are the highest forms of temporal structuring activity. Impairments include Broca's aphasia (Geschwind 1970), a profound disorder of articulation and grammar that results from lesions of the left inferior frontal gyrus, and "central motor aphasia" (Goldstein 1948), † a more subtle speech disorder produced by lesions of the more anterior prefrontal regions sparing Broca's area. Central motor aphasia results in a diminished spontaneity of speech and the general impoverishment of verbal expression. Patients tend to speak in short and simple sentences with little or no use of dependent clauses—a diminution of the so-called recursiveness of language (Chomsky 1975b).

Language impairment is more prominent following left frontal lesions. Right frontal lesions tends to affect fluency in nonverbal tasks such as the production or self-ordering of abstract designs (Jones-Gotman & Milner 1977, Petrides & Milner 1982).

As a general statement, Fuster concludes:

†Other theories include that of Luria (1966) who suggested that the difficulty patients have in planning their actions and in anticipating their consequences results from a disorder of the formation of preliminary synthesis of action, and Teuber (1966) who hypothesised a critical role of the prefrontal cortex in the intimate relationship between perception and movement.

†Also called "frontal dynamic aphasia" by Luria (1970).

... all the disorders of language produced by prefrontal damage can be characterised as syntagmatic disorders, that is, disorders of the capacity to form linguistic structures. Whereas posterior prefrontal lesions (Broca) impair the organisation of elementary morphemes, more anterior lesions impair the organisation of more complex language. With quantitative, not qualitative, differences, all prefrontal lesions seem to compromise a (syntagmatic) function of sequential ordering of sounds, words, and sentences in the making of semantic structures, in which meaning is both the goal and the organising principle.

Computer tomography research shows that there is increased metabolic activity in the prefrontal cortex during speech production, as well as activity in the premotor cortex and the mouth-larynx motor cortex (Ingvar & Schwartz 1974, Larsen, Skinhoj & Lassen 1978). However, in the course of inner speech, patches of activation are seen only in prefrontal areas and the supplementary motor area (SMA) (Larsen et al. 1978, Lassen & Larsen 1980).

Changes in personality after damage to the frontal lobes have been reported for well over a century, and no description of what has come to be called the frontal lobe syndrome would be complete without reference to the case of Phineas Gage. The following summary which has appeared in many places is quoted by Kimble (1963) :

Phineas P. Gage, an 'efficient and capable' foreman, was injured on September 13, 1848, when a tamping iron was blown through the frontal region of his brain. He suffered the following change in his personality according to the physician, J.M. Harlow, who attended him. "He is fitful, irreverent, indulging at times in the grossest profanity (which was not previously his custom), manifesting but little deference to his fellows, impatient of restraint or advice when it conflicts with his desires, at times pertinaciously obstinate yet capricious and vacillating, devising many plans for future operation which no sooner are arranged than they are abandoned in turn for others appearing more feasible. His mind was radically changed so that his friends and acquaintances said that he was no longer Gage."

In summary, Fuster (1985, page 169) suggests that “because of the large variety of its afferent connections, the prefrontal cortex, unlike the primary sensory and motor cortex, and unlike the posterior associative cortex, is in the position to integrate vastly diverse kinds of information on both the external and internal milieu of the organism.” He further contends that “the critical dimension is time and that the prefrontal cortex mediates the formation of temporally extended structures of behaviour on the basis of temporally discontinuous items of information, whether they come from the inside or outside the organism.”

The prefrontal cortex evidently take a role within the sensory and motor components of behaviour—prefrontal unit-discharge changes are seen to both succeed sensory stimuli and precede motor actions. The intervention of the prefrontal cortex allows each event in a sequence, whether sensory or motor, to be put in a proper temporal perspective. Thus, on the strength of the evidence, Fuster postulates that the main function of the prefrontal cortex is the temporal organisation of behaviour.

A.7 Motor Areas

Most of the brain deals with motor function. A large part of the cerebral cortex, as well as a number of subcortical structures such as the cerebellum, large portions of the basal ganglia, the brain stem and spinal cord, are all concerned with self-initiated and stimulus-elicited movements.

The primary motor component of the cerebral cortex is the motor cortex, a strip of cortex located just in front of the central gyrus. The motor cortex provides major outputs to the spinal cord and brain stem, and is heavily interconnected with other cortical areas and with the major subcortical structures, the cerebellum and the basal ganglia.

Changes in motor cortical cell activity precede the development of the motor output and relate quantitatively to its intensity (Evarts 1981, Georgopoulos 1990). One major characteristic of the motor output, be it movement or isometric force, is its direction in space—cells in motor cortex are broadly tuned to the direction of a reaching movement

and not the endpoint (Georgopoulos, Kalaska, Caminiti & Massey 1982, Schwartz 1993, Georgopoulos, Kalaska & Caminiti 1985). This means that a cell's activity is highest for a movement in a particular direction (the cell's preferred direction) and decreases progressively with movements farther away from this direction (Georgopoulos 1995). The broad directional tuning of each cell indicates that it participates in movements of various directions, and that a movement in a particular direction will involve activation of a whole population of cells.

In line with these findings, Georgopoulos (1995) suggests a vectorial neural code for the direction of reaching by the neuronal ensemble, a model entirely in keeping with the ABC model. We propose that the output of the motor component is a vector which represents a neuronal code for the manipulation of various muscles to perform a learned action in line with these findings.

A.8 Language Areas

Language seems to be such a fundamental and uniquely human ability that a discussion of language must be an important topic in any model of human cognition. We discuss the mechanisms of language within the ABC model in Section 5.1. In this section we look at some of the neuroanatomical and neuropsychological evidence that relates to language.

The overwhelming view of language held by cognitive scientists is that it is a formal, computational process in which each sentence must be *parsed* to separate the sentence into its syntactic components—the noun, verb, and so on. Caplan (1995), in a review article, discusses some of the research directed at finding such a computational 'module'. Information-processing models of language may be expressed as flow diagrams which indicate a sequence of operations performed by different components in performing a language-related task. One such hypothesised processor is one that is required for syntactic structure parsing. For example, a model developed by Frazier (1987b, 1987c, 1987a) consists of a number of independent modules: one that builds phrase structure, one that assigns co-reference, one that assigns thematic roles and predication, and one

that assigns reference to pronouns and other referential items.

Caplan looks at lesions in various areas of the cortex with a view to finding suitable sites for these 'computational' structures. An examination of the correlations between language impairments and lesion sites suggests that the association cortex in the region of the sylvian fissure is responsible for language processing (for a review, see Caplan 1987).

This region includes Broca's area (areas 45 and 44), the association cortex in the opercular area of the pre- and postcentral gyri, areas 39 and 40, Wernicke's area (41 and 42), and possibly a portion of the adjacent second temporal gyrus. It is claimed that language processing occurs in these areas in the left hemisphere in up to 98% of right-handed individuals (Milner 1974). Further, in up to a third of individuals who are ambidextrous or left-handed, the corresponding areas in the right hemisphere are used (Goodglass & Quadfasel 1954).

Broca's area is a motor area thought to be responsible for speech, being located adjacent to the area of the motor cortex that controls the muscles of the face, tongue, jaw, and throat.

Correlational and path analysis of lesion sites in patients with language impairments suggest that most of language processing is cortically based (Metter, Riege, Hanson, Jackson, Kempler & VanLancker 1988), but the supplementary motor area and several sub-cortical structures such as the caudate, putamen, and parts of the thalamus are also thought to be involved.

Some aphasic patients appear to suffer selective impairments in their ability to construct syntactic structures in sentence production. One such group—the so-called Broca's aphasics who have an expressive language disturbance known as agrammatism—have a tendency to simplify syntactic structures and to omit function words (e.g., articles, auxiliary verbs, etc.) as well as grammatical morphemes (e.g., agreement markers, plural markers, etc.) from their speech. Patients with these symptoms tend to have lesions that include Broca's area. However, the lesions in patients with the larger syndrome of Broca's aphasia often extend well beyond this region.

Specific impairments caused by ablation in Broca's area often occur in the production of the sounds of speech, even though language ability remains normal (Bloom & Lazerson 1988, page 283).

In general, lesions producing a 'parsing' deficit showed a tendency to involve frontal and insular cortex, but exhibit considerable variability in size and location.

Other patients who have difficulty constructing syntactic structures in comprehension tasks and/or in using these structures to determine the proposition expressed in sentences have been described by many researchers (for a review, see Caplan 1992). Many of these patients are agrammatic Broca's aphasics, leading several researchers to suggest that Broca's area is the site of syntactic processing (Mesulam 1990, Domasio & Domasio 1992). However, for a number of reasons, the hypothesis that Broca's area is the sole area responsible for parsing—or for a particular set of parsing operations—can only be said to receive modest support from the existing data on sentence-comprehension impairments in agrammatic patients (Caplan 1995).

Patients with another kind of aphasia (Wernicke's) produce well-formed speech sounds in essentially correct grammatical sequences, but make meaningless sounds such as (Buckingham Jr. & Kertesz 1974):

“I think that there's an awful lot of mung, but I think I've a lot of net and tunded in a little wheat duhvayden”

This type of aphasia is produced by damage to the upper, posterior part of the left temporal lobe (area 22), also known as Wernicke's area.

Broca's area and Wernicke's area are connected by a collection of nerve fibres called the *arcuate fasciculus*. We postulate that this linkage via the arcuate fasciculus provides the means not only for language use but for thinking (or self-talk), as we discussed in Section 3.3.5.

The quest to isolate the 'parsing module' has been unsuccessful, with varied and confusing results. The ABC model takes a completely different approach to language, suggesting that it is learned temporal sequences over multiple hierarchical recurrent

networks.

Much more work is required here, especially in relating the various aphasias to functional areas in the model. However, this will need to be left to future research.

A.9 Binding of Modalities

Damasio (1989) examines the problem of the integration of the various perceptual inputs, and associated motor interactions—the *binding problem*. The perceptual inputs to the brain are widely separated in different regions, and need to be linked to other internal information and to motor outputs. An answer to the binding problem requires an indication of just how the brain achieves integration of the various modalities and actions.

The traditional view suggests that the different sensory inputs meet in so-called multimodal cortices, where the most detailed and integrated representation of reality is achieved. It further suggests that perception is predicated on a unidirectional flow of information which provides a gradual refinement of signal extraction along a cascade aimed towards the integrative cortices in anterior temporal and anterior frontal regions. This view seems intuitively reasonable as “anatomical projections do radiate from primary sensory cortices toward structures in the hippocampus and prefrontal cortices via a multi-stage sequence, . . . and the further away neurons are from primary sensory cortices, the larger their receptive fields become, and the less unimodal their responses are”. (Damasio 1989)

However, Damasio suggests that this traditional view must be rejected for several reasons:

- No appropriate region has been found which receives projections from all sensory input regions. Even for the most frequently mentioned candidate, the anterior frontal cortices, the sensory and motor streams remain segregated in different regions. In other words, there seems to be no structural foundation to support the intuition that temporal and spatial integration occurs at a single site.

- There is neuroanatomical evidence of multi-stage, reciprocating feedback projections at each stage of the chain of forward cortical projections. In fact the cortex appears to be just as rich in feedback as in feedforward projections.
- Bilateral destruction of temporal and frontal integrative cortices does not preclude “perception of reality as a coherent multimodal experience”, nor “disable memory for any form of past integrated experience” nor “interfere with all levels and types of memory”.
- However, damage to sensory association cortices can result in a reduction in the quality of some aspects of perception within the modality served by those cortices, as well as and recognition and recall. For example, lesions in visual association cortex can result in impaired perception of shape, or colour, or spatial placement of the physical components of a stimulus.
- The fact that a patient may lose the ability to perceive colour, and yet be able to perceive shape and motion normally suggests separate maps for these perceptual modes.
- Damage within some sectors of modal association cortices may disturb recall and recognition of stimuli presented through that modality, even when basic perceptual processing is not compromised. For example, an inability to recognise faces even though the perception of faces is not impaired. Further, the patient may be unable to discriminate familiar from unfamiliar faces at a covert level.

Damasio proposes a new view on binding which we feel is still too rooted in a representationalist viewpoint to be accepted. However, we share his rejection of the traditional view of binding. Just as was the case in Chapter 4 where we rejected the traditional view of spatial fusion in vision, here again we reject the need for an iconic fusion of multiple perceptual inputs into some overall ‘representation of reality’.

The ABC model achieves coordinated actions through temporally associating sensory inputs with appropriate learned behaviour. The vector inputs from the various senses may be regarded more as contextual inputs such that certain combinations of inputs (including internal sources) provides a context for particular actions. There is no need

to bring together all of the inputs into a combined representation. Indeed, some actions, such as orienting to certain sounds in the environment, may require inputs from only a reduced set of the inputs (here auditory and proprioception). Further, sub-sets of the inputs may be extracted for particular tasks, such as separating the foveal and peripheral visual inputs.

The model works by linking inputs to learned outputs via self-organisation, and the manner in which the various components may be linked is very flexible. Association layers provide temporal cohesion, as does the learning of integrated temporal sequences. The process of concatenating vectors allows the vectors to provide each other with a (mutual) context.

A.10 Discussion In Relation To the ABC Model

The neurobiological evidence presented in this appendix is indeed supportive of the ABC model. No finding contradicts the model, and the general structures and connectivities are mirrored in the model.

Obviously, the difficulty in determining some semblance order in a structure as fluid and detailed as the brain is not to be underestimated, and much further work in neuroanatomy and neuropsychology is required to tease out the actual structures.

Appendix B

Temporal Learning Experiments in Detail

In this appendix we present the details of a number of temporal learning experiments that were performed in addition to those described in Chapter 2.

The complete ABC model is described in Chapter 3, and one of the major components of this model is temporal learning. It is important, therefore, to examine the temporal learning capabilities of the proposed model. In order to achieve this understanding, we first look at a number of smaller-scale experiments that were conducted with several reduced models. These simulations were conducted in order to test the temporal learning characteristics of the full architecture. In the following sections we look at these temporal simulations.

B.1 Temporal Learning—Finite State Automata

The architecture of the full temporal model was explained in Section 2.3. However, before we look at the full temporal model, let us first look at a slightly different model. This model is similar to that used by Jordan (described briefly in Section 2.2.1).

The aim here is to build a network that is able to learn a Finite-State Automata (FSA) language. FSAs are very simple computing machines, capable of parsing finite-state

grammars—the lowest on Chomsky’s hierarchy of grammars (Winograd 1983). The importance of this experiment is to show that the simple recurrent ABC temporal network is capable of learning a FSA in a *general* structure, as opposed to the more specific networks as described by Pollack (1991) and others.

A Finite-State Transition Network (FSTN) can be regarded as a description of a language. It can also be interpreted as a specification of an FSA to recognise (or in some cases, generate) elements of the language. A transition network consists of a set of *states* connected by *arcs*. The transitions between states are directional, and occur when a new input is to be parsed (or generated). Usually there is an *initial* state and a *final* state.

B.1.1 Stopwatch Example

As a simple example, consider the case of a stopwatch. The stopwatch transition network is shown in Figure B.1.

The training data is a set of state transitions. For example, if the stopwatch is in the *rest* state and the *sw1* button is pressed, the watch starts the clock mechanism and moves to state *count*. A subsequent press of the *sw1* button will stop the mechanism

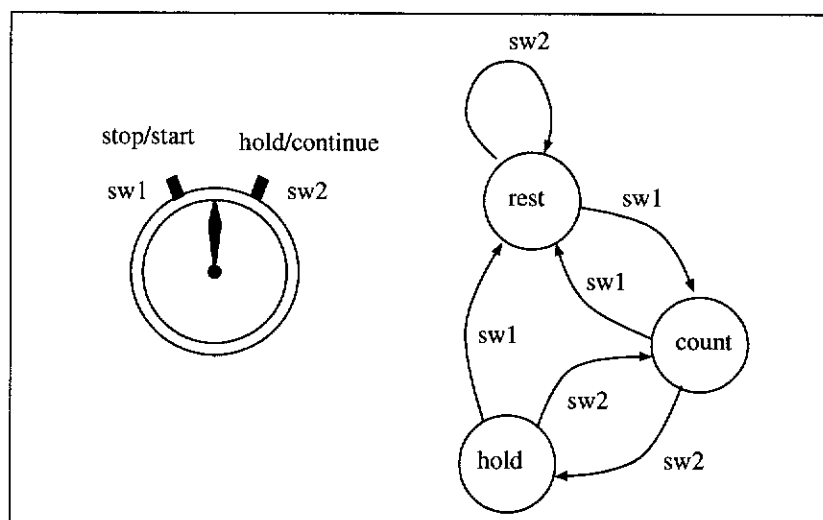


Figure B.1: Stopwatch FSTN.

Current State	Input	Next State
rest	sw1	count
rest	sw2	rest
count	sw1	rest
count	sw2	hold
hold	sw1	rest
hold	sw2	count

Table B.1: State Transitions for Stopwatch FSA.

and return the watch to the *rest* state. Pressing the *sw2* button in the *rest* state has no effect, returning to the same state, whereas pressing this button while in the *count* state will cause the hand to stop while counting continues—the *hold* state.

As can be seen, the *sw1* input starts and stops the overall counting mechanism while the *sw2* input starts and stops the hold action. The full set of state transitions and associated inputs is shown in Table B.1.

Consider the network arrangement as shown in Figure B.2. For convenience, let us refer to this overall structure as an FSAMap.

To train the system, pairs of input states and button press vectors (which have been

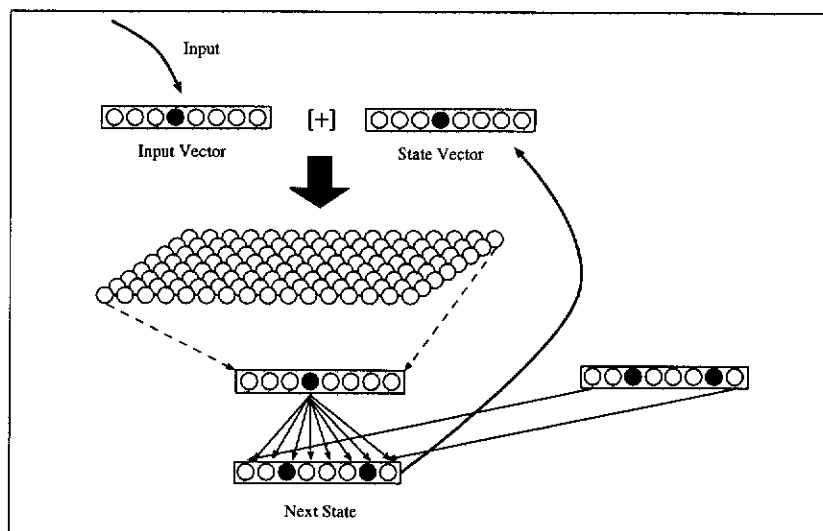


Figure B.2: Finite State Automata—FSAMap.

State	Vector
rest	0.0 1.0 1.0 0.0 1.0 0.0 1.0
count	0.0 1.0 1.0 1.0 0.0 1.0 0.0
hold	1.0 0.0 0.0 1.0 0.0 1.0 1.0
Input	Vector
sw1	1.0 0.0 1.0
sw2	0.0 1.0 0.0

Table B.2: State and Input Vectors.

assigned to random vectors containing ‘bits’ of either 1.0 or 0.0) were presented as inputs, and the system trained to generate the vector appropriate to the next state as its output. The input vector values in this case are shown in Table B.2.

Once the system was trained, another program was used to test the learned FSTN. An initial state is provided as input, and then a series of button-press inputs. The system is tracked to check that it follows the required state transitions. For example, an input sequence of rest sw2 sw1 sw2 sw2 sw1 sw2 should follow the state transition sequence rest rest count hold count rest rest.

The internal process followed by the FSAMap is simple: the vector corresponding to the initial state is concatenated with the vector of the first button-press input. The

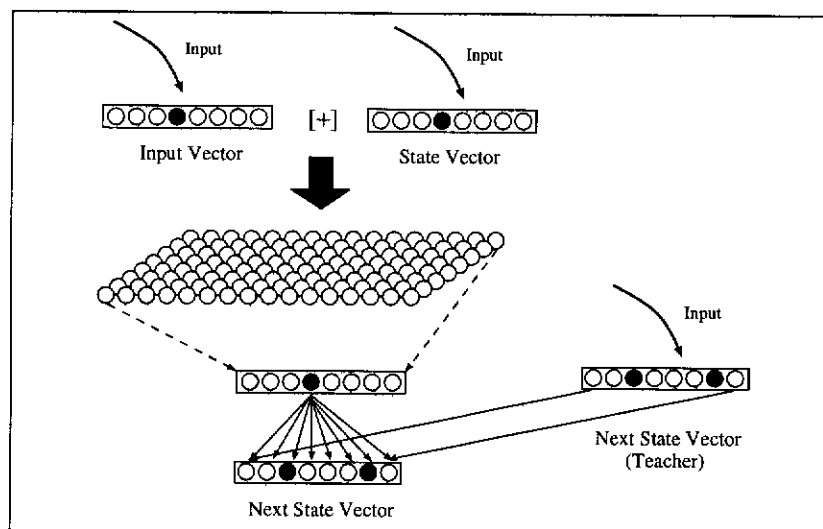


Figure B.3: Training the FSAMap.

combined vector is mapped to the SOM, and a winning node determined. In training mode, the SOM weights are adjusted using the Kohonen algorithm. The outputs of the surface nodes are calculated and combined with the next-state weights to calculate the next state vector. Again, in training mode, the weights are updated via Hebbian learning—using the *self-supervised* algorithm.

Training the FSAMap is shown in Figure B.3. The three input vectors (*state*, *input*, and *next-state*) are used to calculate a set of SOM and motor action weights. Running with a previously trained FSAMap is illustrated in Figure B.4. Here the input vector is combined with the current state vector, and the combined vector passed through the system to calculate a vector for the next state. This state vector is then (recurrently) used as the state vector for the next input.

The components of the FSAMap are illustrated in Figure B.5.

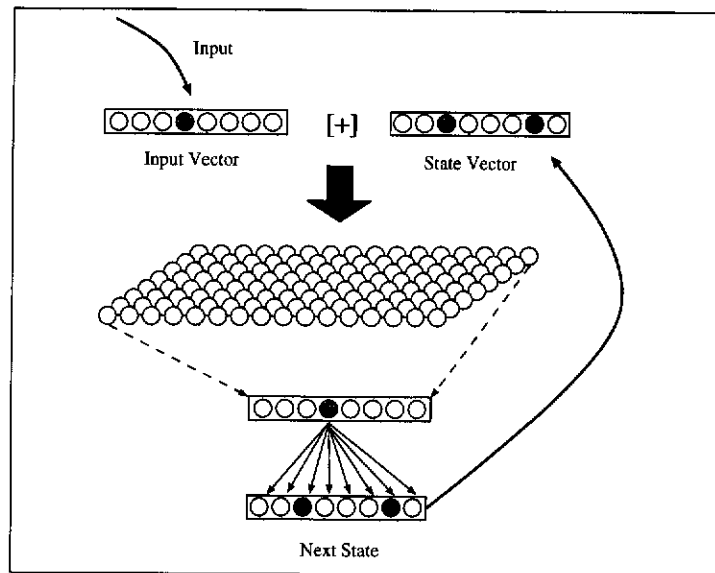


Figure B.4: Running the FSAMap.

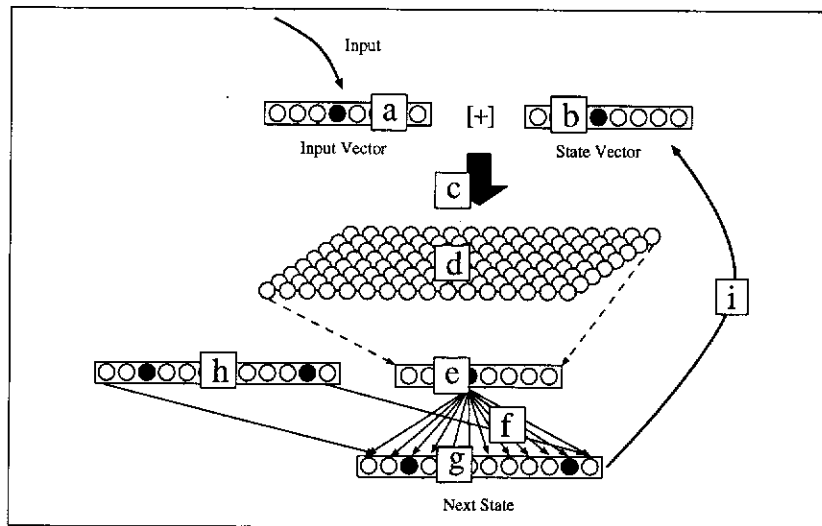


Figure B.5: FSAMap Dissected.

These components are:

- a *input* (button-press) vector
- b *current-state* vector
- c mapping to the SOM surface
- d SOM surface neurons
- e *output* vector (vectorised copy of SOM surface output values)
- f *next-state* weights updated via Hebbian learning
- g *next-state* vector
- h teacher copy of *next-state* vector for Hebbian learning
- i recurrent link

Once trained, the FSAMap was able to perform the task of emulating the stop-watch FSTN to 100% accuracy; given an initial state and an input, the system could correctly determine the next state. This state vector, when recurrently returned and combined with another input, will determine the next state, and so on indefinitely. The FSAMap system was able to learn the given FSA language and behave accordingly.¹⁶

It is important to recognise that no real distinction needs be made between the training and recognition stages of the method. The process is continuous, with recognition and

learning occurring simultaneously, albeit at a lower learning rate for later epochs. The software simulation however, was done in these two stages for convenience.

The first program performs both recognition and learning for a certain number of epochs until the system has learned. For the FSAMap, this was usually set to 200 epochs. A separate program was then used to read in the learned weights for both the SOM and the motor action weights, and to test the learned system against a series of test inputs. The second program performs no learning. †

B.1.2 Number Example

A second FSA problem presents a more difficult example—the learning of a language to determine if a series of digits is a valid number sequence; in integer, real or scientific notation. The FSTN is shown in Figure B.6. This example is both more complicated, and exhibits the effective use of an alternate (error) path to return to a particular state.

Input characters are ['-', '+', '0'-'9', '.', 'e', 'x'] where 'x' is used to indicate the end of a sequence.

The training triplets are similar to the stopwatch example; for example, an input of `start 6 st-1` indicates that if the FSAMap is in state `start` and receives an input of `6`, then it should move to state `st-1`.

Once trained, and given an initial state (`start`), subsequent input sequences such as:

```
6 . 0 x
- 3 5 0 e 1 x
+ . 2 x
+ 2 1 7 0 8 3 . 5 2 1 6 e - 2 1 x
```

are to be parsed correctly, and the state returned to the `start` state, while sequences such as:

†This separation between learning and testing is the case for most of the temporal learning experiments in this chapter. Even though the initial program performs both learning and recognition, it may be more convenient to refer to it as the *learning* program, with the second termed the *testing* program.

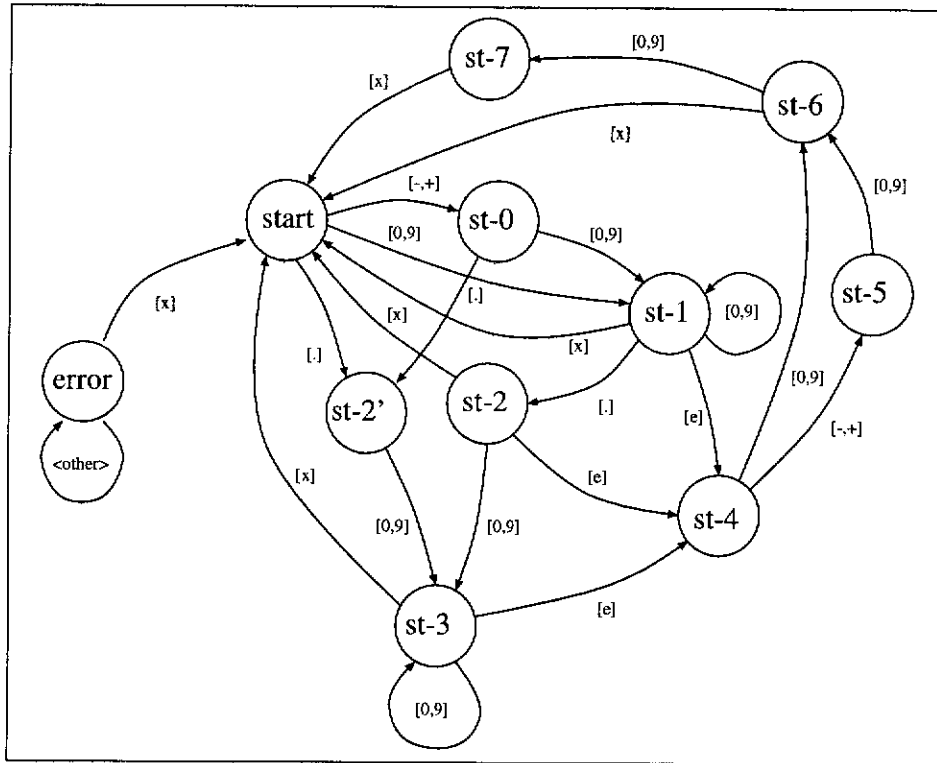


Figure B.6: Number FSA.

```

1 . 2 e 1 0 2 x
9 . . 3 x
. x
    
```

should go to the **error** state upon detecting an error, before returning to the **start** state (following the input of an **x**) in order to accept another number.

An FSAMap network was trained to perform the number parse for binary inputs (digits of just **0** and **1** as well as all other characters). This network was successful in parsing a binary number (in integer, real and scientific forms).¹⁷

An initial attempt to train the network for decimal numbers (as opposed to binary numbers as above) ran into problems of separation of the winning nodes. This separation of winning nodes for different inputs is essential if the correct transitions are to occur. The problem was that the same node was being selected for two state transitions, which obviously cannot be allowed as it leads to incorrect behaviour of the network.[†]

[†]For the other temporal experiments discussed later, this selection of the same winning node for

In order to overcome this problem, the *exclusion* modification (see Section 2.3.6) was made to the learning program to prevent the same node from being selected twice in the one epoch. This modification enabled the revised system to better separate the winning nodes, thus ensuring more ‘deterministic’ behaviour.

With the exclusion modification, the [0-9] Number FSA worked exactly as required. There were 165 state transitions (15 inputs for 11 states), and so represents a reasonably large FSA problem.¹⁸

Note that a return to state **start** from state **st-1** indicates that an integer has been parsed, a return to **start** from states **st-2** or **st-3** signifies a real number, and a return to **start** from states **st-6** or **st-7** signifies a number in scientific notation.

The *error exit* state transitions provide a means of correcting for invalid inputs. If the FSAMap is in a certain state, and an input is received for which there is no explicit state transition in the FSTN, then a transition will be made to the **error** state. Once in this state, all inputs except for an **x** will be absorbed, and the **error** state maintained. An entry of **x** from the **error** state will result in a transition to the **start** state for the entry of the inputs for the next number.

The error handling is built by explicit state transitions. For example, the inputs of -, + and . are not allowed from state **st-2**. The explicit training state transitions for state **st-2** are shown in Table B.3 (a), whereas those for the **error** state are shown in Table B.3 (b).

B.1.3 Bidirectional Link Example

The FSTN shown in Figure B.7 was discussed by Sharkey & Sharkey (1996) in relation to the perceived severe limitations of the SRN architecture to handle embedded sequences. We discuss embedded sequences in Section B.2.6, but here we will simply use this example as another test of the explicit ‘rule’ learning architecture of the FSAMap. The example demonstrates that the network is able to learn the loop FSA exactly.

two inputs also had to be avoided as much as possible, but it was not as important as in the FSA case where the program behaviour must be exactly deterministic.

Current State	Input	Next State	Current State	Input	Next State
st-2	-	error	error	-	error
st-2	+	error	error	+	error
st-2	0	st3	error	0	error
st-2	1	st3	error	1	error
st-2	2	st3	error	2	error
st-2	3	st3	error	3	error
st-2	4	st3	error	4	error
st-2	5	st3	error	5	error
st-2	6	st3	error	6	error
st-2	7	st3	error	7	error
st-2	8	st3	error	8	error
st-2	9	st3	error	9	error
st-2	.	error	error	.	error
st-2	e	st4	error	e	error
st-2	x	start	error	x	start

(a) **st-2** Transitions.
(b) **error** Transitions.

Table B.3: Error State Transitions.

Training involved 48 explicit state transitions (including error recovery) derived from the 8 states and 6 possible inputs at each state. The FSAMap was trained for the standard FSAMap regime of 200 iterations of the complete set of state transitions.

Following training, the weights obtained were saved and used in a further testing program. The system was tested with 10,000 generated strings, (of which 15 were unique),

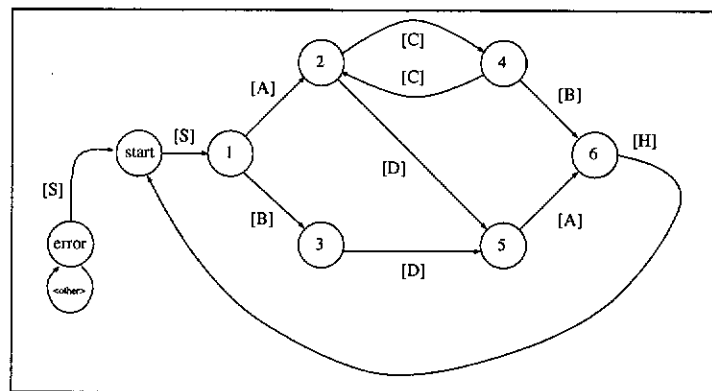


Figure B.7: Bidirectional Link FSA.

involving some 53371 state transitions. The maximum string length (resulting in part from the **C** character loop between states **2** and **4**) was 21 characters, including the initial **S** and final **H**.¹⁹

The FSAMap was able to learn the FSTN and provide 100% correct transitions in the test runs.

B.1.4 Reber Example

Another FSA cited in the literature is that shown in Figure B.8. This example is cited by Cleeremans in his book on implicit learning (Cleeremans 1993), which deals not only with implicit learning by artificial neural systems, but also discusses the implicit learning of ‘rules’ by humans. Again we use this network as an example testbed for the FSAMap architecture.

In this case there are 56 explicit state transitions (including error recovery) involving 8 states and 7 possible inputs at each.

Once trained for the standard epoch regime, the resultant weights were tested with 10,000 generated test strings, (of which 651 were unique), involving some 79,895 state transitions. The maximum string length was 31 characters, resulting from the loops at states **1** and **2**, and the loop between states **2**, **4** and **3**.²⁰

Again the FSAMap was able to correctly determine the next state in 100% of the test strings.

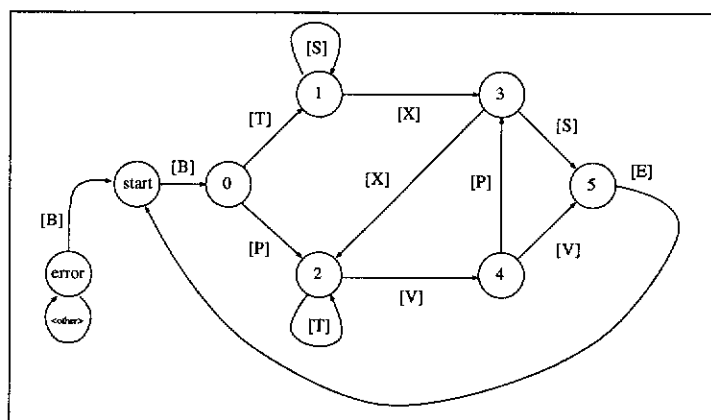


Figure B.8: Reber FSA.

A second test run of 2,368 unique strings again produced no errors and was 100% correct in all transitions.

A test of the error recovery was undertaken, but this was not rigorous. Ten error strings were tested, and the system was able to successfully recover to the start of the next string.

B.1.5 Actions

The FSAMap architecture may be extended to initiate some action or a set of actions whenever a state transition occurs. The action could be to turn some device on or off, or perhaps to output some information.

A set of “action vectors” can be trained directly from the SOM surface (equivalent to the next-state vector arrangement), or even as another layer of nodes *following* the next-state vector, as shown in Figure B.9.

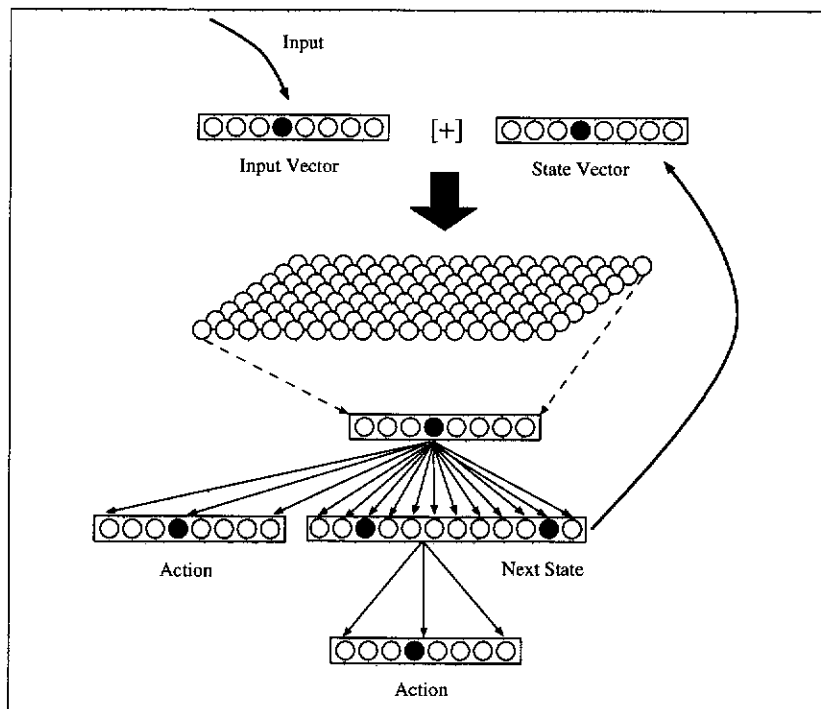


Figure B.9: Finite State Automata with Action Vector.

B.1.6 Soft Programming

As an aside, the learning of state transitions and associated actions opens up the possibility of an alternative to the ‘rigidness’ of programming. Programming on serial digital computers is fraught with problems. If just one bit of information is in error due to a malfunction or any other reason, then the computer system may come to a halt and need to be restarted. Also, the process of programming is very difficult and error prone.

The above ‘learned state transition’ technique might provide a new computational methodology that is less prone to disastrous and *all-or-nothing* errors. It also may provide a method of getting a computer to behave according to a set of learned ‘rules’ without the need for explicit and complicated programming—a form of learned ‘programming’ without actual coding.

This point is illustrated further in the discussion of hierarchies of FSA units—joining a number of FSAMaps together—as discussed in Section B.3. Further discussion is left to future research.

B.1.7 Discussion—FSAMap

The previous sections examined the case of a recurrent neural network being trained to learn the ‘rules’ of an FSA. Although the rules (the state transitions) were learned, they were supplied as data in an *explicit* form.

In a sense, this is essentially the same as the concept of an *expert system*. In both cases the ‘rules’ are supplied in explicit form as data, and it is the task of the system to learn these rules and to take appropriate actions when given examples at run time.

Others have examined the learning of FSAs by recurrent neural nets. Most are of two types: those that take a *hybrid* approach; for example Das et al. (1992) who use a recurrent neural network connected to an external stack memory through a common error function: and those that try to learn a *minimal* network of nodes to perform the FSA; for example, Porat & Feldman (1991).²¹

FSAs have been studied by a number of authors in relation to language learning (see for example Mozer & Bachrach 1991, Rager 1992).

The biological feasibility of both the hybrid and minimal network methods is questionable, but we make no biological claims for the FSAMap method either.

The FSAMap mechanism performed extremely well, and proved to be robust yet flexible. The method is both simple and relatively efficient, and with the use of appropriate hardware implementations (instead of the relatively slow, serial computational method used here) could scale up to much larger problems.

B.2 Full Temporal Model—Experimental Results

We leave consideration of the FSAMap for the moment, and return our attention to the two-SOM layer network shown in Figure 2.9. In this section, a number of experimental simulations are performed with this larger model. In some of the experiments, we duplicate work done in previous papers on temporal sequence learning in order to compare this network with others. We also obtain some new results which extend previous work on sequence learning and generation.

B.2.1 Temporal XOR

The first experiment performed on this network is the perennial XOR, this time a temporal version of the XOR problem. The XOR problem is a standard for both historical and practical reasons: historical in that it was the inability of the perceptron to learn the spatial XOR problem that held back research into neural nets for many years, and practical, in that the XOR problem requires a non-linear partitioning and hence is a good test of the ability of a network to learn non-linear functions. The XOR truth table is:

Input	Output
0 0	0
0 1	1
1 0	1
1 1	0

Learning the temporal XOR was first performed by Elman (1990, page 185) who obtained results indicating that his SRN could learn the XOR table. Our experiment is essentially the same as Elman's: a continuous string of 3,000 bits is presented to the network, where each sequence of three bits satisfies the XOR rule; that is, 1000 blocks of three bits from the set 000, 011, 101, 110 were selected at random and assembled as the input stream. As stated by Elman, the bit value of the first and second bits, the fourth and fifth bits, and so on, will be random. However, every third bit will be predictable based on the previous two bits.

The network was trained over the standard number of epochs. [†] Both of the mapping surfaces were set at 3 by 3 nodes.

The network was able to learn the XOR sequence such that it could correctly determine the required bit (based on the previous two) in 100% of cases in the later training epochs. Of course, a determination was attempted by the program for every pair of inputs in the continuous stream, not just for the first two of every block of XOR examples. For those pairings that were not governed by the XOR rule, the selection was made on the accumulated statistics of the random combinations, and obviously the determination could be incorrect. However, for every block based on the XOR rule, a determination based on the first two inputs corresponded with the third on every occasion. ²²

The winning nodes of the SOM surfaces are shown in Figure B.10. This figure indicates that, once trained, an input of 1 would always select node (0, 0) on the initial SOM map, whereas an input of 0 would always select node (0,2). Similarly, the sequence 01 would always select node (0, 0) on the second SOM surface, and so on.

Following training, the weights associated with the two SOM surfaces and the motor map were saved. As a further test of the learning, a separate program was developed which read in these weights, accepted stream inputs and determined appropriate outputs in the same manner as the learning program. The only essential difference was that this test program did not modify the weights, but simply used them to determine a result. A further five sets of 2000 randomly selected bits were generated and used as test streams.

[†]In most of the temporal experiments, a standard epoch regime was used of 100 epochs for the first map, 200 for the second, and 300 for the motor weights.

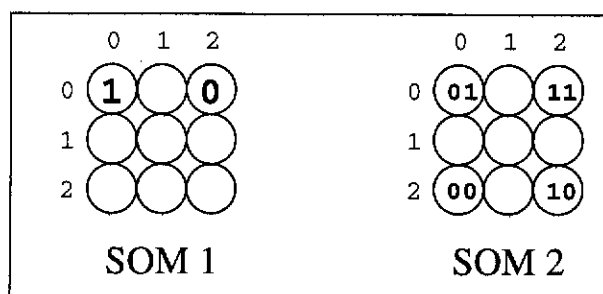


Figure B.10: XOR Winning Nodes.

The behaviour on each of these test sets was interesting. On each set, the first four iterations produced one or two incorrect determinations which violated the XOR rule. For example, given two successive inputs of 1 the program might produce an output of 1. However, following this 'warming-up' period, the results produced in the next 1996 iterations were in all cases 100% correct. It seemed that the network needed a few iterations to get 'on track'.

Of the 10,000 tests over the five sets, 9 errors were produced, all within the first four iterations.

In another test of the XOR sequence learning task, we examined the case of only providing one set of the four (3-bit) XOR patterns as training examples. This presents a problem in that the system does not just attempt to learn after each set of two inputs to calculate a third, but rather attempts to learn after every input. Thus the sequence 101000011110 will contain sets of three bits that violate the XOR scheme—for example, the second, third and fourth bits (010) violate the XOR conditions—so do the third, fourth and fifth (100). If the 12 possible combinations of 3 contiguous bits in the above example are examined, and remembering that the input stream is wrapped around, the following table is obtained:

Bits	Count	Validity as XOR
000	2	yes
001	1	no
011	1	yes
010	2	no
101	2	yes
100	1	no
110	1	yes
111	2	no

In the case of the above sequence, there are more examples of incorrect XOR strings (010 and 111) in the total contiguous input, than there are correct examples! Thus, the XOR scheme should not be able to be learned from this input stream alone, as was found to be the case.

In fact, of the 24 possible arrangements of the 4 XOR strings to form a continuous 12-bit input stream, none are free of this problem. All have combinations which violate the XOR pattern (8 have 8 correct and 4 incorrect combinations, while 16 have 6 correct and 6 incorrect combinations, as in the example above).

To get the system to learn the XOR pattern with just one set of bit strings it was necessary to introduce a separating character (in this case a full stop), to break up the input stream. The input supplied to the network was thus

1 1 0 . 1 0 1 . 0 1 1 . 0 0 0 .

When the network was trained with this pattern over the standard epoch regime, the XOR scheme was also learned: whenever the system received two numerical bit characters in sequence (following a full stop), it correctly predicted the appropriate next bit. ²³

B.2.2 Elman's Consonant/Vowel Sequences

Another of the experiments performed by Elman (1990, page 187) is reproduced in this section. In this case, the input is in the form of a random sequence of characters (**b**, **d**, **g**), with each of these consonants being followed by a particular vowel. Every **b** is followed by exactly one **a** character; every **d** is followed by two **i** characters, and each **g**

is followed by three **u** characters. An input string in this scheme might be, for example, **diibaguuubadiidiiguuu ...**

The first training set consisted of 1000 randomly selected consonants and the appropriate vowels. Following training, the network always selected a **b** whenever one of the consonant characters (b,d,g) was required because of a slight statistical bias in favour of **b**,[†] but then once it knew the actual consonant character, either **b**, **d**, or **g**, it always produced the next 'vowel' sequence exactly as required. That is, if the actual character on the input was a **b**, then the network consistently determined **a** as the next (one) character, then selected **b** as the next; if the actual input character was **d** then the next three selections were always **i i b**; and similarly, whenever a **g** occurred as input, the next selections were always **u u u b**. Thus the system had exactly learned the sequences as required.²⁴

The accuracy of learning was maintained over a number of runs. Whereas Elman's system could only give a statistical indication of the learning of the sequence, we were able to get 100% accuracy in the vowels at the end of the training epochs.[‡]

Keeping in mind the fact that the sets of characters are presented to the learning network in a contiguous stream, the sequence may be drawn as the FSA shown in Figure B.11.

Examination of this figure reveals that a number of paths end up in the same state; for example, paths **id**, **ad** and **ud** all end at state **st-d**, paths **ib**, **db** and **ub** all end at state **st-b**, paths **ag**, **ig** and **ug** all end at state **st-g**, and paths **ba**, **ii** and **uu** all end at state **start**. This structure is shown if we examine the winning nodes of the two SOM surfaces, shown in Figure B.12.

Looking at the winning nodes for the second SOM, which records the transitions across input characters, we see that the network has generalised to the extent that most paths leading to the same state share the same winning node. For example, the character

[†]The counts were: b 342, d 333, g 325.

[‡]This is, of course, predicated on the selection of reasonable input vector codes. The values used, as described previously, were chosen to have a reasonable separation in the space 2^n using the Hamming metric. The issue of vector selection was not examined in detail.

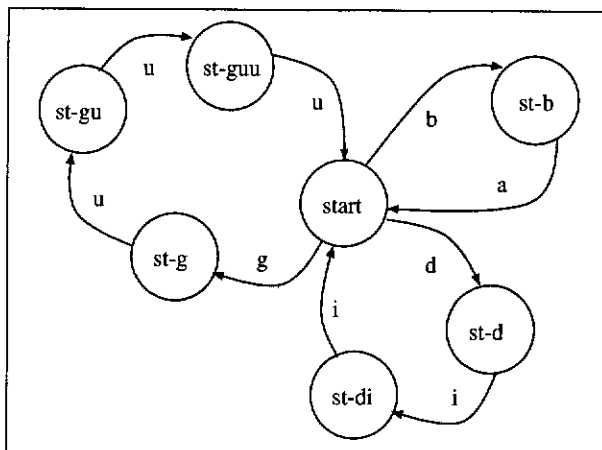


Figure B.11: Consonant/Vowel FSA.

sequences **ag**, **ig** and **ug** all share the node (0,0). These sequences will all expect a **u** to be next. The same can be said about the sequences which end at state **st-b** and share node (0,2)—they all expect an **a** character next.

The sequences which end at the **start** state (**ba**, **ii** and **uu**) are not able to generalise as the subsequent character in the sequence is not predictable. Other states which are not able to be generalised are **st-gu** (reached following a sequence **gu**), **st-guu** (sequence **uu**), and **st-di** (sequence **di**). As expected, these sequences are found on separate winning nodes in Figure B.12.

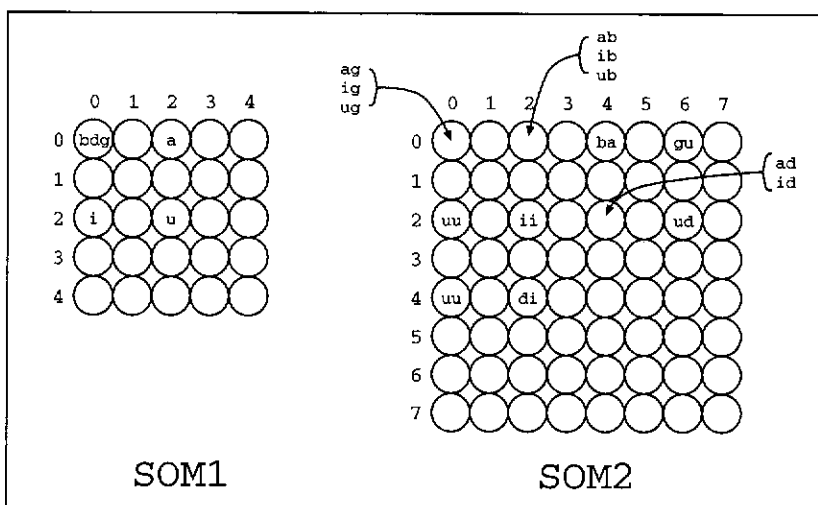


Figure B.12: Consonant/Vowel Winning Nodes.

Notice also that the system has learned to differentiate the two **uu** transitions. This effect is discussed more fully in Section B.2.5, where we examine the extent to which this differentiation may be continued.

One may be tempted to conclude that the limited size of the second SOM array may have contributed by forcing the generalisation. However, this is not the case. Even with a larger surface of nodes the generalisation still occurs.

Notice also that a form of generalisation has also occurred on the first SOM. This is in some part due to the similar vector values for these characters as used by Elman (see following table), and duplicated here.

Char	Vector Used
b	1 0 1 0 0 1
d	1 0 1 1 0 1
g	1 0 1 0 1 1
a	0 1 0 0 1 1
i	0 1 0 1 0 1
u	0 1 0 1 1 1

To test this observation, a new set of random vectors was chosen (6-bit, Hamming distance of at least 3 between each) and another run of the learning process, with a slightly modified set of input strings, was conducted. In this new set of 999 consonants, exactly 333 of each consonant were randomly interspersed.²⁵

The results were essentially the same as before. Although the same degree of generalisation did not occur (due to the more separated vectors and hence less generalisation on SOM1), the second SOM did exhibit some form of generalisation with the winning nodes for the sequence transitions with a common successor grouped close together on the map.

B.2.3 Bidirectional Link Revisited

We examined this FSA previously in Section B.1.3 where we explicitly learned the transition rules using an FSAMap network. In this section we seek to learn the ‘rules’ implicitly by learning directly from example input strings.

The point about this network is that it can generate a variable number of embedded

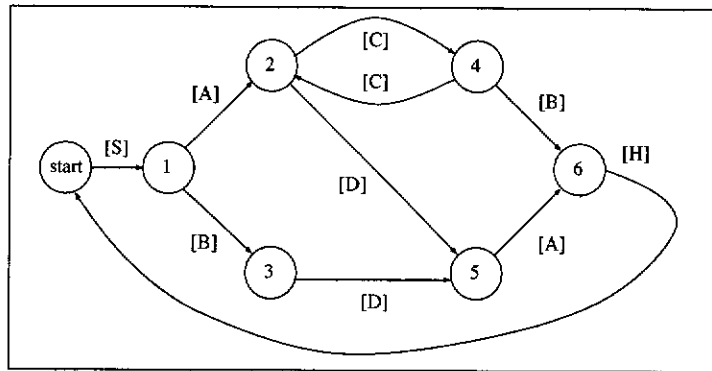


Figure B.13: Bidirectional Link FSA Revisited.

C characters. Various studies attempting to learn this (or similar) sequences with SRNs have shown that an SRN is unable to generalise the loop between states **2** and **4**, and instead learns specific path information (Sharkey & Sharkey 1996, Cleeremans et al. 1989, Cleeremans 1993). Testing the system on the same number of (or less) **C** characters than are found in the training data gives correct results, but any attempt to test on a longer sequence of **C** characters results in failure, with the system ‘losing its way’. For example, if the network is trained on up to say 3 **C** characters in sequence, and then tested on data with 4 **C** characters in a row, it will just retrace the 3 **C**s path and then become lost. The SRN network typically records specific paths as in Figure B.14 (a), whereas what is required is for the system to be able to generalise to the loop transition diagram as shown in Figure B.14 (b).

Sharkey & Sharkey (1996) found that an SRN could not learn the FSTN of Figure B.13

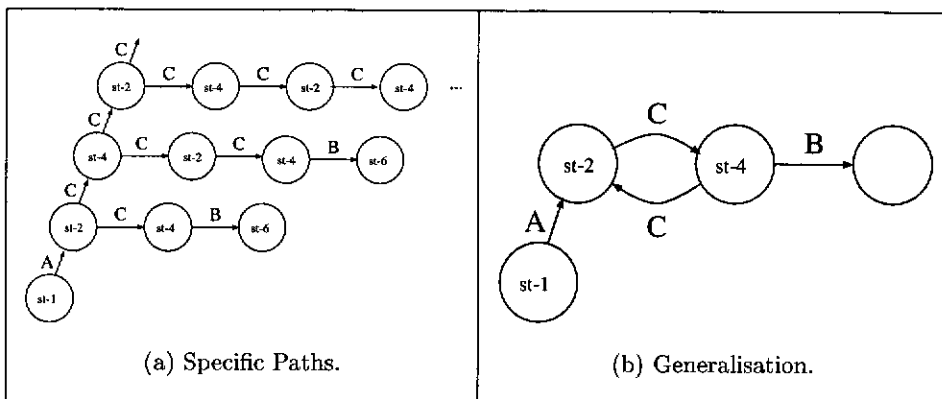


Figure B.14: Specific Path Information versus Generalisation.

at all. They concluded that this was a problem that could not be avoided with SRNs when trying to learn a bidirectional link. If the system does not generalise but records specific paths only, then no matter how many embedded **C**s the network is trained on, the final **C** must predict a **B** only. Thus a test using more embedded **C**s than the training set will fail.

Sharkey & Sharkey could only get an SRN network to learn the embedded loop FSTN by explicit restriction and training of the hidden units. By constructing a solution in this way, they were able to show that the network architecture was capable of learning the required generalised transitions, but the backpropagation algorithm could not do so.

There are a number of ways in which we might seek to train a network to recognise a sequence that includes ambiguous transitions; for example, the transition from state **st-1** on Figure B.13 can proceed to either **st-2** or **st-3**. We would expect that a trained network would indicate that both of these transitions are allowed, and that once in state **st-1** only the characters **A** or **B** are allowed. All other characters should be disallowed. But how do you get a network to indicate that it has learned this multiple-choice transition? Consider the following possibilities.

local representation: Use a supervised mode and train the network to compute a weighted response for each possible next character. This is a form of *local* representation, and for the bidirectional link example, the output vector learned following an **HS** transition would be like that shown in Figure B.15 (a), indicating that outputs **A** or **B** are allowed next.

distributed representation: Pre-select the input and output vectors to uniquely indicate which character or characters are allowed. This is a form of *distributed* representation, and the vector used to indicate a choice of **A** or **B** could be the one shown in Figure B.15 (b). The code used is externally defined and interpreted.

self-supervised: Allow arbitrary vectors for the input and output, and use a self-supervised mode to train the network to reproduce the vector code of the *most likely* next character. Here the codes are not arbitrary, but reflect the input vector

values, but the problem is that only one of the multiple choices can be indicated. This case is indicated in Figure B.15 (c).

context: As with the self-supervised case but allow some other ‘context’ component to bias the choice of which path to indicate. For example, in the case of the transitions in Figure B.13 we could associate the transition **SA** with context **left** and the transition **SB** with context **right**. The addition of the context component reduces the choice of the next character to a single character—the output following an **HS** transition, given a **left** context, is an **A**, as indicated in Figure B.15 (d).

Most of the literature uses one of the first two solutions, especially the first. However both have problems. The local representation method is often the easiest to understand and implement, but is not very realistic for human and animal cognition. Local representations are arbitrary and artificial. The distributed representation method is cumbersome and again rather artificial as the various output combinations need to be determined and a representation agreed upon in advance.

On the other hand, the self-supervision method is appealing in that it only requires the network to learn the actual vectors that appear on the input lines and to reproduce them on the output lines. However, the difficulty is that only one vector at a time may be reproduced on the output lines. The extension to the self-supervised method is to

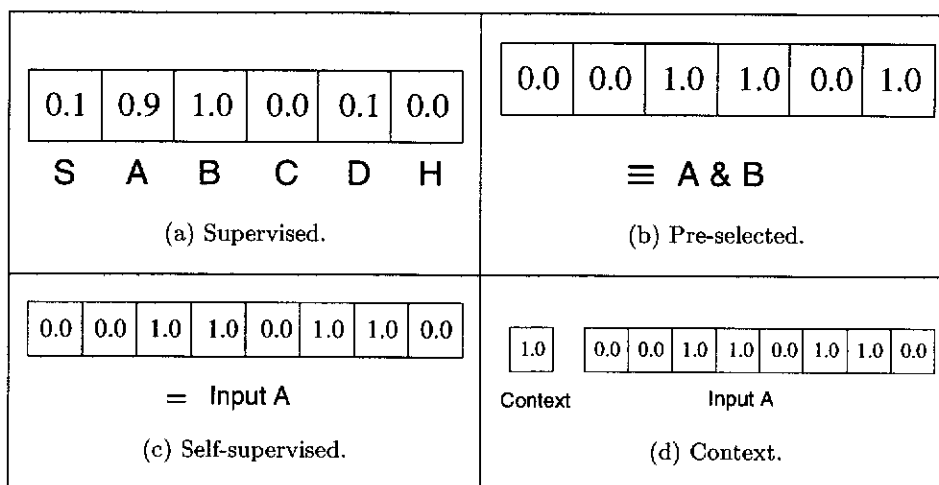


Figure B.15: Multiple-choice Transition Learning.

use some other variable, an indicator of some 'context', to select between the possible alternatives. This alternative is the most biologically realistic, and is considered in future sections.

B.2.3.1 Self-organised Motor Action—SOMA

Let us examine the learning of the bidirectional link FSTN using the self-supervised method. For convenience, we will refer to this architecture as SOMA (Self-organised Motor Action). Using the standard temporal learning architecture described previously, the network is trained to reproduce the vector code of the most likely next character. Thus the same set of vectors applies to both the inputs and outputs.

The training data consisted of 100 random sentences of which 7 were unique, as is shown in the following table. †

S	A	C	B	H					
S	A	C	C	C	B	H			
S	A	C	C	C	C	C	C	B	H
S	A	D	A	H					
S	A	C	C	D	A	H			
S	A	C	C	C	D	A	H		
S	B	D	A	H					

The results following training are shown in the following table. The table gives the selected next character for each of the possible pairs of input characters. For example, given a sequence of **HS** as input, the SOMA network determined that an **A** is the most likely next character (based on the actual input data).

†Note that the set does not include the case of 5 **C** characters in sequence.

Character Input 1	Character Input 2	Selected Next Char	Options
A	C	C	B,C
A	D	A	A
A	H	S	S
B	D	A	A
B	H	S	S
C	B	H	H
C	C	C	C,B,D
C	D	A	A
D	A	H	H
H	S	A	A,B
S	A	D	C,D
S	B	D	D

As can be seen from the table, all of the selected next characters are correct, being the most statistically likely of the possible options based on the input data. ²⁶

The same form of generalisation as in the Elman Consonant/Vowel example was observed on the SOM surfaces, shown in Figure B.16.

By following the trace of the winning nodes on the SOM2 surface, and comparing these with the input data sequences, it is possible to show that the system actually learned a slightly different FSTN to the original, as shown in Figure B.17.

Sharkey & Sharkey ran through a test of their constructed solution with a sequence of some 80,000 embedded Cs—a slight overkill to say the least. The current model was tested with sequences up to SA{19 Cs}BH and SA{20 Cs}DAH, with success in each case.

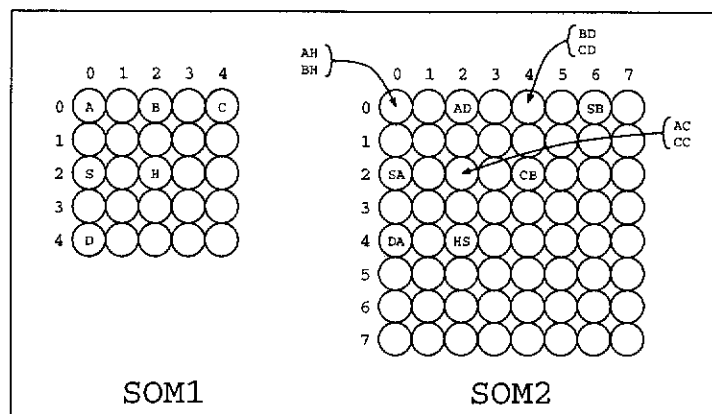


Figure B.16: Generalisation of SOM Nodes.

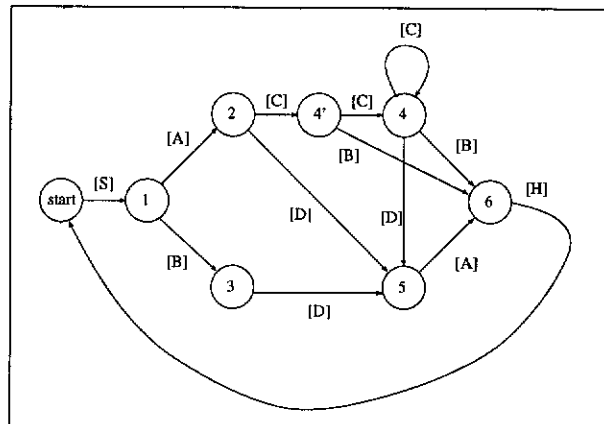


Figure B.17: Actual FSTN Learned By System.

The sequence of winning nodes showed that the generalisation of the embedded Cs did indeed occur. A sequence of embedded Cs always used the same winning node (or nodes) on every presentation, and the extra Cs beyond those used in the training sequences also used those same winning nodes.

The possible transitions terminating in each state are shown in the following table.

State	Transitions
start	AH, BH
st-1	HS
st-2	SA, CC
st-3	SB
st-4	AC, CC
st-5	AD, BD, CD
st-6	CB, DA, DA

These then, are the possible generalisations that the system could make in learning the bidirectional link FSTN. Over a number of runs, the following was observed:

- the SOMA network always generalised the two CC transitions
- it usually generalised the two DA transitions, but did separate on a number of occasions
- it usually generalised (to varying degrees) on to the same (or neighbouring) winning nodes those paths which ended at the same state

For the current FSTN, the following generalisations were observed:

State	Paths Ending	Frequency
start	AH, BH	often
st-2	SA, CC	never
st-4	AC, CC	occasionally
st-5	AD, BD, CD	mostly
st-6	CB, DA	neighbours

A trial was conducted with 500 randomly generated sentences of which 9 were unique, as shown in the following table. This represented a more complete set of training data examples.²⁷

S	A	C	B	H					
S	A	C	C	C	B	H			
S	A	C	C	C	C	C	B	H	
S	A	C	C	C	C	C	C	B	H
S	A	D	A	H					
S	A	C	C	D	A	H			
S	A	C	C	C	D	A	H		
S	A	C	C	C	C	D	A	H	
S	B	D	A	H					

On this occasion, the learned FSTN is that shown in Figure B.18. Again this structure was obtained by determining the correspondence between the allowed transitions and the winning nodes obtained from the output of the network. This FSTN is closer to the original.

The final winning nodes on both SOM surfaces in this case are given in Figure B.19.

If one follows the links of the winning nodes on the second SOM surface, the actual transition links on that surface are shown in Figure B.20. The numbers in brackets refer to the index of the winning nodes on the 8 by 8 map. As this figure shows, a number

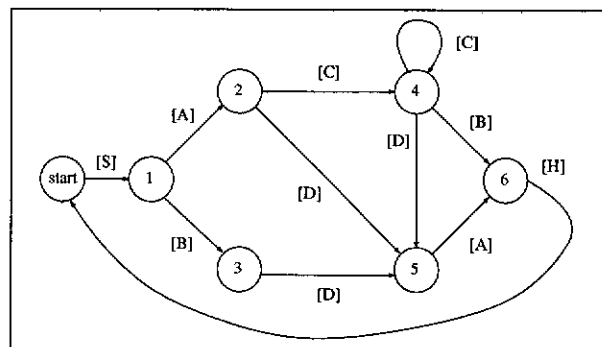


Figure B.18: Actual FSTN Learned By System—Extended Learning.

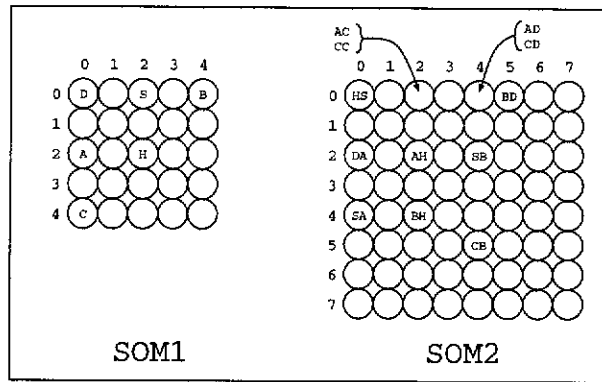


Figure B.19: Generalisation of SOM Nodes—Extended Learning.

of nodes partake in what was previously thought to be a single state; for example, the (0, 5) node is used whenever a **BD** sequence is followed by an **A** character, whereas the (0, 4) node is used whenever an **AD** or **CD** sequence is followed by an **A**. These two nodes provide the ‘functionality’ of state 6. As they are near neighbours on the SOM surface, it is possible that they be further generalised to the name node with subsequent learning.

It is less clear whether the nodes (5,4) and (2, 0), which provide the behaviour of state 6, would ever (or even *should* ever) be capable of further generalisation.

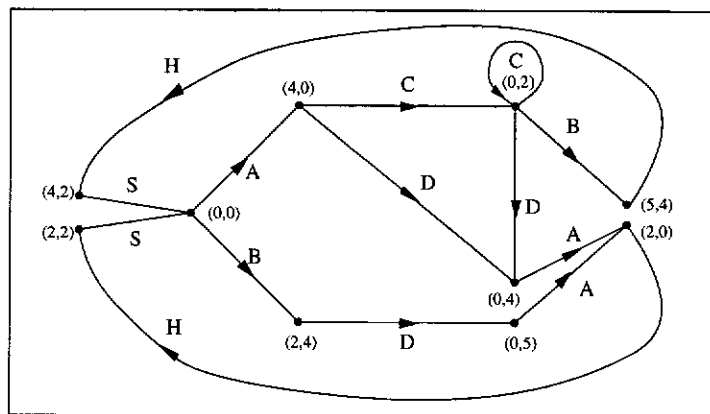


Figure B.20: Actual FSTN Learned By System—Winning Node Links.

B.2.3.2 Separating the C Loop

The problem with the FSTNs as learned so far, (and as illustrated in for example Figure B.18), is that they do not *count* the number of **C** characters in order to differentiate between an odd number of **C**s (which is followed by a **B** character), and an even number of **C**s (which is followed by a **D** character). The original FSTN of Figure B.13 does this.

In order to learn the original FSTN, it is necessary that the two **C** end points are separated. Because of the strong generalisation capabilities of the SOMA network, this is unlikely (if not impossible) in the current circumstances. If, however, two successive **C** characters are *forced* to take on different winning nodes, then it is possible for the two **C** end points to separate.

An additional run was thus performed with the addition of a refractory period (see Section 2.3.6). This run did indeed separate the **CC** loop. However, (given the test data used), the network learned other (spurious) generalisations as well, and so did not learn the original FSTN exactly. Following the flow of transitions on the SOM2 surface as input characters are processed gives the graphical representation of the recurrent links on that surface as shown in Figure B.21.

The numbers in the brackets indicates the actual winning node for the particular transition on the SOM2 surface. The network has indeed generalised the loop as the input

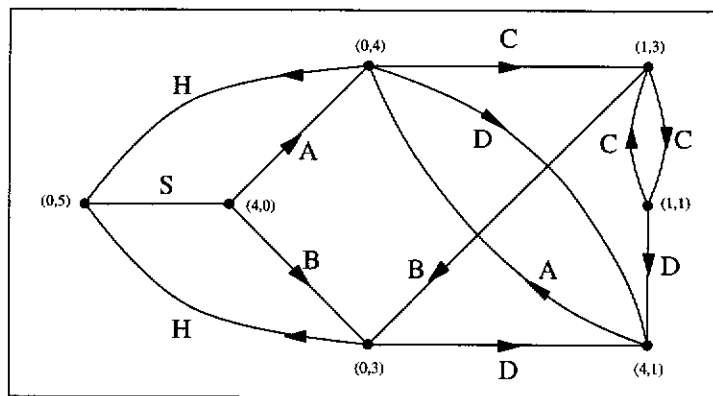


Figure B.21: Actual FSTN Learned By System—Separated C Loop.

of multiple **C**s simply oscillates between the nodes (1,3) and (1,1). Notice also that this network correctly separates the cases of odd and even numbers of **C**s. An even number of **C**s is always followed by a **D**, whereas an odd number of **C**s is always followed by a **B**.

One could imagine how this network could better generalise to be closer to the original. For example, the nodes at (0,4) and (1,1) could merge to correctly position the **C** loop. But the process would also require one of the two **A**s entering node (0,4) and one of the **B**s entering node (0,3) to each separate. The system has over-generalised in both cases.

B.2.3.3 Node Generalisation

It is reasonable to ask just what generalisation is acceptable, and just how closely should the learned structure match that of the original FSTN. Is it reasonable to expect an exact duplication? Consider the generalisation of this bidirectional link FSTN over a number of epochs, as shown in Figure B.22.²⁸

In the process of generalisation, the total number of winning nodes used by the SOMA network tends to decrease, but at times actually increases, as indicated in the following table.

Epoch	Winning Nodes/Paths	Epoch	Winning Nodes/Paths
5	78	45	26
10	66	50	24
15	50	55	24
20	39	60	24
25	45	65	27
30	112	70	24
35	73	75	12
40	139	80	12

Note that the number of winning nodes/paths is the total number of sequence paths that pass through a winning node—several sequence paths may use the same winning node. The winning node/paths for epoch 80 are:

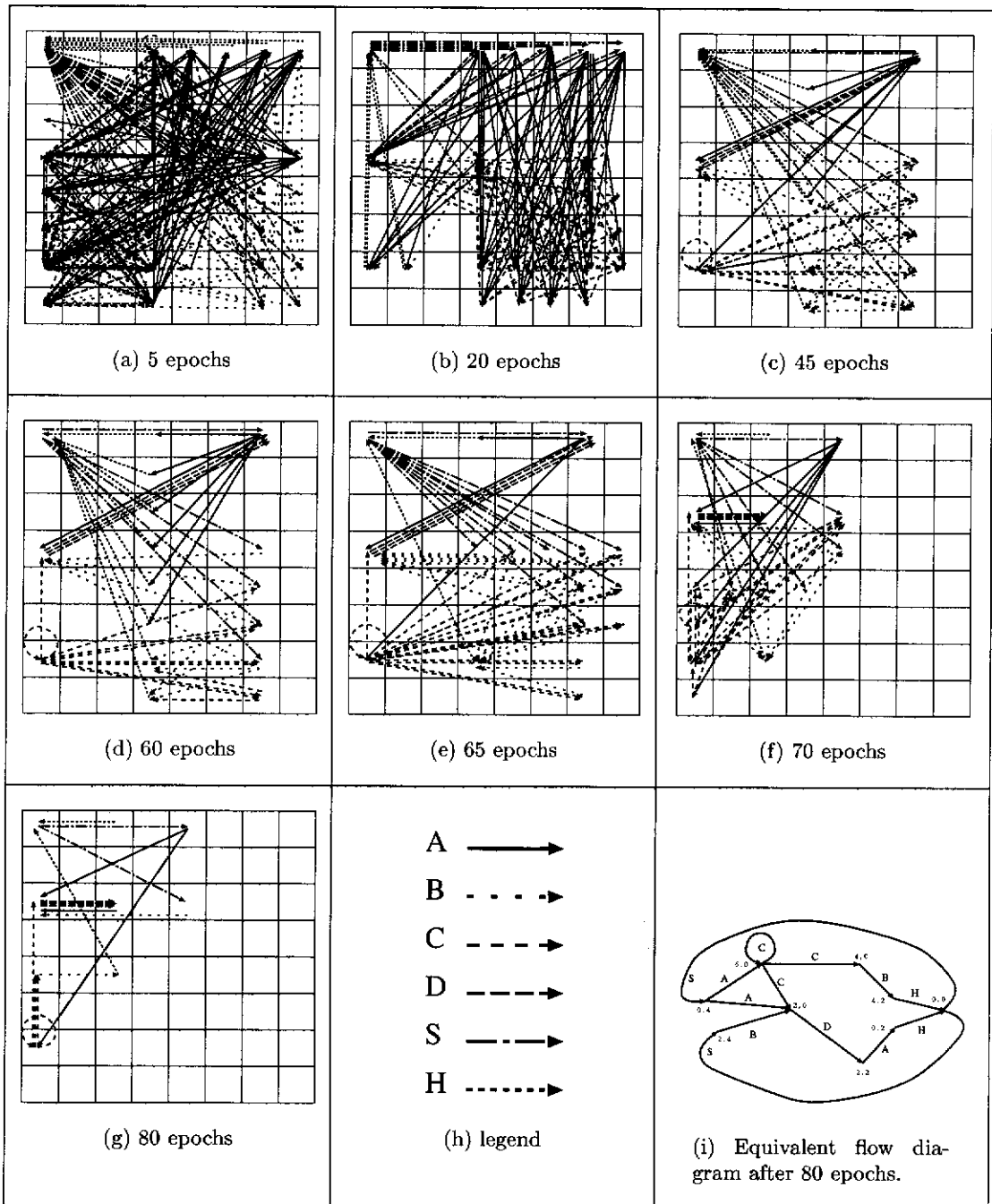


Figure B.22: Bidirectional Loop—Generalisation Over Epochs.

Sequence	Node	Sequence	Node
A	C	6,0	C
A	D	2,0	C
A	H	0,2	D
B	D	2,0	H
B	H	4,2	S
C	B	4,0	S
		2,2	A
		0,0	H
		0,4	A
		2,4	B

which corresponds to the flow diagram in Figure B.23. This also happens to be the final flow diagram in this example, as no further changes occur after epoch number 80. The equivalent flow diagram is shown in Figure B.23.

B.2.3.4 Flow Diagrams

The flow diagram shown in Figure B.23 was obtained by tracing the winning nodes and sequence transitions on the second SOM surface. We are thus able to build the flow diagram of the ‘rules’ learned by the system, and so *automatically* document these rules. For example, for the previous example, at the completion of 300 epochs the winning nodes and sequence transitions are shown in the following table.

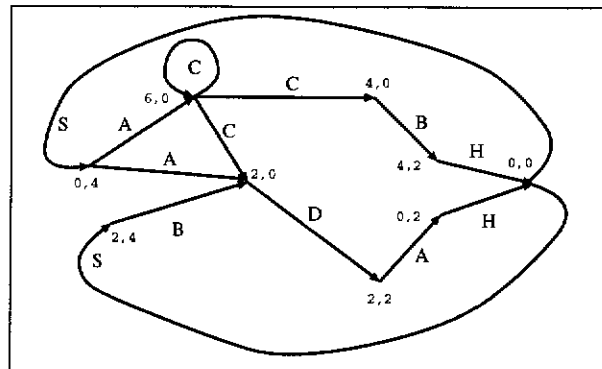


Figure B.23: Flow Diagram After 80 Epochs.

First Winning Node	First Sequence Transition	Second Winning Node	Second Sequence Transition
0,0	H-S	0,4	S-A
0,0	H-S	2,4	S-B
0,2	A-H	0,0	H-S
0,4	S-A	2,0	A-D
0,4	S-A	6,0	A-C
2,0	A-D	2,2	D-A
2,0	B-D	2,2	D-A
2,0	C-D	2,2	D-A
2,2	D-A	0,2	A-H
2,4	S-B	2,0	B-D
4,0	C-B	4,2	B-H
4,2	B-H	0,0	H-S
6,0	A-C	4,0	C-B
6,0	A-C	6,0	C-C
6,0	C-C	2,0	C-D
6,0	C-C	4,0	C-B
6,0	C-C	6,0	C-C

The table indicates that an **S** character is to be found between nodes (0,0) and (0,4) and also between nodes (0,0) and (2,4). Both follow an **H** character, but the node at (0,4) is succeeded by an **A** character whereas the node at (2,4) is succeeded by a **B** character. Linking these transitions together, it is easy to construct the flow diagram shown in Figure B.23.

The actual generalisation learned may also depend on initial random values of the weights, and on a number of other parameters. For example, varying the number of refractory periods produced the flow diagrams shown in Figure B.24. Other training sets and parameter settings might produce slightly different flow diagrams, as was found on a number of occasions during the running of the simulations reported here. [†] The flow diagram for a refractory period of three was also obtained but not displayed here. The overall pattern was similar, but more winning nodes were needed.

The point to be made here is that there may be many alternative flow diagrams, all of which satisfy the required training sequences, but each possibly over-generalised in some manner, based on the actual training data. Further analysis of the SOMA algorithm

[†]Note that the dotted line in Figure B.24 (b) indicates the likely locations of the **B** link following a sequence of five **C** characters. This sequence was not part of the training data.

and method needs to be performed to determine the conditions under which various over-generalisations are learned. Further research into over-generalisations performed by human subjects in this task could also be conducted. It may turn out that similar over-generalisation is part of human cognition.

As put by Rumelhart (1989, page 156), “for most problems there are enough degrees of freedom in the network that there are a large number of genuinely different solutions to the problems, and each solution constitutes a different way of generalising to the unseen patterns.”

B.2.3.5 Differential Learning Rates

The FSTN to be learned in this example is relatively simple, and the SOMA system is able to learn the transitions quickly. As such, it is a good ‘low level’ test of the assumptions made in Section 2.3.5 regarding differential learning periods for the two SOM surfaces and the motor weights. In this test, the number of epochs was set at 200 for all three sets of weights.

The SOMA system run over 200 epochs succeeded in learning the identical flow diagram as that shown in Figure B.23. Thus, so long as the number of epochs is sufficient to ensure that the temporal sequence is able to be learned and stabilised at the SOM1 level, subsequently at the SOM2 level, and then finally at the motor weight level, the absolute requirement for incremental learning rates is not necessary.

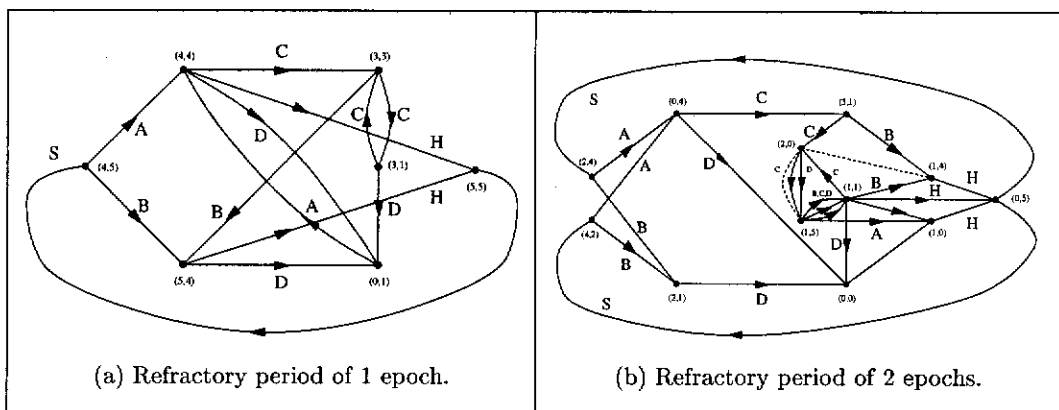


Figure B.24: Bidirectional Loop—Generalisations Over Refractory Period.

However, in order to *ensure* stability, the standard regime of 100, 200 and 300 epochs was adhered to in most test simulations.

B.2.4 Reber FSA Revisited

We return again to the Reber FSA discussed in Section B.1.4 where it was able to be learned using the FSAMap network and explicit state transition ‘rules’. In this section we examine how it is learned implicitly using example data strings only. The string generating FSTN is shown in Figure B.25. Test examples were generated with equal probability of producing an exiting character from each state.

As with the previous Embedded Loop FSTN, the first test of the sequence learning was made with a SOMA system which generated the *most likely* next output character or characters in a *local* representation. In cases in which there are multiple choices, the program must indicate those choices by giving them a high score. Disallowed choices should be given a low score. The results of this run are shown in the following table.

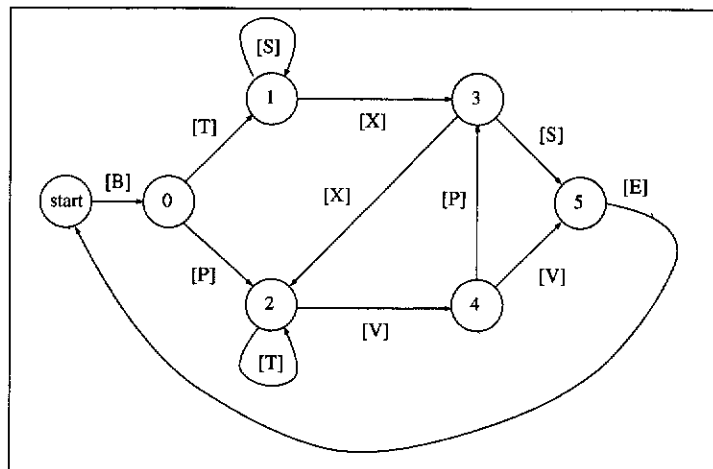


Figure B.25: Reber FSA Revisited.

Character Input 1	Character Input 2	Selected Next Char	Actual Options
B	P	V	T,V
B	T	T	S,X *
E	B	T	T,P
P	S	X	E *
P	T	T,X	T,V *
P	V	V	P,V
P	X	X	T,V *
S	E	B	B
S	S	X	S,X
S	X	X	S,X
T	T	T	T,V
T	V	V	P,V
T	X	X	S,X
V	E	B	B
V	P	X	S,X
V	V	V	E *
X	S	E	E
X	T	V	T,V
X	V	P	P,V
X	X	V	T,V

Most next character selections are correct, but those that are in error are marked with an asterisk. These errors can be explained by the SOMA system over-generalising, as will be shown subsequently.

A SOMA network was trained on 500 randomly generated strings of which 117 were unique, the longest sequences including 16 **S** repetitions, 7 **T**, and 3 repetitions of the loop **VPX**.²⁹

The first thing to note is that the same generalisation of winning nodes as observed in previous experiments was also found here. The winning nodes on the second SOM surface at the end of training are shown in Figure B.26.

Again we can piece together the actual FSTN that has been learned by the system. First we look at the output following learning. If we examine the size of the output 'concept' vector elements, we will know which characters are permitted to follow every other character according to the 'rules' learned by the network. The 'rules' are shown in the following table.

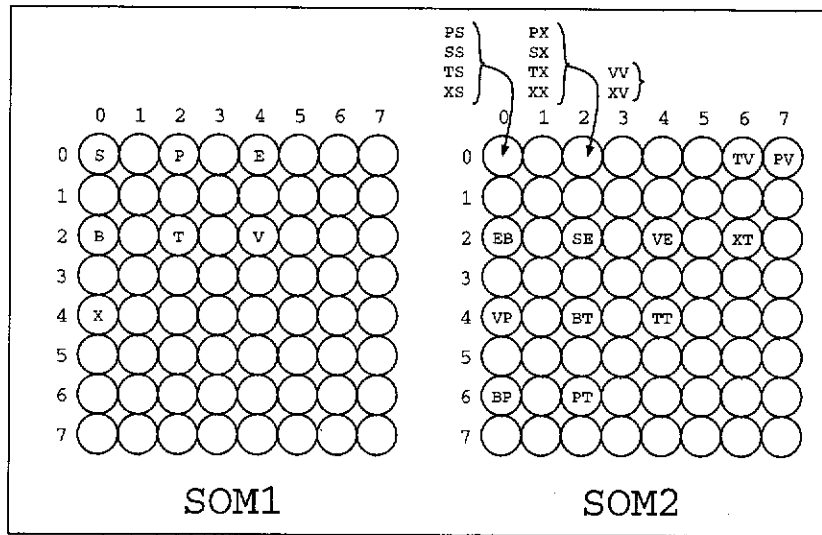


Figure B.26: Generalisation of Reber Sequence Winning Nodes.

This	B	T	P	S	V	X	E	Rule
B	0.00	1.00	0.94	0.00	0.00	0.00	0.00	B → PT
P	0.00	0.95	0.00	0.00	1.00	0.00	0.00	P → TV
V	0.00	0.00	0.95	0.00	0.98	0.00	0.00	V → PV
V	0.00	0.00	0.96	0.00	0.92	0.00	0.95	V → EPV
E	1.00	0.00	0.00	0.00	0.00	0.00	0.00	E → B
B	0.00	0.95	0.99	0.00	0.00	0.00	0.00	B → PT
T	0.00	0.00	0.00	1.00	0.00	0.94	0.00	T → SX
X	0.00	0.90	0.00	0.93	0.95	0.92	0.00	X → STVX
S	0.00	0.00	0.00	0.95	0.00	0.94	0.97	S → ESX
E	1.00	0.00	0.00	0.00	0.00	0.00	0.00	E → B
...								

By examining these generated ‘rules’ we find the following set:

Character	Allowed To Follow
B	P, T
E	B
P	S, X
P	T, V
S	E, S, X
T	S, X
T	T, V
V	E, P, V
V	P, V
V	S, T, V, X

Aggregating common exit character sets, and linking together gives an FSTN as shown in Figure B.27.

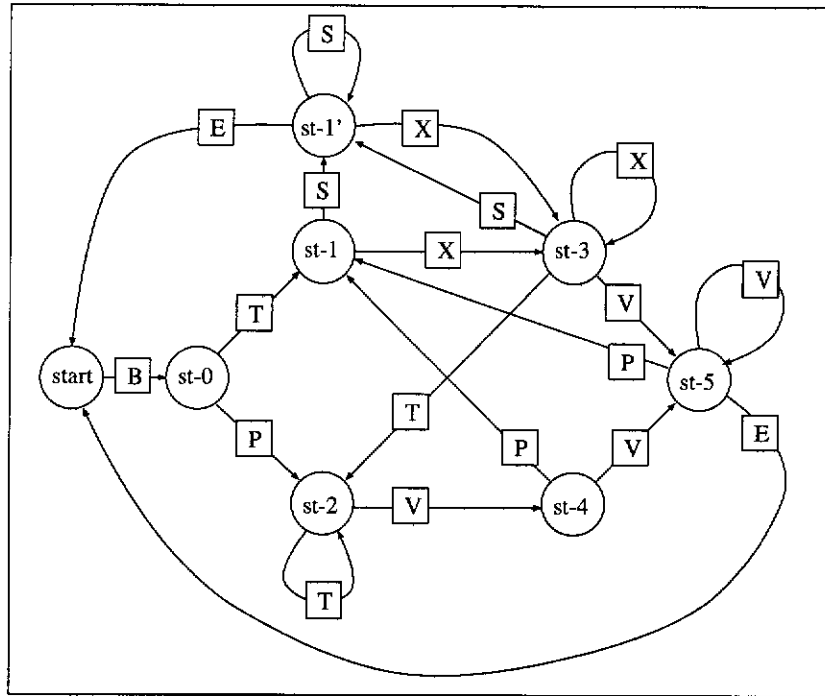


Figure B.27: Actual FSA Learned for the Reber Sequences.

Again this is actually an over-generalisation of the FSTN shown in Figure B.25. It allows all of the test sequences but also allows some not found in the test sequences; for example **BTSE** is allowed in this FSTN but not in the original. †

It is easier to appreciate the generalisation that is occurring if we note the paths ending in certain states for the second SOM surface as shown in Figure B.28.

We can trace the path of some of the allowed strings. For example,

BTXSE	start → st-0 → st-1 → st-3 → st-1' → start
BTSXSE	start → st-0 → st-1 → st-1' → st-3 → st-1' → start
BTSSXSE	start → st-0 → st-1 → st-1' → st-1' → st-3 → st-1' → start

The system definitely does perform generalisation of the loops as the same winning nodes are used. Again, it will be up to later research to discover the conditions under which further generalisation and/or specialisation may be performed with further training.

†The extent to which this *extra* string would be accepted by the learned network was not examined.

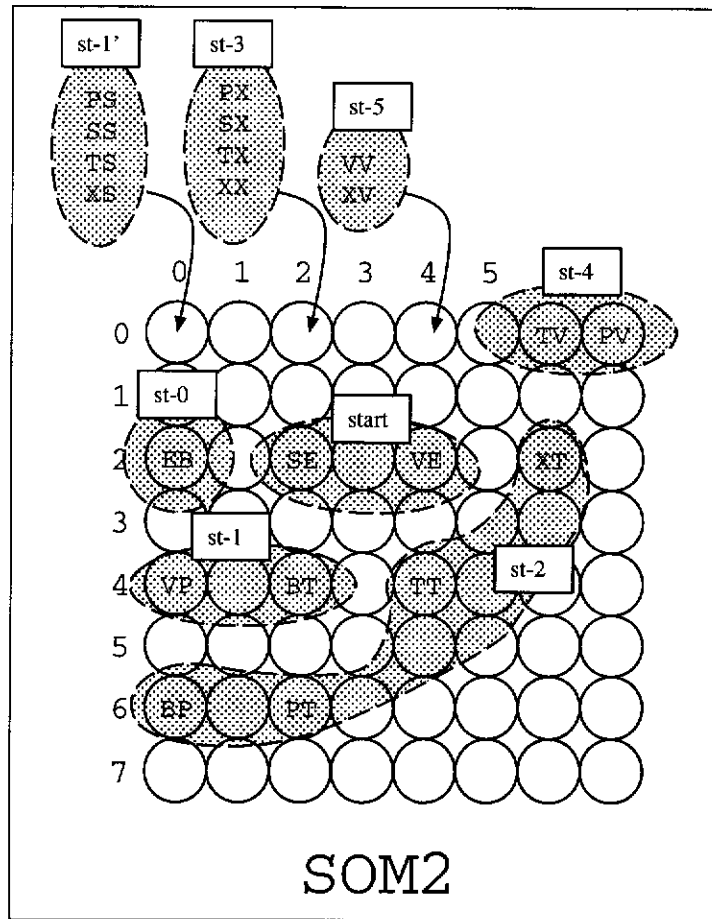


Figure B.28: Generalised Terminating Sequences At Each State.

B.2.5 Counting and Memory

Given that the SOMA network is capable of retaining information about past inputs, and is able to use this information to learn temporal sequences, it is appropriate to ask how far back this 'memory' extends. In order to test this proposition, input of the form

6 a a a a a 6 a a a a a 6 a a a a a 6 a a a a a ...

was presented to the system. To learn this sequence, the SOMA system will need to 'count' across the **a** characters. †

†Note that the number **6** in this case has no significance, and is just another input character represented by a given input vector. Any other character would have sufficed.

The SOMA network was able to learn with counts of 3, 4, 5 and 6 (above) with the usual training regime. Thus the system was able to look back at least 6 input characters or words. However initial attempts to get the system to memorise a seven character sequence met with failure.

The way in which the system was able to learn these sequences resulted from it being able to learn and select different winning nodes for each of the **a** to **a** transitions. A typical example of the SOM2 winning nodes in the case of the six character sequence is shown in Figure B.29.

In order to extend the memory capacity of the SOMA system, the extension to the SOMA model discussed in Section 2.3.6—a *refractory period*—was used. That is, neurons which have just fired are unable to fire again for some period. The modification to the SOMA program prevented any neuron which was a winning node at time t from again being selected as a winning node at time $t+1$. It was thought that this might be able to provide further ‘across time’ cues to extend the memory period.

This proved to be the case. With this simple extension, the SOMA system *was* able to extend the counting period to 7. The sequence:

7 a a a a a a 7 a a a a a a 7 a a a a a a 7 a a a a a a ...

was able to be learned and predicted to 100% accuracy.

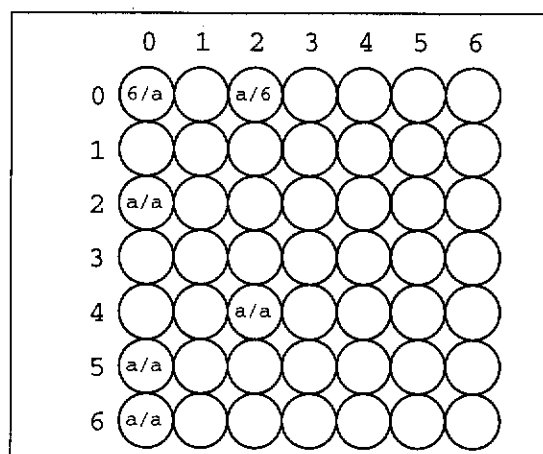


Figure B.29: Counting Via Temporal Memory.

Despite considerable efforts, it was not possible to get the system to ‘count’ to eight.

An analysis of some of the parameters of the model showed that the σ component of the Kohonen exponential output function $\exp(\frac{-x^2}{2\sigma^2})$ was extremely important in this sequence learning task. Using the SOMA model from the refractory period simulation, and keeping all other parameters constant, the following table was generated.³⁰

n	$\sigma \geq$	$\sigma \leq$
7	0.386	0.396
6	0.382	0.462
5	0.004	0.485
4	0.004	707.1

This indicates the permissible range of values for sigma in order to be able to learn the sequence of n characters. The range is very wide for small values of n , but becomes exceedingly small for seven characters.

The table strongly suggests, at least for this combination of parameters and experimental setup, that the limit is seven, and this is supported by anecdotal evidence resulting from the failure to extend the learning further despite numerous (but not rigorous) attempts at other parameter settings.

It is too tempting to forego a passing reference to the similarity between the limit of seven found here, and the magic number 7 ± 2 referred to by Miller in his famous paper on the limits of the human capacity for processing information (Miller 1994). In this paper, Miller refers to numerous examples of human capacities of judgement, span of attention and memory, and so on, which seem to be limited.

One of these is the capacity of short-term memory, and it appears that humans have an ability to recall about six or seven items if (say) they are asked to read a random list of digits (of increasing size) and to repeat the sequence a short time later. Individuals will vary widely, some managing only four or five, while others may be able to recall ten or more. But without resorting to some form of learning strategy, the limit appears to be around seven, and this number appears to be the limit for other forms of short term memory as well.

B.2.5.1 Other Sequences

Other combined counting exercises that included strings such as

01a2bb3ccc4dddd5eeee

were tried with success; that is, all of the memories could be stored simultaneously.

Other sequences that were able to be learned include

01a2aa3aaa4aaaa5aaaa

(that is, a number of ‘counting’ sequences are included), and

01a2ba3cba

This latter sequence is interesting as it is context sensitive. For example, a **2** follows an **a** provided the **a** follows a **1**, but if the **a** follows **2b** then it is followed by a **3**. And if the **a** is preceded by **cb** then it is to be followed by a **0** (assuming wrap-around of input sequences).

B.2.5.2 Short-Term Memory and Chunking

It is not appropriate at this stage to say anything other than speculate on a mechanism for the limitation of short term memory, given the veracity of the ABC model.

In forming a ‘temporary’ link between a number of already stored attractors (for example, digits, if we are trying to remember say a phone number) the link traverses around the second SOM layer as shown in Figure B.30. A sequence that is too long will be overextended because of the need to pass through an ‘exponential-like’ function (or neurobiological equivalent) at each traversal. It is not appropriate to say any more than to highlight this possibility and to suggest that it is worthy of more research in the future.

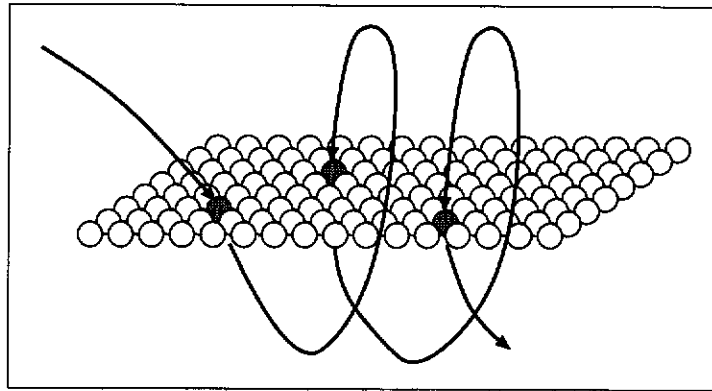


Figure B.30: Short Term Memory—Possible Reason For Limitations.

Of course, this limitation may be overcome by *chunking*, the labelling of a sequence (of perhaps in turn up to seven items) into a new item that can itself be included as an item in a sequence, and so on. That is, the ABC model also suggests a possible mechanism for chunking as shown in Figure B.31.

In Figure B.31 the strings x_i and y_i are initially linked into separate sequences. By linking A to the x_i and B to the y_i , and learning a new sequence $ABC \dots$, it will be then easier to remember this smaller sequence (provided the connections between A and x_i (and between the x_i themselves), and between B and y_i (and between the y_i), is maintained). The extended sequence $x_0, x_1, \dots, x_i, y_0, y_1, \dots, y_j$ is thus able to be recalled by the A and B links.

Note that Figure B.31 is a highly stylised diagram, and is only suggestive of a process at this stage. Further research is required to examine this hypothesis.

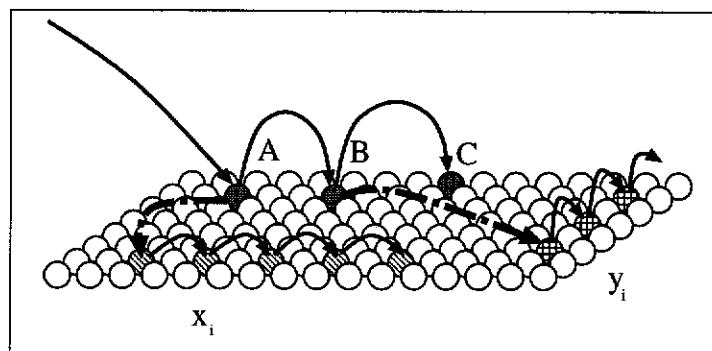


Figure B.31: Possible Mechanism For Chunking.

B.2.6 Embedded Sequences

Embedded sequences have been studied in Cleeremans (1993) and Cleeremans et al. (1989) in relation to implicit learning. The chosen paradigm for the study of implicit learning by Cleeremans is the learning of grammars by both computer and humans, and the network model used is a simple recurrent network (SRN).

A number of different finite-state grammars are studied, but much attention is given to the FSTN shown in Figure B.32. This FSTN includes two embedded clauses which are identical, and which are simplified versions of the FSTN studied in Section B.2.4, the difference being that the **S** and **T** loops have been removed. However, the **XVP** loop remains.

The two component sub-clauses differ in that the upper sub-clause is preceded and succeeded by a **T** character, whereas the lower sub-clause is preceded and succeeded by a **P** character. In order to correctly predict the character following an exit from a sub-clause, the system must be able to remember the lead-in character.

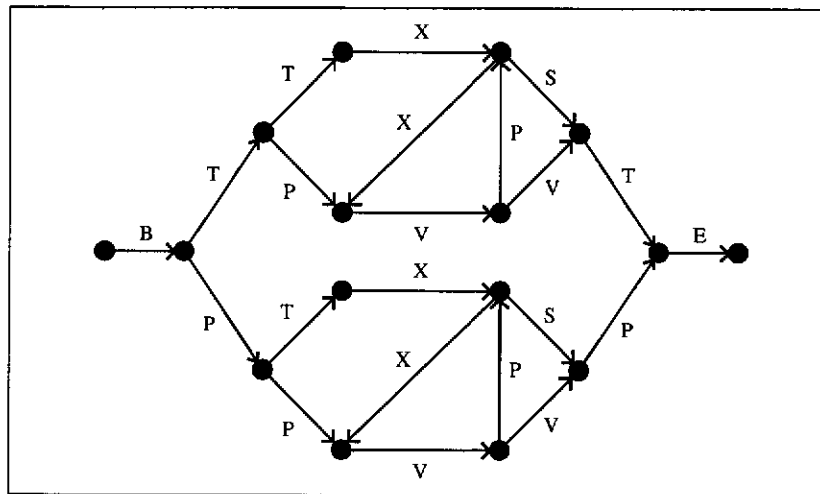


Figure B.32: Complex Finite-state Grammar With Embedded Clauses.

B.2.6.1 Learning the Sub-clause

As a first step, let us first examine the internal component common to both the top and bottom sub-clause structures. This is shown in Figure B.2.6.1.

A SOMA network was trained on 100 random sentences generated from the FSTN of Figure B.2.6.1. Of these sentences, twelve were unique. The longest sequence was B T X X V P X V P X V P X V V E. A refractory period of one was used to separate consecutive identical characters onto different winning nodes. ³¹

The SOMA system was able to learn the FSTN sequence, and it did so in a manner that is worthy of some discussion. The winning nodes on both SOM layers are shown in Figure B.34.

As discussed previously, the actual structure of the ‘rules’ within Figure B.2.6.1 that are learned by the system can be analysed at a ‘neuron’ level by following the sequence of winning nodes and constructing a diagram of the linkages between these winning nodes. First let us examine the flow of winning nodes for the first SOM surface, as shown in Figure B.35. The first SOM surface records the winning nodes for each input character.

In Figure B.35, the numbers in brackets beside each letter indicate the actual winning node for that character. Notice that the **P** corresponding to node (0,3) is only used if it is the second **P** within the refractory period range, (that is, the sequence **B P V P**). Also, the **X** winning node at (4,4) is similarly only used for the second **X** within the

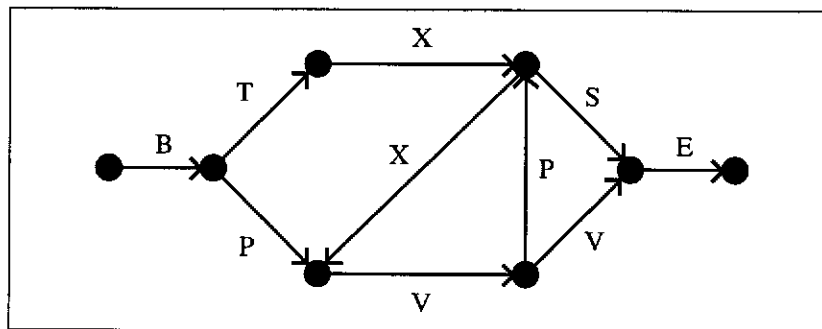


Figure B.33: Complex Finite-state Grammar—Internal Sub-clause.

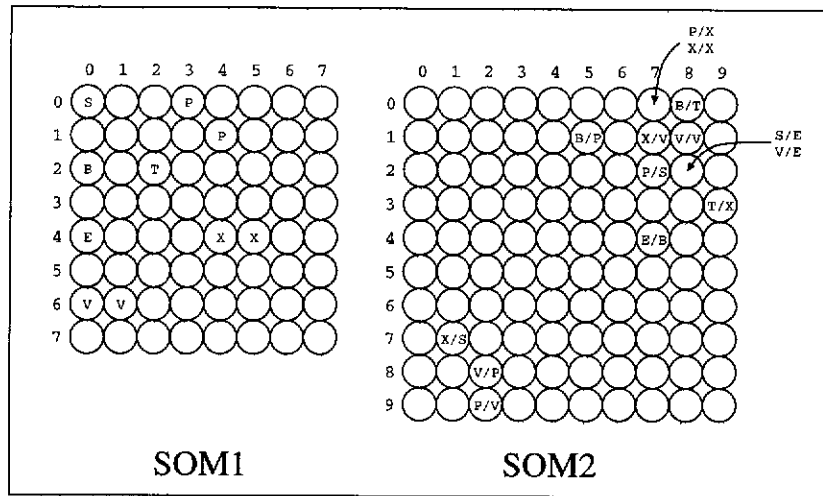


Figure B.34: Complex Finite-state Grammar—Internal Component Winning Nodes.

refractory period, (that is, the sequence **B T X X**). The loop path **XVP** is conducted on the nodes **X** (4,5), **V** (6,0) and **P** (1,4).

These particular nodes are selected by chance in that the original best selections for **X** and **P** depend on randomly allocated weights on the initial SOM surface. The SOMA algorithm will continue to select these particular nodes unless redirected to another node by the refractory period extension. A different mix of input data would produce a different result.

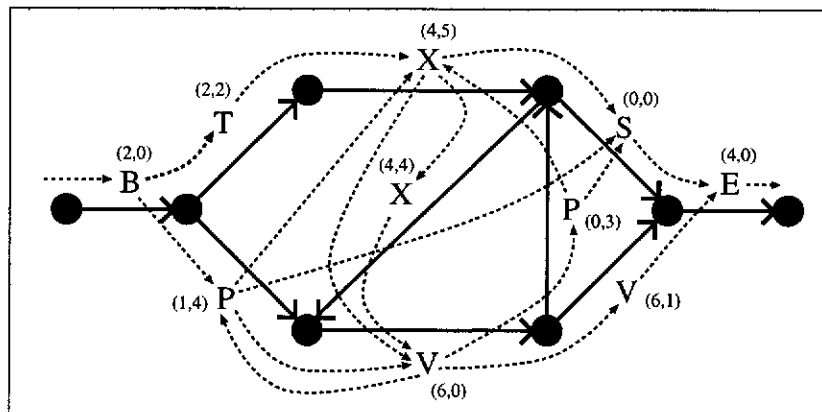


Figure B.35: Complex Finite-state Grammar—Linked Character Winning Nodes.

The linkage of winning nodes on the second SOM surface is also worthy of some reflection. The structure learned by the SOMA network on the second SOM layer is shown in Figure B.36. Again this was determined by following the sequence of winning nodes used in learning the input sequences.

As can be seen, the learned structure is very similar to the FSTN of the training examples. It is possible to surmise how further generalisation and merging of winning nodes following additional learning of input examples could reproduce the original structure. For example, the nodes (0,7) and (2,7), which are near neighbours on the SOM surface, could merge with further training. However, it is not clear whether the nodes (8,2) and (1,8), and also (1,7) and (9,2) could merge, as they are more widely spread on the SOM surface. It is also not clear whether this merging is necessary.

B.2.6.2 Learning the Full Grammar

Cleeremans was unable to get an SRN to learn the full grammar of Figure B.32 if the exit characters at each node were selected with equal probability in the training sentences. That is, if in generating the training examples, a choice is to be made between two subsequent characters exiting from a particular state on the FSTN, and each is given a probability of being selected of 50%, then the training set generated will not statistically differentiate between the upper and lower sub-clauses, and an SRN will not be able to learn the full FSTN. It will not successfully predict the correct trailing character before the terminating E character. Information about the lead-in

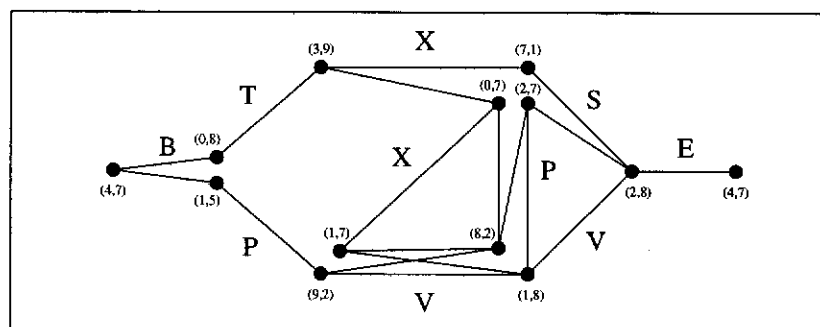


Figure B.36: Complex Finite-state Grammar—Linked Transition Winning Nodes.

character is *not* locally relevant in predicting the successor of any character within the embedded sub-clause. The SRN network will indifferently predict both possible terminating characters regardless of the lead-in character.

If, however, a slight statistical bias was introduced which tended to favour the top arcs for the upper sub-clause and the bottom arcs for the lower sub-clause, then the predictive performance of the SRN network improved dramatically.

As with Cleereman's experiments, and despite considerable effort, the SOMA network could not be made to learn the FSTN of Figure B.32. The SOMA network tended to over-generalise the **T** and **P** characters which occur both *within* the embedded characters and as the *enclosing* characters, assigning them to the same winning nodes.

One of the proposals of this thesis is that language is *not* simply serial string processing, but that other parallel (contextual) components are important in language processing. Cleeremans, along with most of the linguistics literature, make the assumption that only serial processes are involved.

In this case, it would be appropriate to see if the SOMA network can disambiguate the two sub-clauses by using context. In order to test this hypothesis, an additional 3 bits of context were added to the 8 bits used to identify each input character, and the concatenation of these two components used as the input vector to the first SOM surface.

The three extra bits of context were added to every character in each sentence of the training set such that if the sentence sequence involved the upper sub-clause (say **BTTXSTE**), then it had the context of 000 added, whereas if the lower path was taken (for example, the sentence **BPTXSPE**) then every character had a context of 111 added. The contextual bits were only added as an additional input to the first SOM surface.

With this added context, the SOMA network *was* able to disambiguate the two sub-clauses, and to correctly predict the training lead-in character. The learned flow diagram is shown in Figure B.37. ³²

Note that the system has over-generalised differently in the two halves of the FSTN

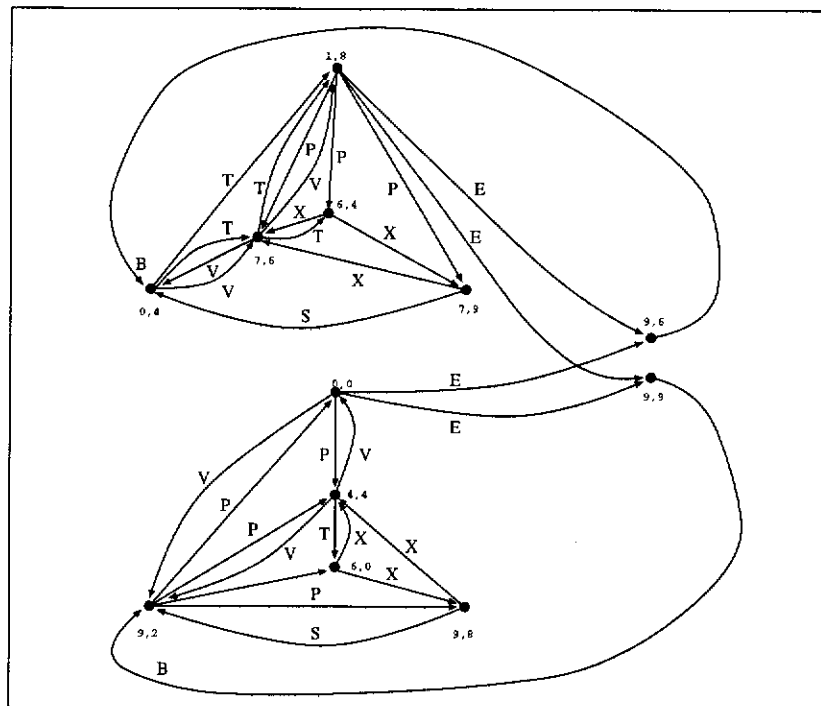


Figure B.37: Learning Embedded Sequences Using Context.

as the randomly selected training data resulted in slightly different sequences being learned on each half.

However, the SOMA network was able to correctly separate the end nodes, as is to be expected from the experimental setup.

B.2.6.3 Enumeration and Generalisation

One of the major differences between the current SOMA method, and the previous SRN networks that have been popular in the study of sequence learning, is the ability of the SOMA network to *generalise* on the hidden layer.

Simple recurrent networks (SRN) (and by extension FFNNs) tend to not generalise on the hidden layer but to *enumerate* the possibilities. This was described in Section B.2.3 where it was pointed out that the SRN network learned specific path information rather than generalising.

This enumeration is also apparent from the work of Elman in his simple sentence

learning task (Elman 1990). Figure B.38 is reproduced from page 206 of that article. It shows a hierarchical cluster analysis of the hidden unit activation vectors in response to some occurrences of the input word **boy**. Upper-case indicates the actual input, while the lower-case words are the context words in each case.

There is structure in the subtrees—the tokens of **boy** which occur in a sentence-final position are clustered together, as are the tokens of **boy** which appear in sentence-initial positions. The tokens are grouped by context (Elman 1990, page 205). However, the tokens do remain *separated*, with a distinct (but similar) representation for each. In this sense the sentence paths (context) have all been enumerated, and the generalisation is achieved by proximity

The generalisation in the SOMA network is much stronger than that found on the SRN hidden layers. The self-organisation of the lateral connections on the SOM layers ensures that each token is associated with a small number of nodes (usually one if no refractive period is included).

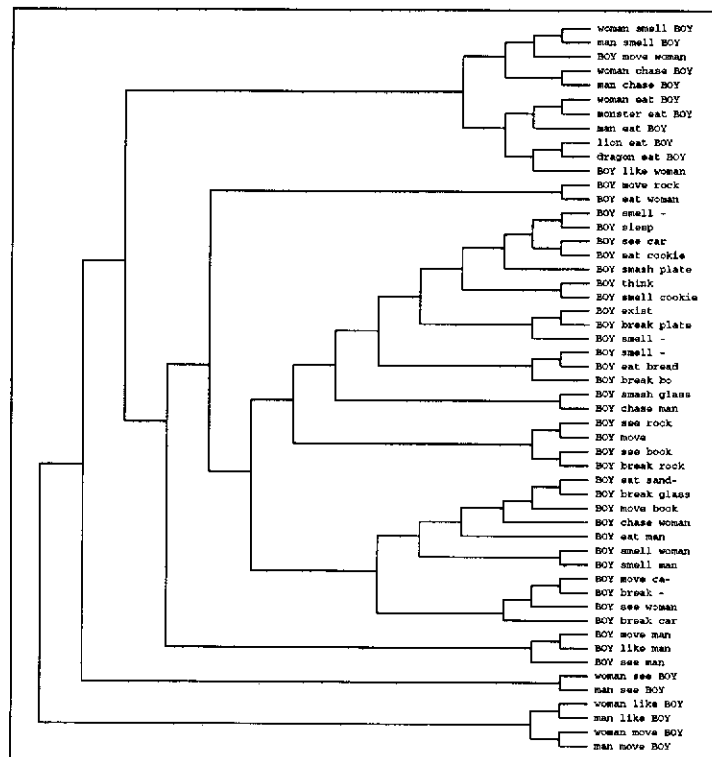


Figure B.38: Hierarchical Cluster Diagram—SRN.

B.2.6.4 Explanation and Verification

There has been a reluctance in some quarters, and especially in industry, to use artificial neural networks in a production environment. This is because of the inability of the users to actually *look at* and *check* the ‘rules’ that have been learned. The preference is to explicitly *program* the rules (which have been determined by prior theoretical and experimental research) using a computer language, or to enter the rules as *data* in an expert system.

The learning of rules by an artificial neural network (ANN) is *implicit*. This is opposed to the *explicit* use of rules as is the case with programming or expert systems. The learning of rules in machine learning is considered acceptable as the rule *components* are already in some semantic form, and the machine learning task is simply to combine these semantic components to cover the training examples. The learned rules are then deemed to be *human readable*. Users are not comfortable with the perceived *black box* learning of ANNs. They are unsure of the possibility of the net failing for certain inputs, with the perception that the results produced could be incorrect. For example, there is an understandable reluctance to use an ANN in the design of a bridge for both technical and legal reasons.

Some methods have been developed to extract rules from ANNs (see, for example, Sestito & Dillon 1991, Sestito & Dillon 1994), but in general the methods have been unsuccessful. Further, these methods to date have been directed towards learning rules from FFNNs.

This disinclination to accept decisions except in the presence of explicit rules has an interesting corollary in the case of knowledge acquisition for use in expert systems. Here so-called knowledge engineers work with domain experts in an effort to cast the ‘rules’ that the expert uses into symbolic form. Those who have worked as knowledge engineers (including the author) invariably find that the human experts often have no real idea as to why they make certain decisions, especially in the case of perceptual choices—they do *not* appear to have access to explicit rules either. Despite this, an expert’s opinion is generally highly regarded because of their *previous successful behaviour*.

However, for various reasons, it is important that we be able to extract the rules that have been learned by a neural net—to perhaps put them into some explicit form, but at least to check that the learned rules are valid and will do the task required efficiently and without error.

The SOMA method has an especially powerful means of extracting the learned rules as the discussion in previous sections has shown. The processes involved in the SOMA method, especially the generalisation at the hidden layer (the second SOM surface) are much more ‘atomic’ (symbolic) than in the case of FFNNs (where statistical methods are usually required). An analysis of the winning nodes for both surfaces, in conjunction with the ‘symbolic’ inputs, allows us to put the learned rules into a verifiable form.

It is beyond the domain of this thesis to take this matter any further at this stage, but rather to leave it to future research. It is likely however, that a full analysis tool could be developed to explicitly check and enumerate the rules learned by the SOMA network. It may also be possible to include an *error exit* facility (as was used in the explicit FSA experiments described previously) should the network encounter an error condition, such as receiving spurious inputs.

B.2.7 English Sentences

A number of researchers have looked at sentence processing using neural networks (Ritter & Kohonen 1989, Elman 1990, Elman 1991, Elman 1993, St. John & McClelland 1990, McClelland & Kawamoto 1986).

In this section, we look at the performance of the SOMA network in learning simple English sentences. The sentences generated were similar to those used in Elman (1990).

The initial trial consisted of 100 simple English sentences, such as *a man smelt some cookies*. Ninety-nine of these were 5 word sentences (article noun verb article noun), with one sentence having a two word verb (listened to). The words of the sentences were randomly selected from the following tables of nouns, verbs, and articles. There were 6 classes of noun, with both singular and plural nouns included. The nouns used are shown in Table B.4 (a). The verbs used were all transitive, and only included

third-person and past tense. These were divided into 4 separate classes as shown in Table B.4 (c). A reduced set of articles, as shown in Table B.4 (d) was used.

The sentences were forced to maintain some semantic ‘reality’ by combining the various noun and verb classes according to the ‘connecting rules’ shown in Table B.4 (b).

The random selection of 100 sentences gave a total of 22 nouns, 9 verbs, and 3 articles as shown in the following table.

<table border="1"> <thead> <tr> <th>Type</th> <th colspan="2">Words</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>man</td> <td>men</td> </tr> <tr> <td></td> <td>woman</td> <td>women</td> </tr> <tr> <td>2</td> <td>cat</td> <td>cats</td> </tr> <tr> <td></td> <td>mouse</td> <td>mice</td> </tr> <tr> <td>3</td> <td>book</td> <td>books</td> </tr> <tr> <td></td> <td>rock</td> <td>rocks</td> </tr> <tr> <td>4</td> <td>lion</td> <td>lions</td> </tr> <tr> <td></td> <td>tiger</td> <td>tigers</td> </tr> <tr> <td>5</td> <td>cup</td> <td>cups</td> </tr> <tr> <td></td> <td>plate</td> <td>plates</td> </tr> <tr> <td>6</td> <td>cookie</td> <td>cookies</td> </tr> <tr> <td></td> <td>bread</td> <td>bread</td> </tr> </tbody> </table> <p>(a) Nouns.</p>	Type	Words		1	man	men		woman	women	2	cat	cats		mouse	mice	3	book	books		rock	rocks	4	lion	lions		tiger	tigers	5	cup	cups		plate	plates	6	cookie	cookies		bread	bread	<table border="1"> <thead> <tr> <th>Subject Noun Type</th> <th>Verb Type</th> <th>Object Noun Type</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>1</td> <td>6</td> </tr> <tr> <td>1</td> <td>2</td> <td>1</td> </tr> <tr> <td>1</td> <td>2</td> <td>2</td> </tr> <tr> <td>1</td> <td>2</td> <td>3</td> </tr> <tr> <td>1</td> <td>2</td> <td>4</td> </tr> <tr> <td>1</td> <td>2</td> <td>5</td> </tr> <tr> <td>1</td> <td>2</td> <td>6</td> </tr> <tr> <td>1</td> <td>3</td> <td>2</td> </tr> <tr> <td>1</td> <td>4</td> <td>3</td> </tr> <tr> <td>1</td> <td>4</td> <td>6</td> </tr> </tbody> </table> <p>(b) Rules.</p>	Subject Noun Type	Verb Type	Object Noun Type	1	1	6	1	2	1	1	2	2	1	2	3	1	2	4	1	2	5	1	2	6	1	3	2	1	4	3	1	4	6
Type	Words																																																																								
1	man	men																																																																							
	woman	women																																																																							
2	cat	cats																																																																							
	mouse	mice																																																																							
3	book	books																																																																							
	rock	rocks																																																																							
4	lion	lions																																																																							
	tiger	tigers																																																																							
5	cup	cups																																																																							
	plate	plates																																																																							
6	cookie	cookies																																																																							
	bread	bread																																																																							
Subject Noun Type	Verb Type	Object Noun Type																																																																							
1	1	6																																																																							
1	2	1																																																																							
1	2	2																																																																							
1	2	3																																																																							
1	2	4																																																																							
1	2	5																																																																							
1	2	6																																																																							
1	3	2																																																																							
1	4	3																																																																							
1	4	6																																																																							
<table border="1"> <thead> <tr> <th>Type</th> <th colspan="2">Words</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>ate</td> <td>devoured</td> </tr> <tr> <td></td> <td>caught</td> <td>swallowed</td> </tr> <tr> <td>2</td> <td>saw</td> <td>imagined</td> </tr> <tr> <td>3</td> <td>heard</td> <td>listened to</td> </tr> <tr> <td>4</td> <td>touched</td> <td>smelt</td> </tr> </tbody> </table> <p>(c) Verbs.</p>	Type	Words		1	ate	devoured		caught	swallowed	2	saw	imagined	3	heard	listened to	4	touched	smelt	<table border="1"> <thead> <tr> <th>Type</th> <th colspan="2">Words</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>the</td> <td>the</td> </tr> <tr> <td>2</td> <td>a</td> <td>some</td> </tr> </tbody> </table> <p>(d) Articles.</p>	Type	Words		1	the	the	2	a	some																																													
Type	Words																																																																								
1	ate	devoured																																																																							
	caught	swallowed																																																																							
2	saw	imagined																																																																							
3	heard	listened to																																																																							
4	touched	smelt																																																																							
Type	Words																																																																								
1	the	the																																																																							
2	a	some																																																																							

Table B.4: English Sentence Components.

Nouns		Verbs	Articles
book	man	caught	a
books	men	devoured	some
bread	mice	heard	the
cat	mouse	imagined	
cats	plate	listened to	
cookie	rock	saw	
cookies	rocks	smelt	
cup	tiger	swallowed	
cups	tigers	touched	
lion	woman		
lions	women		

The sentences were generated by randomly selecting a sentence 'rule', and then randomly selecting appropriate nouns, verbs and articles to make up the sentence. The sentences were processed as a contiguous stream, so that the end word of one sentence would be followed by the first word of the next sentence. The last word of all of the sentences wrapped around to precede the first.

The results obtained were excellent. After training for the standard epoch regime, the results as shown in Table B.5 were obtained: ³³

The table indicates that the system was very capable of predicting the next word, and especially word type. For example, of the 200 nouns in the sentences, 34 were predicted exactly based on the learned statistics. In a further 139 cases, a noun other than that expected was selected. Thus, in 86.5% of cases, the system was able to correctly predict that a noun should occur as the next word.

Noun Expected			Verb Expected		
Exact Word	Other Noun	Other Type	Exact Word	Other Verb	Other Type
34	139	27	36	64	0
173		27	100		0
86.5%			100%		
Article Expected			Preposition Expected		
Exact Word	Other Article	Other Type	Exact Word	Other Prep	Other Type
98	86	16	1	0	0
184		16	1		0
92%			100%		

Table B.5: Simple English Sentences.

The situation for verbs and articles was even better, with the exact verb being predicted in 36% of cases, and an alternate verb selected in the other 64% of cases—the correct selection of a verb in 100% of cases.

Other studies using FFNNs have used a much larger number of sentences. However, even with this sized example, we run into significant time problems. †

The SOMA architecture does require more computational resources than the FFNN alternative. This is because the SOM surface matrices are generally larger than the corresponding vectors of the FFNN model, with corresponding increases in numbers of weights. Both increased compute time and memory are required. We discuss this point, and the need for massively parallel hardware simulation devices in Section 6.1.

Further trials of larger sentence data sets will need to be set aside for future research with a more powerful computer.

B.2.8 Other Language Sentences

One of the strengths of the full ABC cognitive model is that it is not predicated on any particular language. The only requirement is that temporal sequences are able to be learned. Thus the temporal learning that is described in this chapter could be applied equally to other human languages.

Word order restrictions vary considerably in human languages. Some have a free word order (for example Latin, Navajo), whereas some, such as English, French and Vietnamese have SVO as their basic pattern. Other orderings are SOV (Japanese, Tibetan, Korean) and VSO (Welsh, Tongan). These orderings account for some 85-90% of the world's languages. The other subject, verb, object orderings (VOS, OSV, and OVS) are rarer, but are found in multiple languages (see Crystal 1992, page 98).

Some initial trials using the SOMA network in the learning of German and Japanese have been conducted with similar success. These results are not reported here because of time and space limitations.

†For example, this example took some two days to complete on a Silicon Graphics Challenge—a not insignificant machine.

Future research will include other language learning, including positional independent languages such as is found in some Australian Aboriginal languages.

B.2.9 Embedded English Sentences

In another influential paper on temporal learning, Elman (1993) conducted a number of experiments on embedded English sentences. He showed that the SRN architecture was capable of predicting the successive word options following training with sentences generated by a phrase structure grammar. The phrase structure grammar used by Elman was almost identical to the one described in Table B.6.

In the usual parlance of linguistics, **S**, **NP**, **VP**, **RC**, **N**, **VT** and **VI** refer to a sentence, noun phrase, verb phrase, relative clause, noun, transitive verb and intransitive verb respectively. The | symbol specifies alternatives, and the italicised words are the terminal symbols of the grammar.

We can express these grammatical ‘rules’ in the form of an FSTN as shown in Figure B.39. The subscripts in this figure refer to the fact that the nouns and verbs involved in valid constructs must match for the singular/plural number case. For exam-

S	→	NP VP ‘.
NP	→	N RC
RC	→	<i>who</i> NP VP <i>who</i> VP NP
N	→	<i>boy</i> <i>girl</i> <i>dog</i> <i>cat</i> <i>boys</i> <i>girls</i> <i>dogs</i> <i>cats</i>
VP	→	VT N VI
VT	→	<i>chases</i> <i>likes</i> <i>hears</i> <i>sees</i> <i>chase</i> <i>like</i> <i>hear</i> <i>see</i>
VI	→	<i>smiles</i> <i>walks</i> <i>smile</i> <i>walk</i>

Table B.6: Phrase Structure Grammar for Embedded Sentences.

- one level of embedding

0 - 1 - 3 - 1 - 0
 0 - 1 - 3 - 1 - 2 - 0
 0 - 1 - 3 - 4 - 1 - 0
 0 - 1 - 3 - 4 - 1 - 2 - 0
 0 - 1 - 3 - 5 - 1 - 0
 0 - 1 - 3 - 5 - 1 - 2 - 0

- two levels of embedding

0 - 1 - 3 - 5 - 6 - 7 - 5 - 1 - 0
 0 - 1 - 3 - 5 - 6 - 7 - 5 - 1 - 2 - 0
 0 - 1 - 3 - 5 - 6 - 8 - 5 - 1 - 0
 0 - 1 - 3 - 5 - 6 - 8 - 5 - 1 - 2 - 0

The correspondence between the Phrase Structure Grammar and the FSTN is not exact. For example, the PSG includes the possibility of an infinite number of embedded clauses whereas the FSTN only allows for a maximum of two. Most actual language usage, however, rarely extends beyond one embedded clause. The sentence

The man [the girl [I used to go out with] married] just got drafted.

is double center-embedded. Center embedding is difficult for most speakers and presents performance problems. However, single center-embedded sentences in English usage are common, such as

The girl [I used to go out with] got married.

It is suggested that double center-embedding is not used in most human languages, and so the more powerful grammar suggested by Chomsky (Miller & Chomsky 1963) is not required. The exception is Dutch, which can have a complicated double embedding structure.

The FSTN figure is deliberately limited in two ways: first, there is no intransitive verb path between states **6** and **5**; and second, the noun on exit from state **8** need not be restricted to match the number of the transitive verb leading into state **8** but is rather free. These limitations are removed in a later experiment.

The claim of many in the the linguistic community is that embedded sentences such as these must be parsed, (that is, understood), on the basis of the words in the sentence alone. Miller & Chomsky (1963) argued that a sentence such as

The people who say they want to rent your house next summer while you are away in Europe are from California.

requires a dependency relationship that extends for 17 words, from *people* to *are*, in order to ensure the correct number (plural) is used for the verb. Miller & Chomsky conclude that this makes statistically-based language learning infeasible. They suggest that if the learner is able to use co-occurrence statistics alone, then correct number agreement for this sentence could not be achieved without sampling all the possible sentences which contain *people* and *are* separated by 17 words—a somewhat large number.

As correctly pointed out by Elman (1993, page 87), neural networks do not merely compile statistics for a lookup table procedure, but rather act as function approximators. In addition, the statistical correlations within SRN and SOMA networks extend to more than simply adjacent inputs.

Further, statistics is but part of the processes involved in language understanding. The parallel ‘context’ of just who the statement refers to remains constant throughout the utterance, and so the listener or speaker is able to use this information in determining the number of the verb, thus selecting *is* or *are* as required.

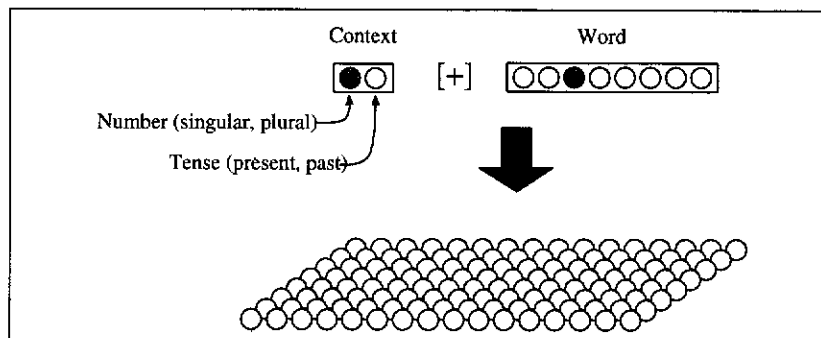


Figure B.40: Context and Embedded Sentences.

Figure B.40 illustrates an example in which the context of number (singular or plural) and tense (present or past) is maintained in another vector component to be mapped onto the initial SOM surface. This is not to imply that the context vector needs to include the exact grammatical constructs (number and tense)—these are arbitrary and human invented. All that is needed is some indication of the *object* or *event* that is the actual context—in the case of the sentence above, the (several) people who wish to rent the flat.

Most discussions of sentence embedding, both cognitivist (e.g. Chomsky), and connectionist (e.g. Elman) fall into the syntax-only category and do not include parallel context. But surely the person hearing or emitting an embedded sentence has available the context information regarding the actual object of the sentence. The appropriate number or tense has been learned previously in sentences *without* embedding—either implicitly as is suggested is the case in normal language learning, or explicitly if the grammatical ‘rules’ are known (having been taught at school). The process of sentence generation and understanding is not a simple serial one—many parallel processes are simultaneously active.

The use of parallel context was illustrated in Section 2.4.5. The use of context in embedded sentence understanding is left to future research. For now, we will limit ourselves to re-examining the experiments and claims made by Elman (1993) in relation to sentence embedding.

One finding made by Elman was that of the “importance of starting small”. In order to learn embedded sentences using an SRN, Elman found that he needed to begin training

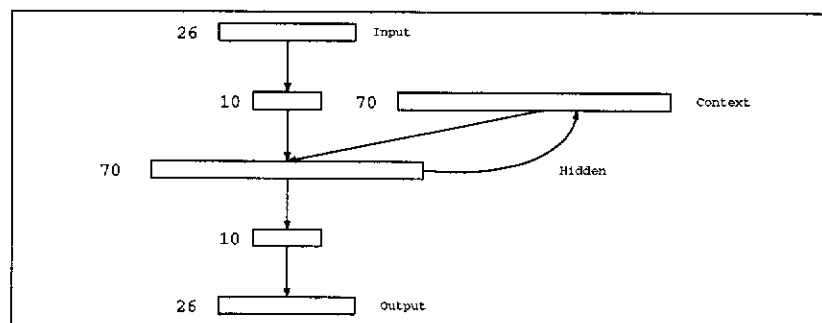


Figure B.41: Elman Embedded Sentence Network.

Phase	Number Of Sentences	Sentence Type
1	10,000	simple
2	7,500 2,500	simple complex
3	5,000 5,000	simple complex
4	2,500 7,500	simple complex
5	10,000	complex

Table B.7: Elman Sentence Build-up.

with sentences that were not embedded. After training on these simple sentences, he was able to build up the percentage of sentences involving embedding.

If the SRN network was given the full set of embedded sentences from the start of training, the network performed poorly, failing to predict the correct number agreement on some of the relative clauses. However, if a training regime such as that shown in Table B.7 was used, then the network was able to learn to correctly predict the number agreement. The SRN setup used by Elman is shown in Figure B.41.

In an attempt to duplicate the Elman experiments, a number of experiments were run with time-dependent data sets. For example, a number of randomly generated sentences with no embedding was used as the first data set, say for the first 50 epochs. This might be followed by a second set of randomly generated sentences which include at most one level of embedding, but with a ratio of say 75% of sentences without embedding and 25% of sentences with embedding. This second data set might be in force between say epochs 51 to 90, when a third set of data came into force. This set could include at most 2 levels of embedding, with relative percentages of 50%, 25% and 25% for the 0, 1 and 2 level cases. At epoch 120, the final set of sentences with relative percentages 17%, 50% and 33% might be used. †

Two experiments were conducted. The first involved a build up of complexity in em-

† The percentages 17%, 50% and 33% correspond to the number of unique sentences at each level of embedding using just one noun, one transitive verb and one intransitive verb. It thus equally samples the embedding sets.

bedding as described above, while the second began learning from the first epoch with the full 17%, 50% and 33% sentence mix.

The results obtained using the SOMA network did not duplicate Elman's results. There did not appear to be a need to build up the complexity of sentences over the learning epochs. We instead found that identical or similar results were obtained regardless of the training regime.

The same experiment was repeated with a 0%, 0%, 100% sentence mix with essentially the same results. This is shown in the following table. The table shows the final winning nodes, and the words which exit *from* each winning node—the 'context word'. Each word is represented primarily by one node, although some nodes have multiple context words (for example, node (16,16) covers both *boy* and *chase*). Further, the run included a refractory period (RP=1) which accounts for the extension of some words onto two nodes, especially in the case of the verbs which may appear in adjacent positions in the data. Because of time restrictions, only 100 random sentences were used. ³⁴

We suspect that the failure to learn without some form of build-up in Elman's experiment was due more to problems associated with the backpropagation algorithm than with any limitation of language learning by children. As Elman himself points out (Elman 1993, page 78): †

Although there is evidence that adults modify their language to some extent when interacting with children, it is not clear that these modifications affect the grammatical structure of the adult speech. Unlike the network, children hear exemplars of all aspects of the adult language from the beginning.

The sentences used by Elman consisted of several nouns and verbs. In order to appreciate the performance of the SOMA network on this task, we start with the simple case

†Elman also examined the option of not staging the input, but rather beginning with a limited memory and subsequently increasing the memory span. This was achieved by eliminating the recurrent feedback after every few words, and then increasing the memory window. Learning was essentially the same as the staged input method. We did not attempt to duplicate this incremental memory model here.

Win	Word Context					
Node	boy	boys	girl	girls	who	.
16,16	6					
6,10		6				
14,12			6			
22,22	1			6		
12,15					4	
23,23						8
6,14			1			
14,8						
6,19						

(a)

Win	Word Context							
Node	like	likes	chase	chases	smile	smiles	walk	walks
16,16			10				1	
6,10								1
14,12				1		5		
22,22			2					4
12,15								
23,23								
6,14				10	3			
14,8	9	2						
6,19	1	10					2	

(b)

Table B.8: Winning Node Word Counts.

of just one noun (boy), one transitive verb (chases), and one intransitive verb (smiles). In this case there are only 12 unique sentences. †

†Note that the data sentences are terminated with a full-stop character. This enables each sentence to be delineated.

boy chases boy .
 boy smiles .
 boy who boy chases chases boy .
 boy who boy chases smiles .
 boy who boy who boy chases chases chases boy .
 boy who boy who boy chases chases smiles .
 boy who boy who chases boy chases chases boy .
 boy who boy who chases boy chases smiles .
 boy who chases boy chases boy .
 boy who chases boy smiles .
 boy who smiles chases boy .
 boy who smiles smiles .

A SOMA network was trained over the standard epoch regime. The transitions learned by the network are shown in Figure B.42. Note that the SOMA algorithm used included the refractory period extension (RP=1) and hence the verbs *smiles* and *chases*, which can appear in pairs, are associated with two winning nodes.

The initial training data excluded the sentence *boy who smiles chases boy*. A subsequent run included it, and the only effect on the ‘flow diagram’ was to introduce another path as shown in Figure B.43. ³⁵

The paths through the flow diagram for the 12 unique sentences are shown in Fig-

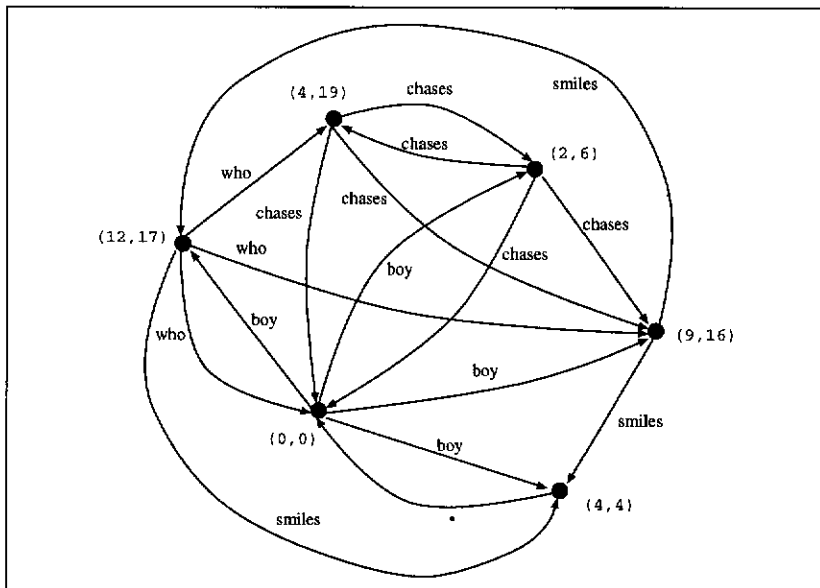


Figure B.42: English Embedded Sentences—Build-up with Restricted Components.

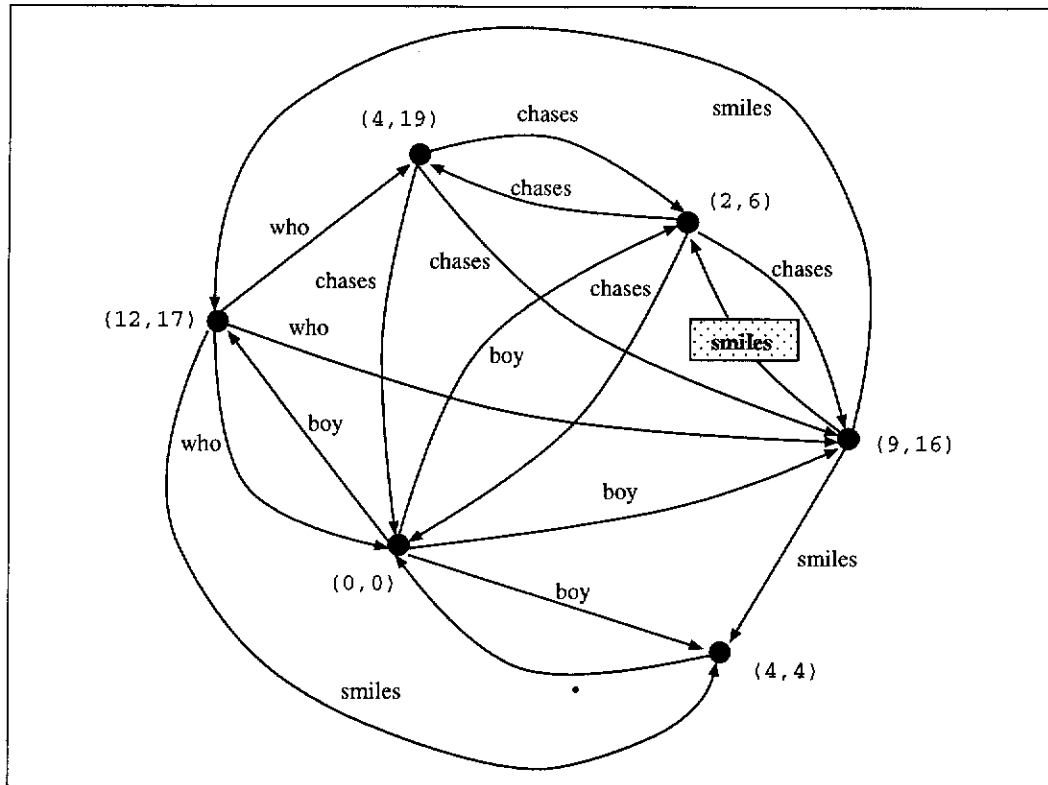


Figure B.43: English Embedded Sentences—All 12 Sentences.

ure B.44.

This experiment was performed both with a build-up of embedding (as described above) such as:

Epoch Number	% Sentences No Embedding	% Sentences 1 Level Embed	% Sentences 2 Level Embed	Number Of Sentences
1	100	0	0	50
50	75	25	0	100
90	50	25	25	150
120	17	50	33	300

as well as with no build up, running straight from the beginning epochs with a mix of embeddings such as:

Epoch Number	% Sentences No Embedding	% Sentences 1 Level Embed	% Sentences 2 Level Embed	Number Of Sentences
1	17	50	33	300

The two training regimes produced essentially the same results. The flow diagram without build up is shown in Figure B.45.

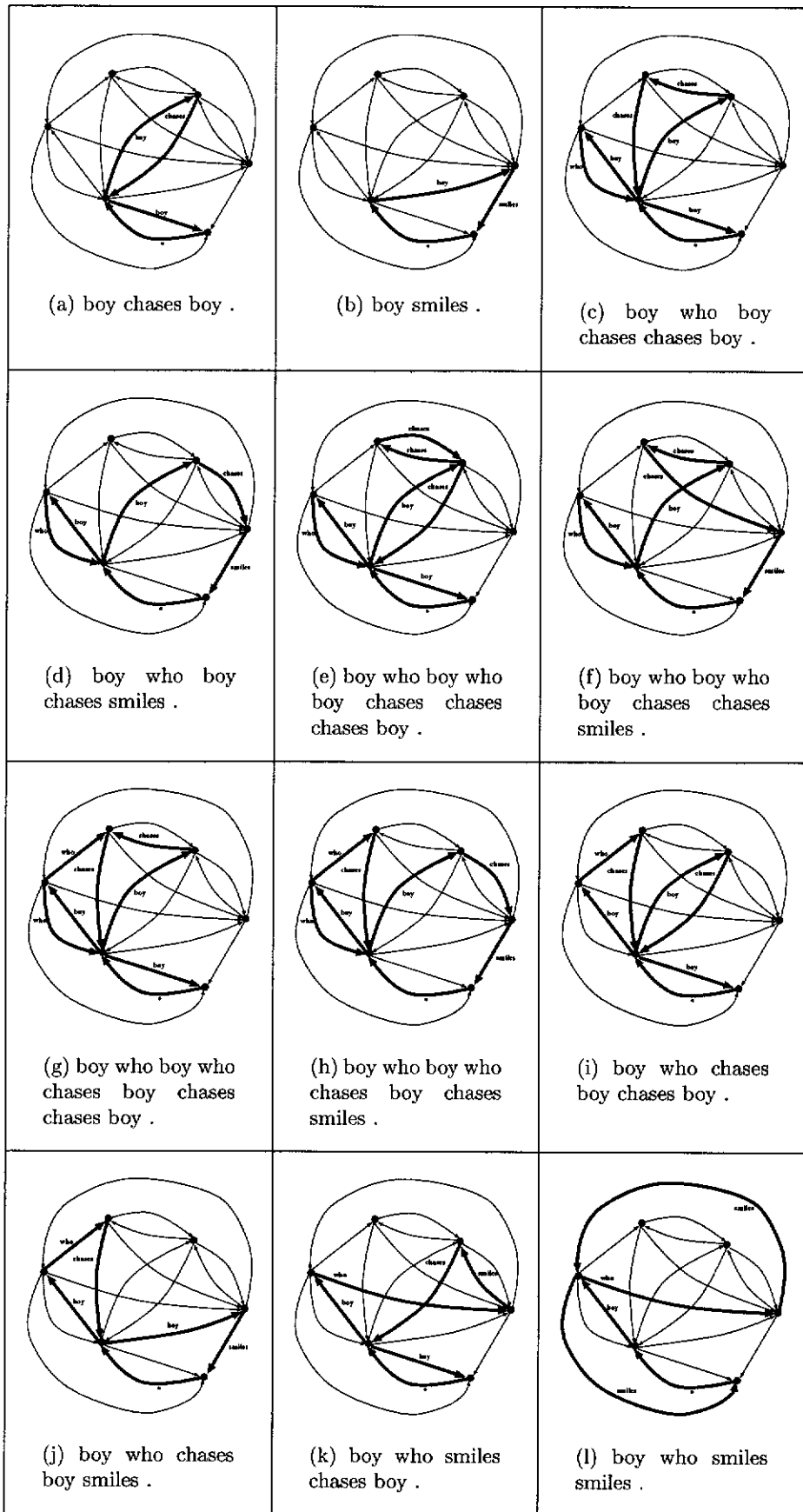


Figure B.44: English Embedded Sentences—Flow Paths for All 12 Sentences.

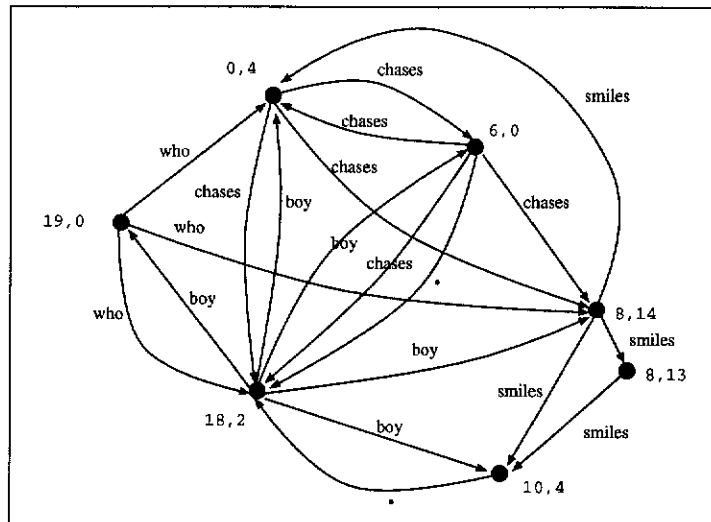


Figure B.45: English Embedded Sentences—No Build-up.

The experiment was repeated using both the singular and plural cases.³⁶ In this case there are 72 unique sentences. Training was conducted on the 68 sentences:

boy chases boy . boy chases boys . boy smiles . boy who boy chases chases boy . boy who boy chases chases boys . - boy who boy who boy chases chases chases boy . boy who boy who boy chases chases chases boys . boy who boy who boy chases chases smiles . boy who boy who boys chase chases chases boy . boy who boy who boys chase chases smiles . boy who boy who chases boy chases chases boy . boy who boy who chases boy chases chases boys . boy who boy who chases boy chases smiles . boy who boys chase chases boy . boy who boys chase chases boys . boy who boys chase smiles . boy who boys who boy chases chase chases boy . boy who boys who boy chases chase chases boys . - boy who boys who boys chase chase chases boy . boy who boys who boys chase chase chases boys . boy who boys who boys chase chase smiles . boy who boys who chase boys chase chases boy . boy who boys who chase boys chase chases boys . boy who boys who chase boys chase smiles . boy who chases boy chases boy . boy who chases boy chases boys . boy who chases boy smiles . boy who chases boys chases boy . boy who chases boys chases boys . boy who chases boys smiles . boy who smiles chases boy . boy who smiles chases boys . boy who smiles smiles .	boys chase boy . boys chase boys . boys smile . boys who boy chases chase boy . boys who boy chases chase boys . boys who boy chases smile . boys who boy who boy chases chases chase boy . boys who boy who boy chases chases chase boys . boys who boy who boy chases chases smile . boys who boy who boys chase chases chase boy . boys who boy who boys chase chases chase boys . boys who boy who boys chase chases smile . boys who boy who chases boy chases chase boy . boys who boy who chases boy chases chase boys . boys who boy who chases boy chases smile . boys who boys chase chase boy . boys who boys chase chase boys . boys who boys chase smile . boys who boys who boy chases chase chase boy . boys who boys who boy chases chase chase boys . - boys who boys who boys chase chase chase boy . - boys who boys who boys chase chase smile . boys who boys who chase boys chase chase boy . boys who boys who chase boys chase chase boys . boys who boys who chase boys chase smile . boys who chase boy chase boy . boys who chase boy chase boys . boys who chase boy smile . boys who chase boys chase boy . boys who chase boys chase boys . boys who chase boys smile . boys who smile chase boy . boys who smile chase boys . boys who smile smile .
--	--

The 4 missing sentences were:

boy who boy chases smiles . boy who boys who boy chases chase smiles . boys who boys who boy chases chase smile . boys who boys who boys chase chase chase boys .
--

The first trial run of this singular/plural example included only sentences *without* embedding. The result was most illuminating, as shown in Figure B.46. As all possible data examples were used in this test set, a symmetrical pattern was obtained.

The flow diagram for a single level of embedding is shown in Figure B.47. Here the data used excluded one sentence of level-one embedding and three of level-two embedding, and so the figure is no longer symmetrical.

Proceeding on to 2 levels of embedding only adds a further 6 links, while retaining the rest of the structure.

Adding further nouns and verbs increases the complexity of the flow diagrams, but still retains an overall simplicity. For example, with 4 nouns (boy-boys, girl-girls, dog-dogs, cat-cats), 4 transitive verbs (chases-chase, likes-like, hears-hear, sees-see), and

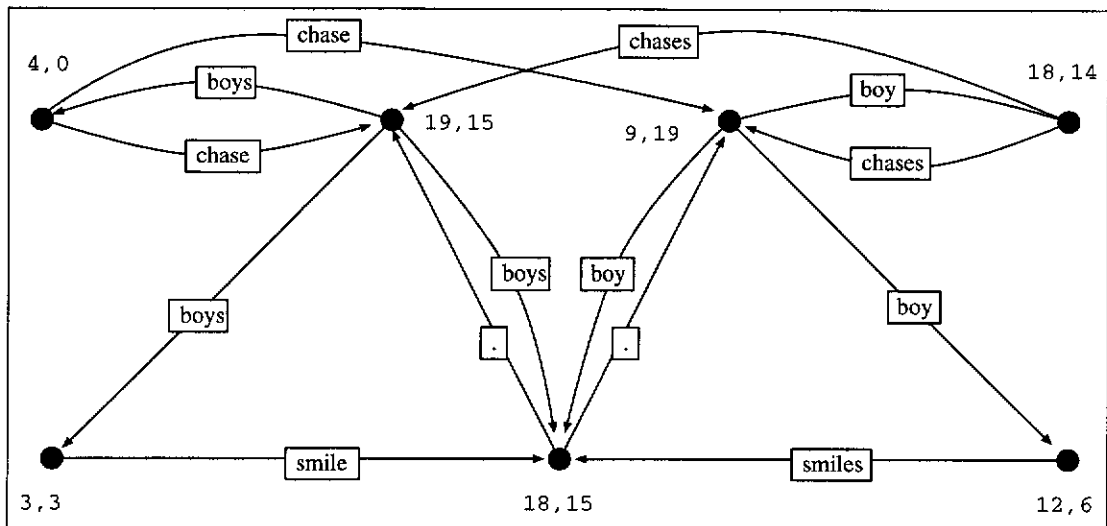


Figure B.46: English Embedded Sentences—Singular and Plural.

2 intransitive verbs (smiles-smile, walks-walk), the flow diagram for the case of both singular and plural forms, with no embedding, produced the flow diagrams as shown in Figures B.48 and B.49. In this case there are too many connections to draw all at once, so instead the individual word connections are shown. The full flow diagram is the combination of all of these connections.³⁷

The data here consisted of 200 randomly generated sentences (of which 135 were unique) out of a theoretical total of 272 possible unique sentences.

A further run with 4 nouns, 4 transitive and 2 intransitive verbs, with up to two levels of embedding (but using a somewhat reduced data set due to processing time restrictions[†]), produced a very similar structure.³⁸

What can we say about these flow diagrams?

The first point to make is that the second SOM surface of the SOMA network self-organizes to a small set of winning nodes. In most of the trials discussed above, a 20 by 20 surface was used, giving a total of 400 nodes. However, only 8 winning nodes were used, for example, in Figure B.47. These nodes correspond to the 8 concepts (words)

[†]This particular run took an elapsed time of approximately two weeks to complete, and so represents the extreme limit with current equipment and algorithm implementations. The program may be able to be optimized to some extent, but this was not attempted.

used in the data sentences.

This means that the ‘semantics’ are kept local, as shown in Figures B.50 and B.51. A ‘concept’ (word) is associated with a transition from particular node (or nodes) on the second SOM surface, to another node on that surface. For example, in Figure B.50 (c), the word **boy** is associated with the node (9,19). All transitions which correspond to the word **boy** in a sentence occur from this node. The following word determines the actual path used on exit from node (9,19).

Thus each words (or verbal concept) is located in one area,[†] with links to other concept areas.

Links to words which are found to follow a particular word in a sentence are enhanced, whereas the links to words which are never or rarely adjacent in a sentences are not.

The concentration of word transitions at a single node (or a few nodes) has implications if we add some local context at the second SOM layer inputs. In Section 2.4.5 we examined the effect of additional parallel context when located at inputs to the first SOM layer. Here we consider the addition of parallel context to the inputs of the second SOM layer as indicated in Figure B.52.

The addition of context at this layer will *bias* the selection of one of the alternate following links. As was shown in Section 2.4.5, the addition of extra parallel context increases the number of possible winning nodes for each word sequence recorded on the surface.

Different context components, with identical word and sequence components, will tend to bias the selection of links to subsequent winning nodes. This is illustrated in Figure B.53, where the context components have been put within quotes to indicate that the exact content of the context is unspecified at this stage. It need not result from symbolic sources, but may result from association links to other modalities and sub-modalities.

[†]The run which generated the links of Figures B.50 and B.51 used a refractive period of two, so some words, (the verbs), which may occur as successive words in a sentence, are associated with two nodes. However, one node appears to take a major share of the transitions for the word.

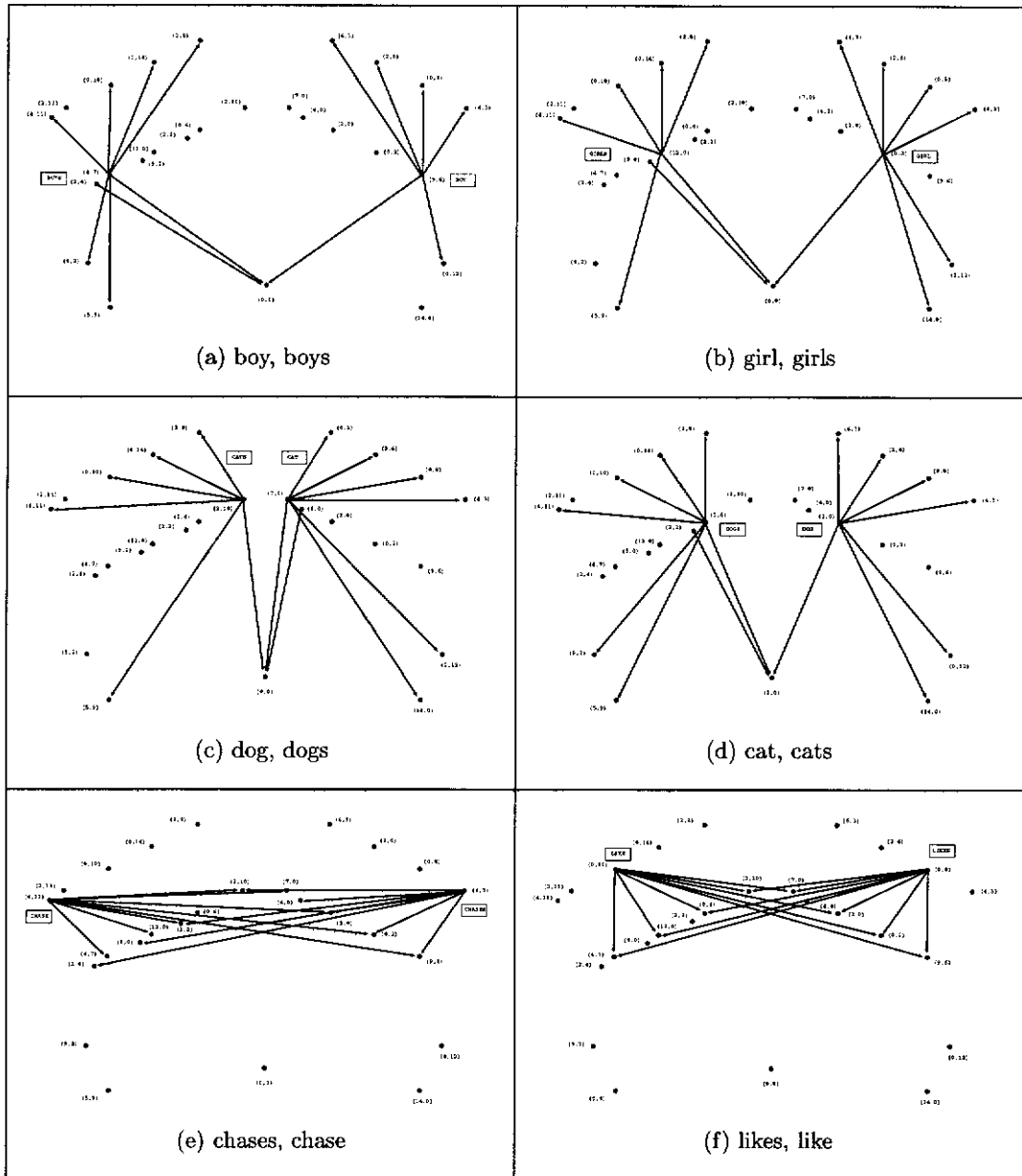


Figure B.48: English Embedded Sentences—4 Nouns, 4 Transitive and 2 Intransitive Verbs.

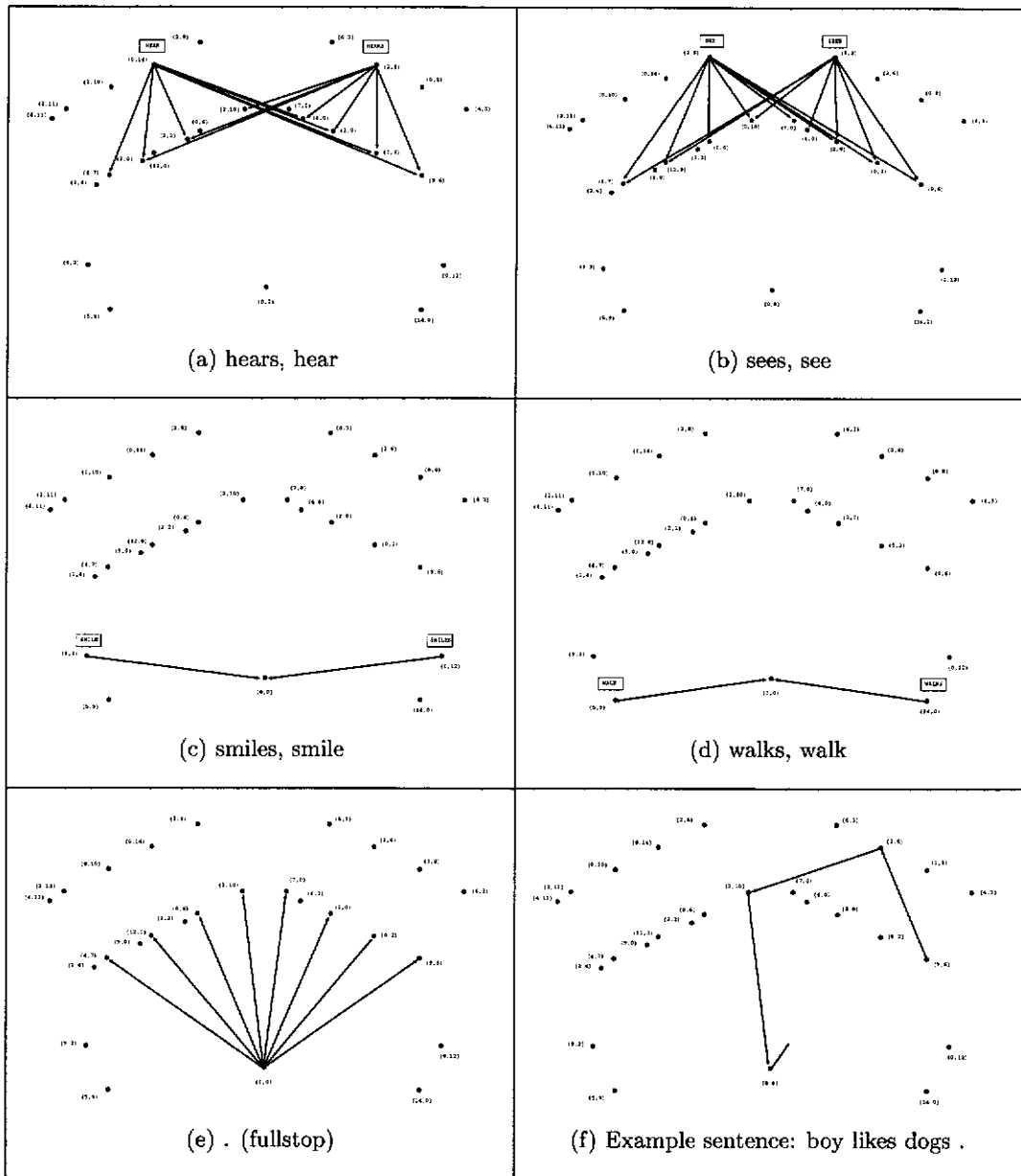


Figure B.49: English Embedded Sentences—4 Nouns, 4 Transitive and 2 Intransitive Verbs (cont).

Win	Word Context									
Node	boy	boys	girl	girls	dog	dogs	cat	cats	who	.
10,26	o									
10,12		o								
2,19			o							
15,16				o						
1,25					o					
0,17						o				
29,7							o			
18,0								o		
8,5									o	
27,11										o

Win	Word Context							
Node	like	likes	chase	chases	hear	hears	see	sees
21,23	o							
22,21	•							
22,3		o						
23,16			o					
22,16			•					
4,23				o				
4,24				•				
24,0					o			
26,0					•			
13,22						o		
13,19						•		
14,28							o	
13,28							•	
6,28								o
5,28								•

Win	Word Context			
Node	smile	smiles	walk	walks
24,23	o			
24,24	•			
9,28		o		
25,4			o	
3,9				o
4,9				•

Table B.9: Word Winning Nodes—4 noun, 4 transitive verbs, 2 intransitive verbs.

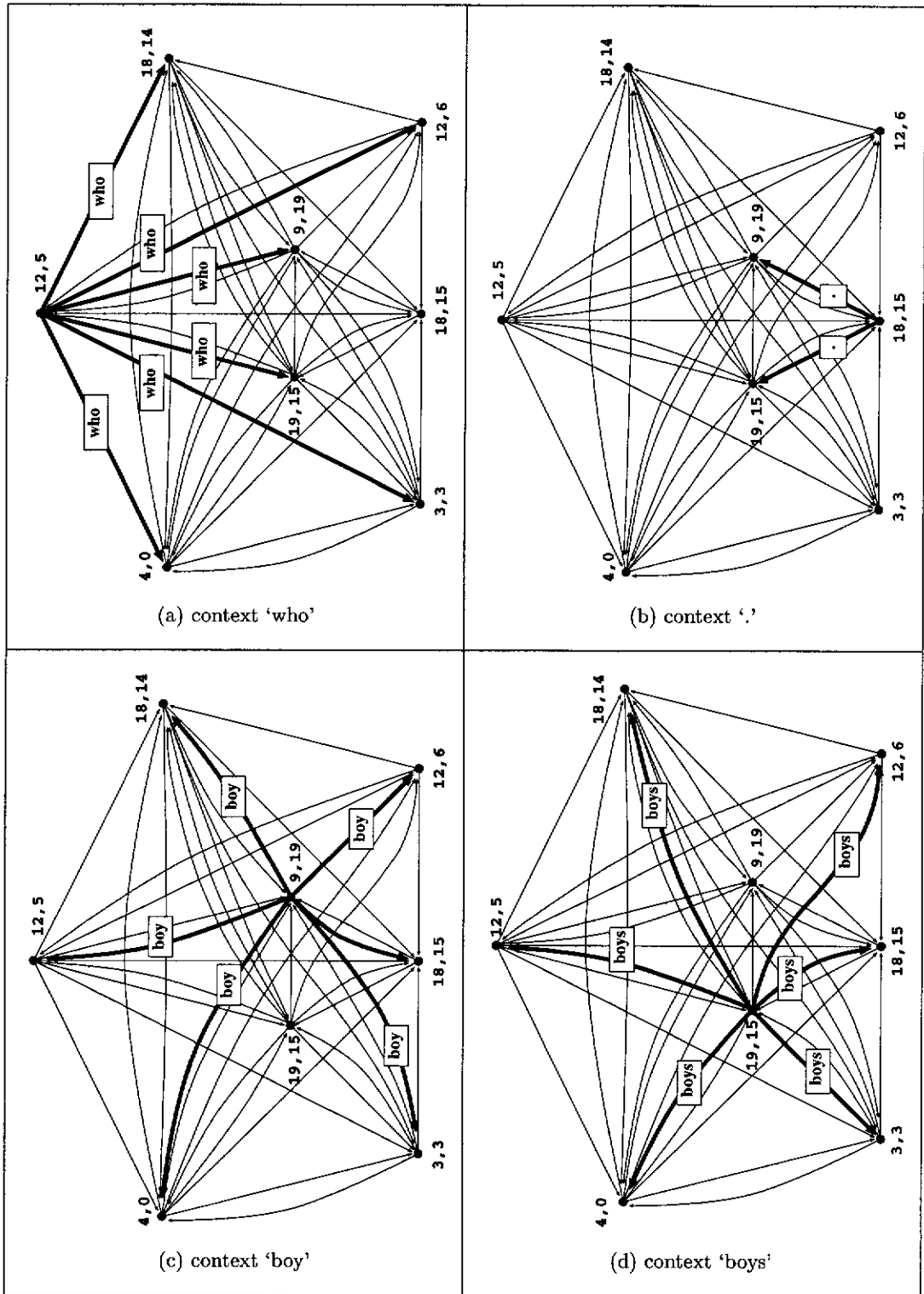


Figure B.50: English Embedded Sentences—'Semantics'.

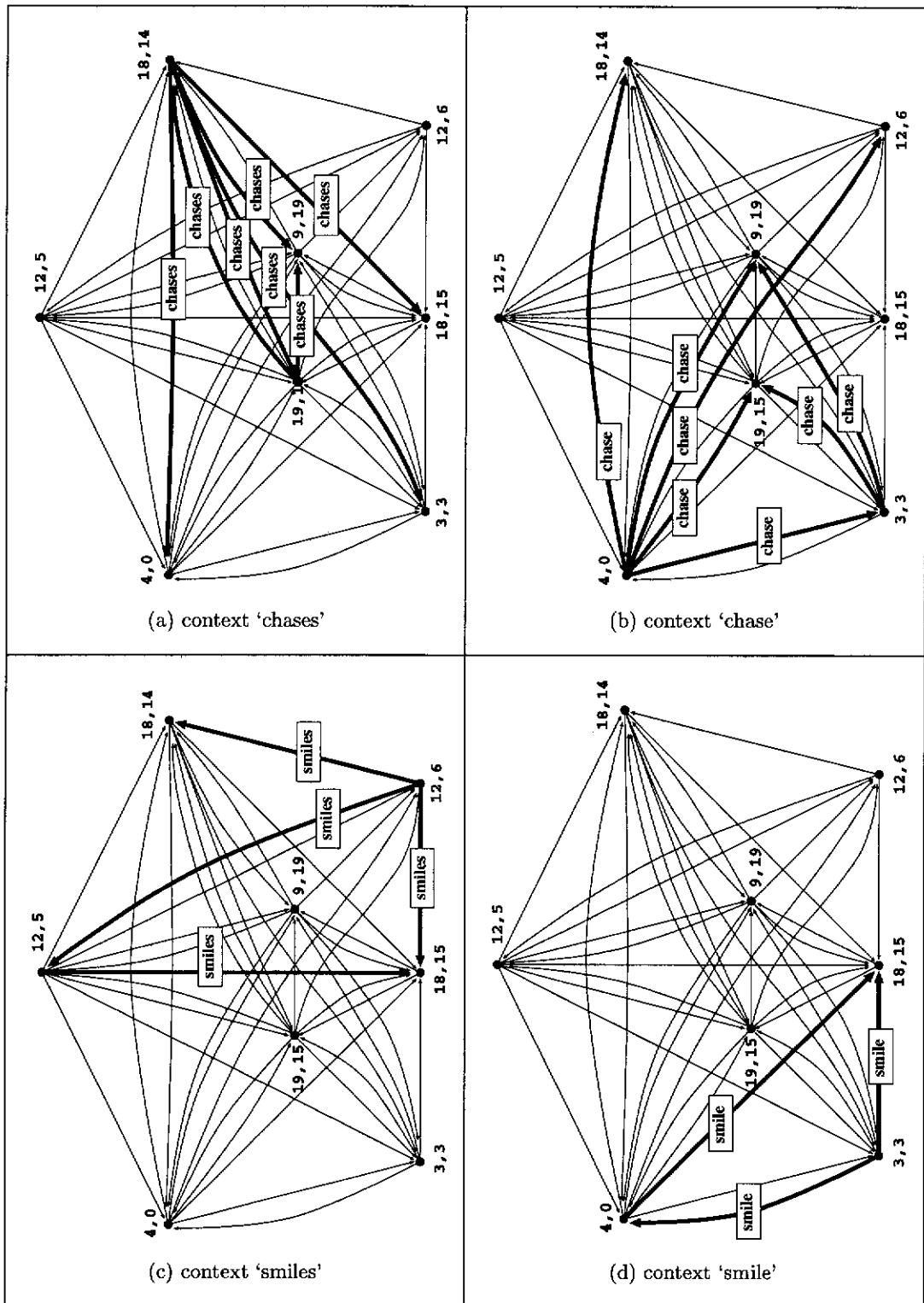


Figure B.51: English Embedded Sentences—'Semantics' (con't).

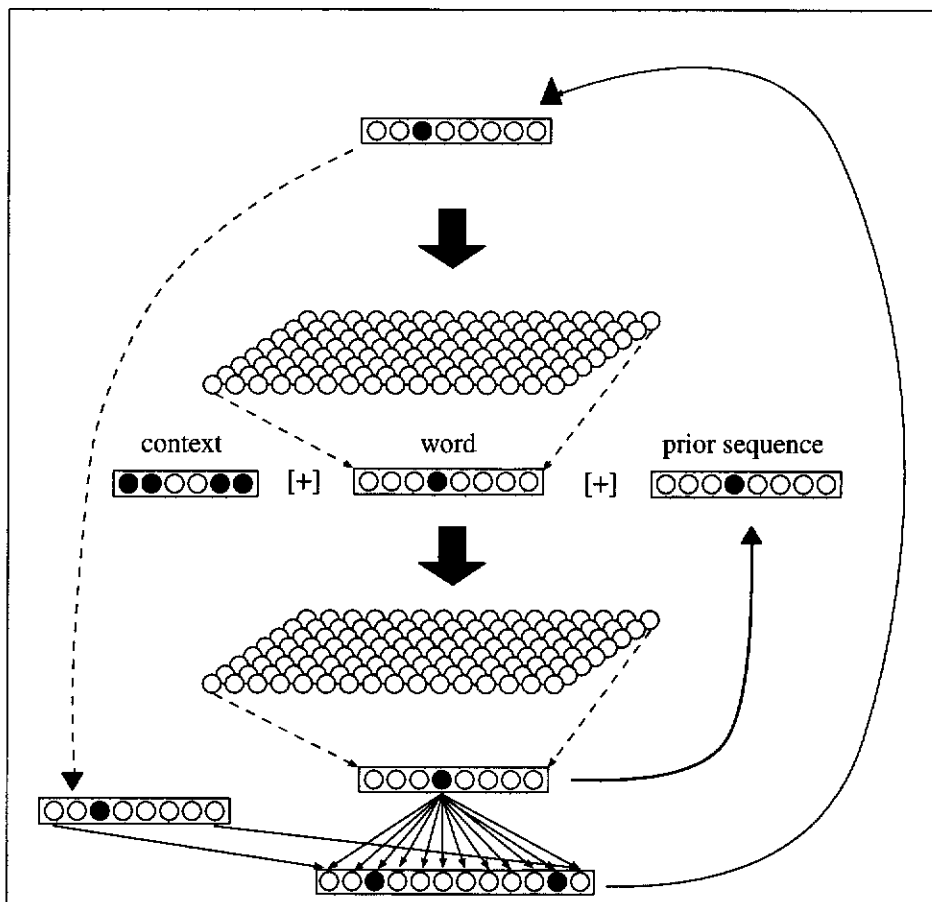


Figure B.52: Semantic Priming on Transition SOM.

The implications of adding contextual input to the second SOM layer were not tested experimentally here, but will be left to future research.

Adding in further nouns and verbs quickly increases the number of possible linkages, as shown in the following table:

Num. Cases	Num. Nouns	Num. Trans. Verbs	Num. Intrans. Verbs	Sentences With No Embedding	Sentences 1 Level Embedding	Sentences 2 Levels Embedding
1	1	1	1	2	6	4
1	2	2	2	12	120	384
1	3	3	2	33	660	5346
1	4	4	2	72	2448	36864
2	1	1	1	6	30	36
2	2	2	2	40	720	3840
2	3	3	2	120	4560	58320
2	4	4	2	272	17952	417792

For example, with 2 nouns (boy, boys, girl, girls), 2 transitive verbs (chases, chase, likes, like) and 2 intransitive verbs (smiles, smile, likes, like) there are 40 unique sentences (as derived from the FSTN of Figure B.39) with no embedding, 720 with 1 level of embedding, and 3,840 with 2 levels of embedding. †

The number of possible sentences increases dramatically, but the number of winning nodes on the SOM surface only increases (approximately) linearly with the total number of concepts, that is, $c \times (n + vt + vi)$.

†The actual calculations are:

$$N_0 = c \times n \times (vi + c \times n \times vt),$$

$$N_1 = c \times n \times (vi + c \times n \times vt) \times (vi + 2 \times c \times n \times vt), \text{ and}$$

$N_2 = c^2 \times n^2 \times vt \times (c \times n \times vt + vt \times n) \times (vi + c \times n \times vt)$, where N_0 , N_1 and N_2 are the number of sentences with 0, 1 and 2 levels of embedding (exclusively), c is the number of cases, n the number of nouns, vi the number of intransitive verbs and vt the number of transitive verbs.

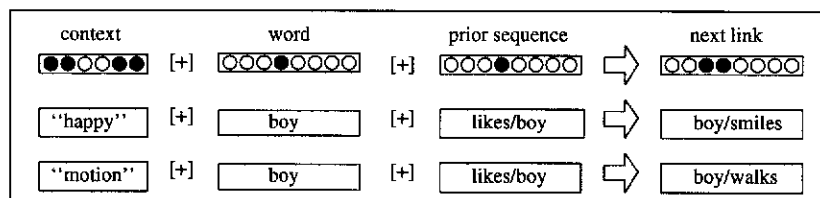


Figure B.53: Context Biased Recurrent Links.

The FSTN of Figure B.39 is deficient in at least two ways: first, there is no allowance for an intransitive verb following state 6, and second, the noun following state 8 can be generalized to be either singular or plural in number. A further experiment added these components to the FSTN (and hence to the test data), and repeated the run using 2 nouns, 2 intransitive verbs, and 2 transitive verbs.

The result obtained was similar to those found previously. In other words, the addition of the new ‘grammatical’ features was able to be incorporated into a similar structure, with particular nodes on the second SOM layer being responsible for certain word transitions.

B.3 Hierarchical Temporal Sequences

In this section we examine the combination of a number of these temporal learning structures (FSAMap, SOMA, LAPS), organized into some form of hierarchy. In order to look at these temporal learning hierarchical structures, let’s first examine the simplest case of hierarchical FSAMaps.

B.3.1 Hierarchical FSAMaps

The current *single-level* FSAMap arrangement is inflexible. If the system is in *state1* say, and it receives an input *input1*, then it will always go to the same state, *state2*, as shown in Figure B.54.

This is not always what is required—we may sometimes want the system to proceed to alternate states depending on some other contextual information, as indicated in Figure B.55. This alternate requirement cannot be met with a single layered FSAMap.

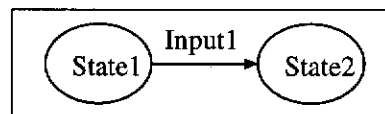


Figure B.54: Fixed Transitions in FSA.

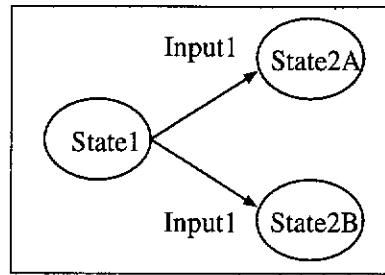


Figure B.55: Variable Transitions in FSA.

The next state chosen may need to be determined not only by the input, but also by some other criteria, such as the state of another FSAMap. For example, if we seek to develop a system to parse the c-like computer language statement

```
for ( expression ; expression ; expression ) statement ;
```

then using a single FSAMap we would need to have three separate parse state transitions, one for each of the three `expression` parts, as is shown in Figure B.56. In this figure the highlighted `expression` states are in reality multiple states that will need to be duplicated three times. What is required is a mechanism whereby only one `expression` state is required, and other contextual information determines the behaviour *following* the parse of each expression.

We can achieve this aim if we introduce a second FSAMap state machine as indicated in Figure B.57.

State transitions on FSAMap 1 are then of the form shown in Figure B.58. The behaviour of FSAMap 1 will depend on the different states of FSAMap 2, as well as the inputs and states of FSAMap 1.

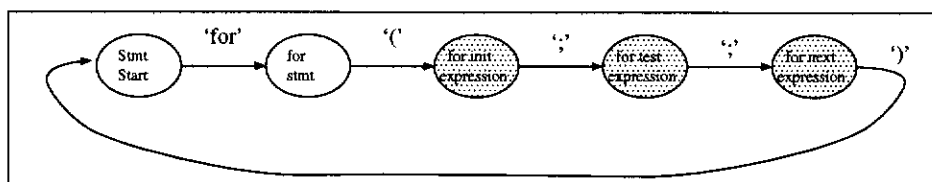


Figure B.56: Parsing a 'for' Statement—One FSAMap.

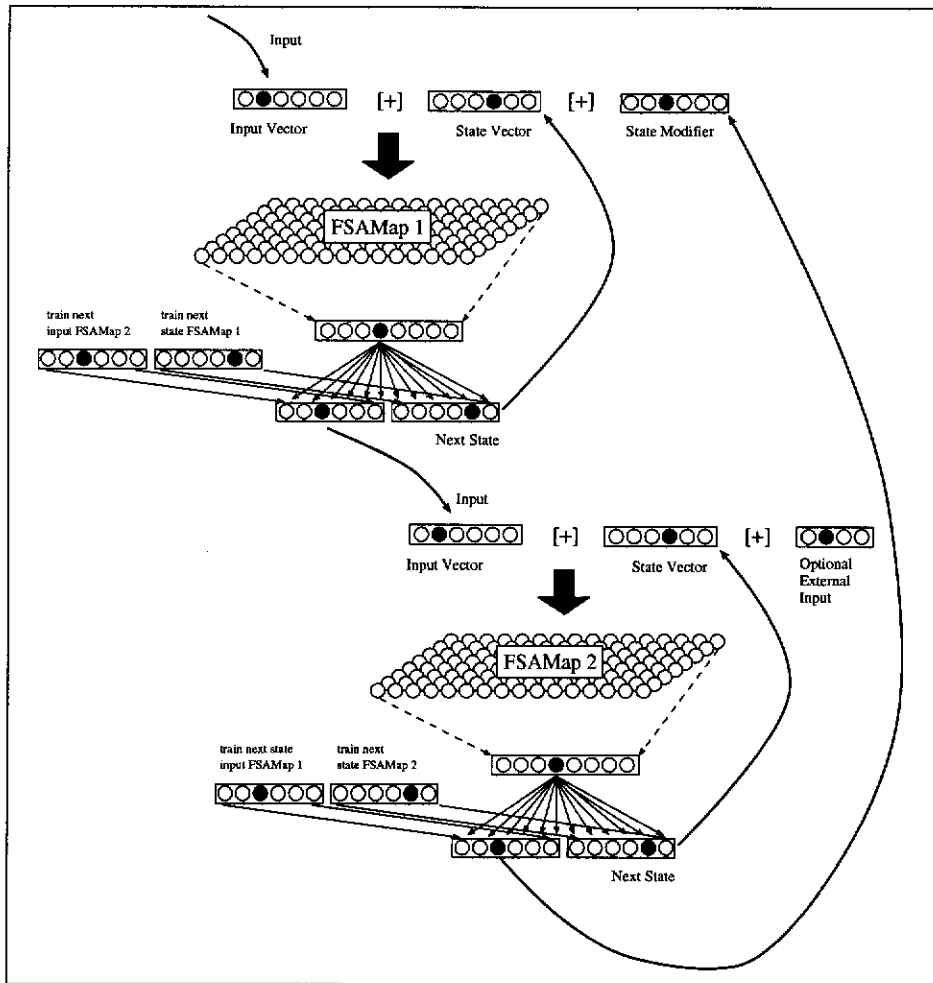


Figure B.57: Hierarchical FSAMaps.

Figure B.57 shows an optional external input vector for FSAMap 2. If this is used, then the two FSAMaps are symmetrical. If we redraw the FSAMaps as black boxes, hiding the internal details and just indicating that each FSAMap can have two inputs and one output (that is, given 2 inputs, the FSAMap will move to a new state and output a vector), then we can see this symmetry more clearly in Figure B.59. The multiple input mechanism allows us to connect a number of FSAMaps together, and the number of inputs per FSAMap may be extended to further increase the possibilities.

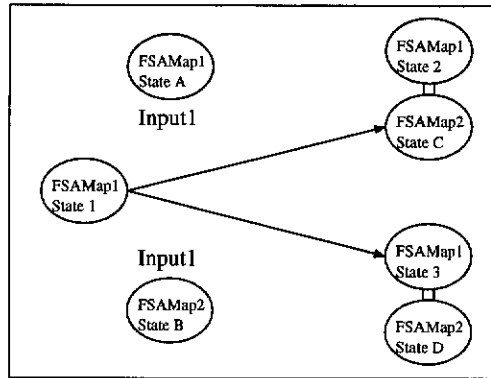


Figure B.58: Variable State Transitions.

B.3.1.1 Multiple FSAMaps—Context-Sensitive Grammar Example

The Chomsky hierarchy of languages³⁹ includes the *context-sensitive* grammar (CSG). An example of a context-sensitive grammar that is often cited in the computer science literature in relation to CSGs is $a^n b^n c^n$ (see for example Tremblay & Sorenson 1985, page 38). Obviously any parser of this grammar must be able to examine the context—the number of *a* characters must be counted in some form in order to be able to parse the *b* and *c* components.⁴⁰ It is one of the claims of this thesis, however, that this categorisation of grammars is not relevant to actual human languages, but only to serially-parsed streams of strings. It does not take into account a number of important issues relevant to human languages, such as meaning and global context over-and-above

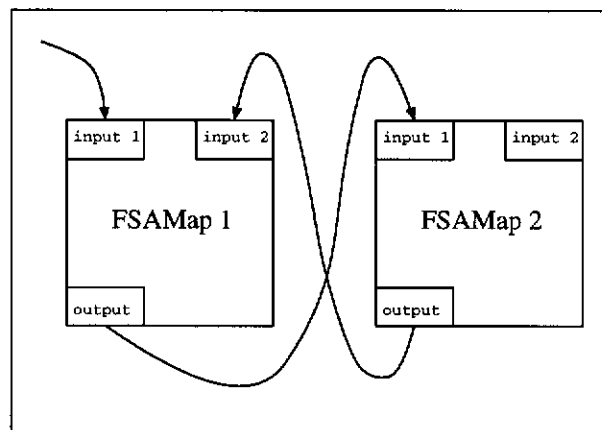


Figure B.59: Linked FSAMaps.

the actual strings themselves.

Given this objection, it is perhaps still relevant to see if the temporal learning architectures discussed here *are* capable of learning these particular types of string grammars.

To implement the grammar $a^n b^n c^n$ we need two counters and a controlling state machine as shown in Figure B.60.

Each counter is a simple device that learns to count up to say 5, then back down to zero. To achieve this, we will need 6 number states (**state0**, **state1**, . . . , **state5**) as well as an error state to indicate underflow, overflow or mismatched indices (**error**). Note that we could differentiate these errors and take separate actions depending upon the type of error, but this distinction is not relevant here.

As well as the inputs 0 , 1 and -1 , we will also need inputs *reset* (which resets the current state back to **state0**), *init* (which initialises the current state to **state1**), and *check* (which checks that the count is zero; if yes, it performs the same action as *init*; if no, it starts the error routine). The state transition table for the counter is shown in Table B.10.

The initial state of the Count FSAMap is set to **state0** and the initial output value **ok**.

The outputs for each counter (**ok** and **err**) are fed back to the main Control FSAMap. The state transition table for this FSAMap is shown in Table B.11 for the non-error

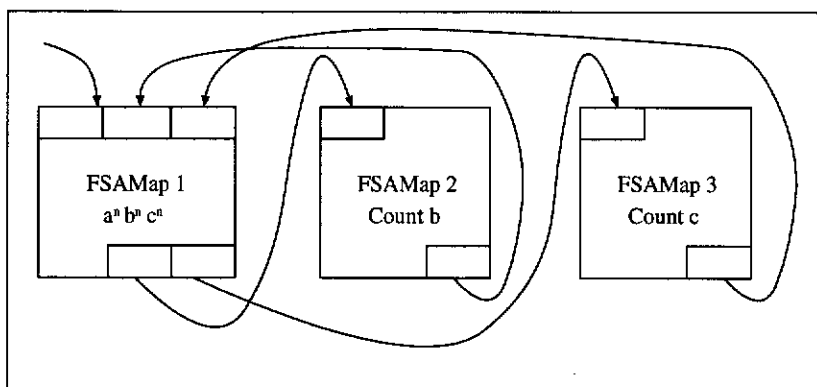


Figure B.60: Context Sensitive Grammar $a^n b^n c^n$ —Three FSA Maps.

Current State	Input	Next State	Output
$state_i$ $i=1,5$	0	$state_i$	ok
$state_i$ $i=0,4$	1	$state_{i+1}$	ok
$state_5$	1	error	err
$state_0$	-1	error	err
$state_i$ $i=1,5$	-1	$state_{i-1}$	ok
$state_i$ $i=0,5$	init	$state_1$	ok
$state_i$ $i=0,5$	reset	$state_0$	ok
$state_0$	check	$state_1$	ok
$state_i$ $i=1,5$	check	error	err
error	0	error	err
error	1	error	err
error	-1	error	err
error	init	$state_1$	ok
error	reset	$state_0$	ok
error	check	error	err

Table B.10: Count FSA—State Transition Table.

flow, and Table B.12 for the error handling state transitions. Note that the ‘-’ indicates the three combinations of at least one error input from the Count FSAs; that is, (ok err), (err ok), or (err err).

Each FSAMap was trained individually on the state transitions and output vectors appropriate to it. Obviously, as the Counter FSAMap is used twice in the multi-level structure, the same learned weights can be used for both. Once both the Control and Counter FSAMaps were learned, their weights were used in another program which

Current State	Input	Input Count b	Input Count c	Next State	Output Count b	Output Count c
start	a	ok	ok	state-a	init	init
state-a	a	ok	ok	state-a	1	1
state-a	b	ok	ok	state-b	-1	0
state-b	b	ok	ok	state-b	-1	0
state-b	c	ok	ok	state-c	check	-1
state-c	c	ok	ok	state-c	reset	-1
state-c	a	ok	ok	state-a	init	check

Table B.11: Control $a^n b^n c^n$ FSA—State Transition Table.

Current State	Input	Input Count b	Input Count c	Next State	Output Count b	Output Count c
start	b	ok	ok	error	reset	reset
start	b	-	-	error	reset	reset
start	c	ok	ok	error	reset	reset
start	c	-	-	error	reset	reset
state-a	a	-	-	error	reset	reset
state-a	b	-	-	error	reset	reset
state-a	c	ok	ok	error	reset	reset
state-a	c	-	-	error	reset	reset
state-b	a	ok	ok	error	reset	reset
state-b	a	-	-	error	reset	reset
state-b	b	-	-	error	reset	reset
state-b	c	-	-	error	reset	reset
state-c	a	-	-	error	reset	reset
state-c	b	ok	ok	error	reset	reset
state-c	b	-	-	error	reset	reset
state-c	c	-	-	error	reset	reset
error	a	ok	ok	state-a	init	init
error	a	-	-	state-a	init	init
error	b	ok	ok	error	reset	reset
error	b	-	-	error	reset	reset
error	c	ok	ok	error	reset	reset
error	c	-	-	error	reset	reset

Table B.12: Control $a^n b^n c^n$ FSTN—Error Handling.

controlled the co-ordination of the three FSAMaps. Input to this program was of the form

```
start st0 st0 a a a b b b c c c a b c a a b b c c a a a b b c c c a a b b c c
```

The first three inputs are the initial states of each of the three components of the combined multi-level structure. [†]

Subsequent entries are input characters which are passed to the Control FSAMap at each time step. These inputs, along with the current outputs from the two Counter FSAMaps, are processed at each time step to evaluate the next state, and the two

[†]As well, the outputs for each FSAMap vectors are initialised within the multi-level program to be *init* for both outputs of the Control FSAMap, and *ok* for each of the Counter FSAMaps. This initialisation could also have been supplied as input data, but was done as part of the program initialisation.

output values appropriate to the Control FSAMap. At the *same time step*, each of the outputs of the Control FSAMap are used as inputs to one of the counters, and along with the current state of the particular Counter FSAMap, is used to also determine the next state and the appropriate output of these FSAMaps. Each FSAMap is thus processed *synchronously*.

The multi-level FSAMap structure so described was able to correctly parse the input data. The above input data includes strings of $a^n b^n c^n$ for $n = 3, 1$ and 2 respectively. These were parsed correctly and the appropriate states of all of the FSAMaps were encountered.⁴¹

The next data string is the invalid sequence (a a a b b c c c) and the combined system correctly determined that there was a missing b following the input of the first c of the sequence. The **check** input passed to the **b** Counter FSAMap causes an error exit to occur. The subsequent c characters of the invalid string are absorbed via the error processing, and the system is able to recover following the input of the first a of the next sequence. The final string sequence is then parsed correctly.

B.3.1.2 Multiple FSAMaps—Computer Language Example

Returning to the c-like language example, we can now devise a state transition diagram for the parsing of the language.

The context component for the **for** statement is shown diagrammatically in Figure B.61. The input component for each state transition comprises two parts; the actual input, which is shown first, and the state from a second FSAMap, which is shown in square brackets. The contextual input from the second FSAMap provides a way of making alternate state transitions at the completion of each of the three expression parses.

The parsing of other statements of the language also uses the state of the second FSAMap to determine the required state transitions.

Say we have a c-like language defined by the description shown in Table B.13. This language is very simple, and probably not very useful. But it does illustrate a number

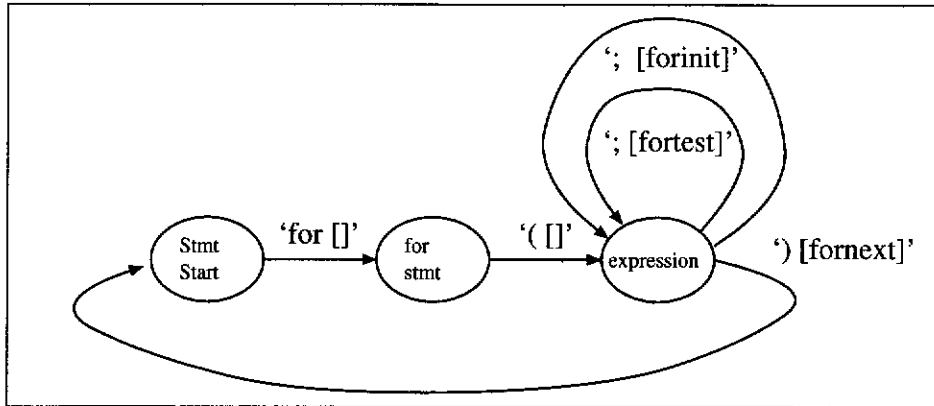


Figure B.61: Parsing a 'for' Statement—two FSA Maps.

<i>statements</i>	:=	<i>statement statements</i>
		<i>statement</i>
<i>statement</i>	:=	<i>read list ; ;</i>
		<i>print list ; ;</i>
		<i>for (expression ; expression ; expression) ; ; statement ; ;</i>
		<i>expression ; ;</i>
		<i>if (expression) ; ; statement ; ;</i>
		<i>while (expression) ; ; statement ; ;</i>
		<i>{ statements }</i>
<i>expression</i>	:=	<i>variable logicalop expraval</i>
<i>expraval</i>	:=	<i>term</i>
		<i>term arithop expraval</i>
		<i>(expraval)</i>
<i>list</i>	:=	<i>variable</i>
		<i>variable , list</i>
<i>term</i>	:=	<i>variable</i>
		<i>number</i>
<i>variable</i>	:=	<i>a b c</i>
<i>number</i>	:=	<i>0 1 2 3 4</i>
		<i>5 6 7 8 9</i>
<i>logicalop</i>	:=	<i>= == < <=</i>
		<i>> >= !=</i>
		<i>++</i>
<i>arithop</i>	:=	<i>+ - * /</i>

Table B.13: C-like Language Description.

of points concerning the use of multiple FSAMaps. Some example statements in the language are shown in Table B.14.

Note the extra semi-colons in the statements that are not found in the standard c-language known and loved by computer programmers. These semi-colons have been included to enable a better synchronization between the two FSAMaps, and reduce the number of the state transitions in each. Admittedly this is somewhat arbitrary, but the point of this exercise is not to define a suitable language for parsing by FSAMaps, but rather to illustrate the possibility of hierarchical, recurrent maps cooperating to learn somewhat difficult and context-dependent languages.

The language described in Table B.13 may also be partially described by the FSTM of Figure B.62.

We will only look at parsing the statements in a very simple manner. A number of extensions could be made; for example, a count of the level of nesting (resulting from the { and } statement block brackets) could be maintained with another FSAMap counter. In this FSAMap, an input of '{' would increment a count (state 0 \Rightarrow 1, 1 \Rightarrow

```
read a , b ;;

for( a = 1 ; a < b ; a ++ ) ;;
{
  c = a + b ;;
}

if( c >= 3 ) ;;
  print a , c ;;

while( c < a + b ) ;;
{
  c = 1 + a ;;
  if( a == c * 2 - b / 3 ) ;;
  {
    c ++ ;;
  }
}
```

Table B.14: Example Code.

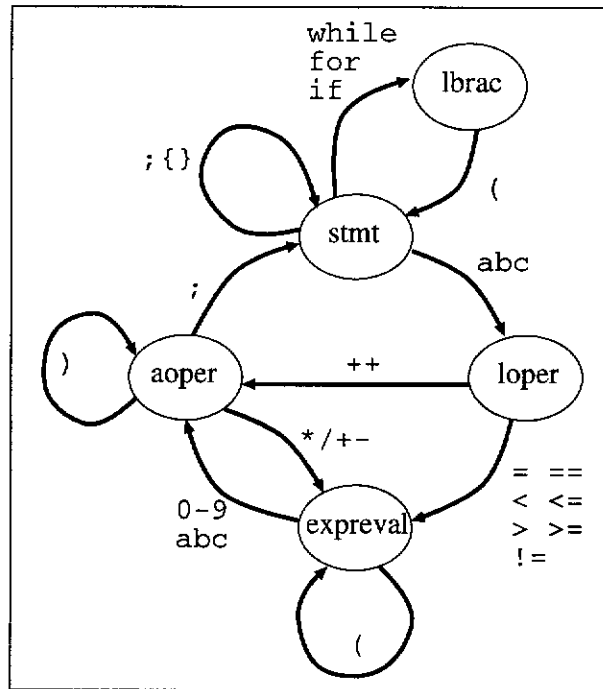


Figure B.62: Partial FSTN for c-language.

2, etc) whereas an input of ‘}’ would decrement the count (state 1 ⇒ 0, 2 ⇒ 1, etc). This expansion was not included here in order to make the problem relatively simple.

The full set of state transitions for the language is shown in the following tables.

Current FSAMap 1 State	External Input	Input From FSAMap 2	Next FSAMap 1 State	Output To FSAMap 2
stmt	print	stmt	list	list
stmt	read	stmt	list	list
list	a	list	comma	list
list	b	list	comma	list
list	c	list	comma	list
list	a	stmt	comma	list
list	b	stmt	comma	list
list	c	stmt	comma	list
comma	,	list	list	list
comma	;	list	stmt	stmt
stmt	;	list	stmt	stmt

Table B.15: State Transitions—Parse *print* and *read* Statements.

Current FSAMap 1 State	External Input	Input From FSAMap 2	Next FSAMap 1 State	Output To FSAMap 2
stmt	a	stmt	loper	expr
stmt	b	stmt	loper	expr
stmt	c	stmt	loper	expr
stmt	a	expr	loper	expr
stmt	b	expr	loper	expr
stmt	c	expr	loper	expr
stmt	;	expr	stmt	stmt
loper	=	stmt	expreval	expr
loper	==	stmt	expreval	expr
loper	<	stmt	expreval	expr
loper	<=	stmt	expreval	expr
loper	>	stmt	expreval	expr
loper	>=	stmt	expreval	expr
loper	!=	stmt	expreval	expr
loper	++	stmt	aoper	expr
expreval	a	expr	aoper	expr
expreval	b	expr	aoper	expr
expreval	c	expr	aoper	expr
expreval	0	expr	aoper	expr
expreval	1	expr	aoper	expr
expreval	2	expr	aoper	expr
expreval	3	expr	aoper	expr
expreval	4	expr	aoper	expr
expreval	5	expr	aoper	expr
expreval	6	expr	aoper	expr
expreval	7	expr	aoper	expr
expreval	8	expr	aoper	expr
expreval	9	expr	aoper	expr
expreval	(expr	expreval	expr
aoper	+	expr	expreval	expr
aoper	-	expr	expreval	expr
aoper	*	expr	expreval	expr
aoper	/	expr	expreval	expr
aoper)	expr	aoper	expr
aoper	;	expr	stmt	stmt

Table B.16: State Transitions—Parse *expression*.

The state transition tables for each statement were developed and tested separately and incrementally. For example, the *print* and *read* statements are relatively separate from the other statements, and so were developed first. Other statements depended

Current FSAMap 1 State	External Input	Input From FSAMap 2	Next FSAMap 1 State	Output To FSAMap 2
stmt	for	stmt	lbrac	forexpr
stmt	if	stmt	lbrac	ifexpr
stmt	while	stmt	lbrac	wexpr
lbrac	(stmt	stmt	same

Table B.17: State Transitions—Parse *for*, *if* and *while* Statements.

upon the *expression* statement, and so this was developed next, along with a number of additional transitions needed because of the interactions between the *print* and *read* statements, and the *expression* statement.

Once each separate component was working, the full set of state transitions were learned on one FSAMap.

The system worked as required.⁴² For example, the code of Table B.14 was supplied to the system, with the results shown in Tables B.20, B.21, B.22, and B.23. Error exits were not included in this example.

A point that could be made is that the c-like language is based on computer languages that were designed to be parsed by a compiler on a *serial* computer. A more appropriate language may be able to be developed that would better suit the state transition table environment of the FSAMap architecture. This will need to be left to future research.

Current FSAMap 1 State	External Input	Input From FSAMap 2	Next FSAMap 1 State	Output To FSAMap 2
stmt	{	stmt	stmt	stmt
stmt	}	stmt	stmt	stmt

Table B.18: State Transitions—Parse Statement Block Brackets.

Current FSAMap 2 State	Input From FSAMap 1	Next FSAMap 2 State	Output To FSAMap 1
stmt	stmt	stmt	stmt
stmt list list list	list same list stmt	list list list stmt	list list list stmt
stmt expr expr	expr expr stmt	expr expr stmt	expr expr stmt
stmt forexpr forexpr forexpr	forexpr same expr stmt	forexpr forexpr forexpr foruntil	stmt stmt expr stmt
foruntil foruntil	expr stmt	foruntil fornext	expr stmt
fornext fornext	expr stmt	fornext stmt	expr stmt
stmt ifexpr ifexpr ifexpr	ifexpr same expr stmt	ifexpr ifexpr ifexpr stmt	stmt stmt expr stmt
stmt wexpr wexpr wexpr	wexpr same expr stmt	wexpr wexpr wexpr stmt	stmt stmt expr stmt

Table B.19: State Transitions—Context FSAMap.

Current FSAMap 1 State	External Input	Input From FSAMap 2	Next FSAMap 1 State	Output To FSAMap 2	Current FSAMap 2 State	Input From FSAMap 1	Next FSAMap 2 State	Output To FSAMap 1
stmt	read	stmt	list	list	stmt	stmt	stmt	stmt
list	a	stmt	comma	list	stmt	list	list	list
comma	,	list	list	list	list	list	list	list
list	b	list	comma	list	list	list	list	list
comma	;	list	stmt	stmt	list	list	list	list
stmt	;	list	stmt	stmt	list	stmt	stmt	stmt
stmt	for	stmt	lbrac	forexpr	stmt	stmt	stmt	stmt
lbrac	(stmt	stmt	same	stmt	forexpr	forexpr	stmt
stmt	a	stmt	loper	expr	forexpr	same	forexpr	stmt
loper	=	stmt	expreval	expr	forexpr	expr	forexpr	expr
expreval	1	expr	aoper	expr	forexpr	expr	forexpr	expr
aoper	;	expr	stmt	stmt	forexpr	expr	forexpr	expr
stmt	a	expr	loper	expr	forexpr	expr	forexpr	stmt
loper	<	stmt	expreval	expr	foruntl	expr	foruntl	expr
expreval	b	expr	aoper	expr	foruntl	expr	foruntl	expr
aoper	;	expr	stmt	stmt	foruntl	expr	foruntl	expr
stmt	a	expr	loper	expr	foruntl	stmt	foruntl	stmt
loper	++	stmt	aoper	expr	fornext	expr	fornext	expr
aoper)	expr	aoper	expr	fornext	expr	fornext	expr
aoper	;	expr	stmt	stmt	fornext	expr	fornext	expr
stmt	;	expr	stmt	stmt	fornext	expr	fornext	stmt

Table B.20: Results Using Example Code (a).

Current FSAMap 1 State	External Input	Input From FSAMap 2	Next FSAMap 1 State	Output To FSAMap 2	Current FSAMap 2 State	Input From FSAMap 1	Next FSAMap 2 State	Output To FSAMap 1
stmt	{	stmt	stmt	stmt	stmt	stmt	stmt	stmt
stmt	c =	stmt	loper	expr	stmt	stmt	loper	stmt
loper	a +	stmt	expreval	expr	stmt	expreval	expreval	expr
expreval	b ;	expr	aoper	expr	expr	aoper	aoper	expr
aoper	;	expr	expreval	expr	expr	expreval	expreval	expr
stmt	;	expr	aoper	expr	expr	aoper	aoper	expr
stmt	}	expr	stmt	stmt	expr	stmt	stmt	stmt
stmt	if	stmt	stmt	stmt	expr	stmt	stmt	stmt
lbrac	(stmt	lbrac	ifexpr	stmt	lbrac	lbrac	stmt
stmt	c	stmt	stmt	same	stmt	stmt	stmt	stmt
loper	>=	stmt	loper	expr	ifexpr	ifexpr	loper	stmt
expreval	3	expr	expreval	expr	ifexpr	expreval	expreval	expr
aoper)	expr	aoper	expr	ifexpr	aoper	aoper	expr
aoper	;	expr	aoper	expr	ifexpr	aoper	aoper	expr
stmt	;	expr	stmt	stmt	ifexpr	stmt	stmt	stmt
stmt	print	stmt	stmt	list	ifexpr	stmt	stmt	stmt
list	a	stmt	comma	list	stmt	list	comma	list
comma	,	list	list	list	stmt	list	list	list
list	c	list	comma	list	list	list	comma	list
comma	;	list	stmt	list	list	list	stmt	list
stmt	;	list	stmt	list	list	list	stmt	stmt

Table B.21: Results Using Example Code (b).

Current FSAMap 1 State	External Input	Input From FSAMap 2	Next FSAMap 1 State	Output To FSAMap 2	Current FSAMap 2 State	Input From FSAMap 1	Next FSAMap 2 State	Output To FSAMap 1
stmt	while	stmt	lbrac	wexpr	stmt	stmt	stmt	stmt
lbrac	(stmt	stmt	same	stmt	wexpr	wexpr	stmt
stmt	c	stmt	loper	expr	wexpr	same	wexpr	stmt
loper	<	stmt	expreval	expr	wexpr	expr	wexpr	expr
expreval	a	expr	aoper	expr	wexpr	expr	wexpr	expr
aoper	+	expr	expreval	expr	wexpr	expr	wexpr	expr
expreval	b	expr	aoper	expr	wexpr	expr	wexpr	expr
aoper)	expr	aoper	expr	wexpr	expr	wexpr	expr
aoper	;	expr	stmt	stmt	wexpr	expr	wexpr	expr
stmt	;	expr	stmt	stmt	wexpr	stmt	stmt	stmt
stmt	{	stmt	stmt	stmt	stmt	stmt	stmt	stmt
stmt	c	stmt	loper	expr	stmt	stmt	stmt	stmt
loper	=	stmt	expreval	expr	stmt	expr	expr	expr
expreval	1	expr	aoper	expr	expr	expr	expr	expr
aoper	+	expr	expreval	expr	expr	expr	expr	expr
expreval	a	expr	aoper	expr	expr	expr	expr	expr
aoper	;	expr	stmt	stmt	expr	expr	expr	expr
stmt	;	expr	stmt	stmt	expr	stmt	stmt	stmt

Table B.22: Results Using Example Code (c).

Current FSAMap 1 State	External Input	Input From FSAMap 2	Next FSAMap 1 State	Output To FSAMap 2	Current FSAMap 2 State	Input From FSAMap 1	Next FSAMap 2 State	Output To FSAMap 1
stmt	if	stmt	lbrac	ifexpr	stmt	stmt	stmt	stmt
lbrac	(stmt	stmt	same	stmt	ifexpr	ifexpr	stmt
stmt	a	stmt	loper	expr	ifexpr	same	ifexpr	stmt
loper	==	stmt	expreval	expr	ifexpr	expr	ifexpr	expr
expreval	c	expr	aoper	expr	ifexpr	expr	ifexpr	expr
aoper	*	expr	expreval	expr	ifexpr	expr	ifexpr	expr
expreval	2	expr	aoper	expr	ifexpr	expr	ifexpr	expr
aoper	-	expr	expreval	expr	ifexpr	expr	ifexpr	expr
expreval	b	expr	aoper	expr	ifexpr	expr	ifexpr	expr
aoper	/	expr	expreval	expr	ifexpr	expr	ifexpr	expr
expreval	3	expr	aoper	expr	ifexpr	expr	ifexpr	expr
aoper)	expr	aoper	expr	ifexpr	expr	ifexpr	expr
aoper	;	expr	aoper	expr	ifexpr	expr	ifexpr	expr
aoper	;	expr	stmt	stmt	ifexpr	expr	ifexpr	expr
stmt	;	expr	stmt	stmt	ifexpr	stmt	stmt	stmt
...					...			

Table B.23: Results Using Example Code (d).

B.3.2 Hierarchical Temporal Sequences—SOMA and LAPS

The full ABC model incorporates a number of SOMA and LAPS modules connected in a hierarchy. As such, separate tests of hierarchically connected combinations of these components were not performed. Tests of this kind are left to future research.

Suffice to say that a number of connected SOMA or LAPS modules will allow for a much richer form of mapping between inputs and outputs, as was shown to be the case with hierarchical FSAMaps.

B.4 Miscellaneous

Some miscellaneous experiments using the full temporal model are described in the following sections.

B.4.1 Testing the Learning of the Alphabet

In Section 2.4.3 on page 55 we experimented with learning the alphabet by reproducing the next letter in the sequence on each pass of the SOMA module. This section describes various tests performed in order to examine the tolerance of the method.

Variations on the theme of the usual alphabetic sequences were tried—for example the network was trained to learn m characters ahead—that is, given A learn C for 2 ahead, given A learn D for 3 ahead, and so on, including counting backwards. Of course this is to be expected as there is nothing significant in the vectors of the standard alphabetic sequence.

A number of experiments were performed in an attempt to disrupt the sequence and thus check the robustness of the method. In fact the sequence generation proved to be extremely robust.

The first test was to just arbitrarily change the sequence in the middle of a run programmatically; for example, whenever the sequence got to 'I', the program substituted the input vector for a 'K' on the next cycle. The sequence simply continued L, M, N,

and so on without a problem. This was true even if the jump was many characters away in the sequence; for example, a jump from I to Q—the sequence would just start up again at R.

Noise was introduced by flipping the ‘bits’ of the vectors associated with each letter. The test program randomly selected n bits out of the 35 for each output vector (after it was generated) [†], and if the corresponding output vector bit was a one it was changed to a zero and vice-versa.

The maximum number of bits flipped in each vector was supplied to the test program, but the actual bits were selected at random. A check was not kept on the status of each bit, so it was possible for the same random number (in the range [1,35]) to be selected twice, thus reversing a previously flipped bit. Thus the number of flips is *up to* n .

With this noise component the network exhibited very interesting, yet entirely expected behaviour.

The bit-flip experiment was of two types; the first combined a ‘jump ahead’ in the sequence (say from I to Q) with a random flip of up to 30 bits in the Q vector.

If the new ‘noisy’ *next* character vector that was applied to the input *resembled* an existing character vector (with a large fuzzy envelope) then the sequence would simply resume from this new character.

For example, if the new character resembled a D then the sequence EFGH ... would resume. If, however, the new character did not ‘adequately’ resemble any existing character, then the previous *sequence* component would force the sequence to continue from the next character that would have been expected had not the ‘jump’ and ‘bit-flip’ occurred; that is, if the character preceding the jump and bit-flip was an I, the sequence would resume JKLM ... and so on.

A statistical analysis of this behaviour was not performed, but it seemed that provided the Hamming distance of the flipped vector was within approximately 10 of an existing character, (and depending on the actual next *sequence transition* vector), then the sequence would take up at the successor of this perceived character.

[†]Refer to Figure 2.16 for the bit structure of the letters.

An example will illustrate this behaviour. The following table gives the Hamming distances from all of the existing character vectors for a flipped vector.

A	B	C	D	E	F	G	H	I	J	K	L	M
18	12	17	18	13	15	13	17	17	15	22	19	23
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
20	12	13	12	14	11	15	15	19	20	21	15	15

The character S is the closest, and as it has a sufficiently low Hamming distance the sequence will resume as T, U, V, ... However, if the Hamming distance to the nearest character is too large, the original sequence will continue unperturbed. For example, the following table illustrates a case for which the Hamming distance to all existing characters was too high.

A	B	C	D	E	F	G	H	I	J	K	L	M
16	18	17	20	21	23	17	19	17	23	18	19	15
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
14	18	21	16	18	21	19	19	19	16	19	21	19

Although the closest Hamming distance is for the character N (and N is chosen as the winning node for the first SOM), its output vector is too divergent, and so the sequence vector becomes more important in the selection of the second SOM winning node. In this case, the next transition in the previous sequence, (here the H/I node) wins on the second SOM surface and the sequence begins again unchanged with J, K, L, ...

The overall impression from this test was one of an extremely robust sequence generator; no matter where the sequence takes up after a disruption, it continues correctly from then on.

This behaviour is of course to be expected. In determining the winning node on the second SOM surface, if the actual letter component is not decisive then the sequence component will determine the winning node.

The second noise experiment involved adding noise to *every* generated character to see the overall effect on the sequencing. A test program was supplied with a seed character and programmed to generate the next say 30 characters in the sequence. After every output vector in the sequence was evaluated, up to n% of the vectors bits were flipped.

The 5% flip case exhibited minimal effect. Twenty four sequences of 30 characters each were generated from random seed characters—a total of 720 characters.

In 5 cases, the sequence PST was found—that is, the characters QR were missing. This resulted from a noisy P becoming more like an R and thus generating an S as the next character in the sequence. In another 2 cases, the sequence PQRQRS was found; that is, an additional QR component in the overall sequence. This is the reverse of the above case in which a noisy R more closely resembles a P, thus generating another Q. Both cases result from the fact that the vectors for P and R are very close in the Hamming distance of their representations.

Thus in only 7 transitions out of 720 was a difference observed—otherwise the 5% flips were absorbed without problem and the sequence generation of the alphabet was as required.

At the 10% level a few more changes were observed, as is shown in the following table.

Invalid Sequence	Original Char	Now Resembles	Times Occurs
EQ	E	P	1
FS	F	R	5
WO	W	N	1
HX	H	W	2
BT	B	S	4
BQ	B	P	1
LD	L	C	1
HN	H	M	1
OR	O	Q	4
FF	F	E	1
LF	L	E	1
NX	N	W	1
OC	O	B	1
PS *	P	R	7
RQ *	R	P	3

The asterisk indicates the two cases described in the 5% case above.

Here, from the 720 character transitions at the 10% flip level, 34 were incorrect. Again, in all other cases, the noise component was absorbed and the sequence continued as required.

At the 15% noise level, 115 out of the 720 transitions were incorrect.

Thus the noise tolerance of the network is seen to be quite impressive. The network is able to recover from minor disturbances and absorb the differences without interruption. For larger disruptions, the network may be pushed out of its original sequence but quickly recovers to resume the sequence from a new character.

B.4.2 Learning Characters and Recognizing Words

In this example we explore the relationship between characters and words. Input data consisted of the 'pictorial' characters as used in learning the alphabet (see Figure 2.15). The SOMA learning program was modified so that at the completion of the input of the characters for a particular word, an extra set of motor weights connected to a 'word' motor vector were updated. This is the only occasion in which words, rather than characters, are used in the system. The mappings on both SOM surfaces is of the continuous stream of characters associated with each word.

To reiterate, the only inputs are characters, and so the only way a word can be learned is by learning a sequence of characters, and associating that sequence with the word. Words are only used to update the weights of the additional 'word' motor action weights, as shown in Figure B.63.

The idea is to see if the system can transfer its learning from characters to words.

The sentences used as input data contained 32 words, 22 of which were unique. The actual sentences are not relevant. As each word was read in as input, the characters that made up the word were split off and entered as input data to the system. Normal learning at all levels was performed for each character. After all characters for a particular word were entered, the system was given an additional 'word' motor weights update phase to remember the whole word.

Also, the system was modified so that instead of learning predict the *next* word (as with most of the examples in this chapter), it was trained to recognize the *current* word.

The only inputs to the system were the characters, with the extra reinforcement of word choices at the completion of the characters for each word.

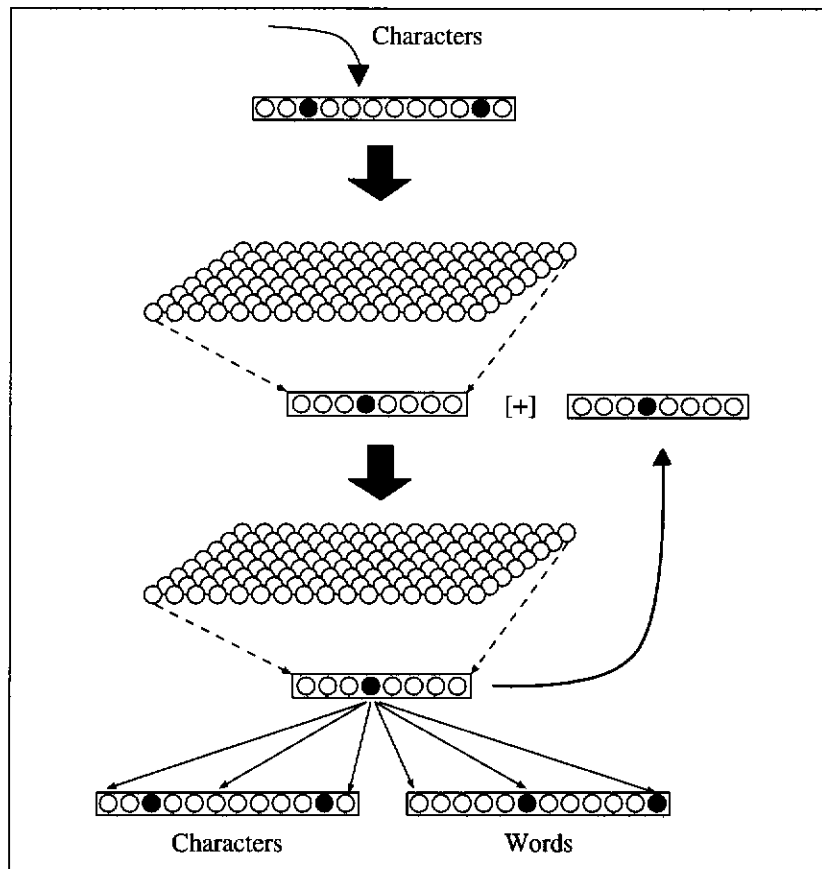


Figure B.63: Learning to ‘Recognize’—Characters into Words.

The results of the learning were checked by getting the program to predict the word just completed following the sequence of characters that make up the word. For example, after the characters **d**, **o**, **g** have been entered and processed by the system, the then current word motor weights are used, along with the output values from the second SOM surface, to determine the best estimate for the actual word.

The results obtained were quite good. The following table shows the results that the system achieved in learning the 32 words. Remember that there were 22 unique words, so the choice at each stage is out of twenty-two possibilities. As can be seen, the system was able to estimate the correct choice of word in 21 out of the 32 cases, with those words not estimated exactly being either the second or the third choice.

Correct Word	21
Second Choice	8
Third Choice	3

The actual words and their statistics are shown in the following table. The columns give the number of correct and incorrect determinations for each word.

Word	Num.	Corr.	Word	Num.	Corr.	Word	Num.	Corr.
a	1	0	dog	1	1	that	2	1
about	1	0	dogs	1	0	the	5	5
ate	1	1	fish	1	1	thinks	1	0
beds	1	1	girls	2	2	two	1	1
bird	1	0	keen	1	0	was	1	1
boy	2	2	loved	1	0	your	3	3
chase	1	0	on	2	2			
dances	1	0	sofa	1	0			

Another experiment was performed with a larger set of sentences; 28 unique words, 17 nouns, 7 verbs (plus 'to' as in 'listened to'), and 3 articles, arranged as before in 30 sentences of 5 words each (except for one which included a two word verb—*listened to*). The words used with their grammatical category are shown in the following table.

Nouns		Verbs	Articles
books	man	heard	a
bread	men	imagined	some
cat	mouse	listened to	the
cats	rock	saw	
cookie	rocks	smelt	
cookies	tigers	swallowed	
cup	woman	touched	
lion	women		
lions			

Learning is again quite good, as shown in the following table.

Word	Total Count	Num Cor	Num Inc	Posn.	Word	Total Count	Num Cor	Num Inc	Posn.
a	22	22			mouse	2	0	2	3,4
books	1	0	1	4	rock	5	5		
bread	1	0	1	6	rocks	1	0	1	3
cat	3	0	3	3*2	saw	9	9		
cats	1	0	1	4	smelt	3	0	3	3*2
cookie	4	0	4	3*4	some	8	1	7	7*2
cookies	2	1	1	3	swallowed	2	0	2	2*4
cup	1	1			the	30	26	4	4*2
heard	2	0	2	2,4	tigers	1	0	1	6
imagined	9	5	4	2*2,3,4	to	1	1		
lion	3	0	3	2,2*3	touched	4	0	4	3*2,3
lions	1	0	1	6	woman	13	7	6	6*2
listened	1	0	1	6	women	4	2	2	2,3
man	10	4	6	6*2					
men	5	2	3	2,2*3					

Remember that here the system is essentially estimating the word from the *last* few characters of the spelt-out word. The table lists the actual words, and the number of correct and incorrect estimates made by the system. The final column gives the position in the sorted list of those words estimated incorrectly. An asterisk indicates that there were a number of occasions in which the word was estimated in that position; for example, 3*2 indicates that there were 3 incorrect estimates for the word in which it was actually found as the second choice.

As can be seen, the system does reasonably well—even if it does not choose the correct word, the statistics collected ensure that the correct word is very close to the top of the sorted list.

The case of the pairings **man/woman** and **men/women** is interesting. In all cases where **man** or **woman** was incorrect, the other was the (only other) choice—this is to be expected as these two words are the only ones ending in ‘an’. In most cases of incorrect selection for **men** and **women**, the other was the chosen word. A few cases resulted in **lion** also being selected due to cross-talk in the transition map.

It is very important that significance is not given to artifacts. The SOMA mechanism described in this chapter is a very powerful sequence learner, and it is important that the results are not misinterpreted. For example, a cursory examination of the following table may give the impression that the network does an excellent job of choosing the correct part of speech—noun, verb, or article.

Noun Expected			Verb Expected		
Exact Word	Other Noun	Other Type	Exact Word	Other Verb	Other Type
22	27	11	14	13	3
49		11	27		3
81.7%			90%		
Article Expected			Preposition Expected		
Exact Word	Other Article	Other Type	Exact Word	Other Prep	Other Type
49	11	0	1	0	0
60		0	1		0
100%			100%		

However, it must be noted that the words were selected on the basis of the *final* few characters given as input, and the words from each grammatical type as used here

tend to end with unique characters. For example, most of the verbs end with 'ed'. Hence, one could claim that the the above table is an artifact. However, cues of this sort *are* used by humans in language learning and use (see MacWinney et al. (1989) for a connectionist analysis of the use of such cues in the learning of German definite articles).

B.4.3 Spelling

This experiment is similar to the previous one except that here, a word and then a series of characters were supplied as input; for example, *cat c a t dog d o g*. This is different to the above case in that both words—via their vectors, and characters—again via their vectors, were supplied as input.

After training, an alternative test program was supplied with various *words* as seeds, and then required to generate the next n symbols. If it has learned the training examples then it will have learned to 'spell'. A standard SOMA program was used for this experiment, and the task was to learn the various sequences.

A possible psychological justification for this method is that learning to spell usually proceeds via a sounding out of the word and its component characters. For example, *cat - c - a - t* spells *cat*

The spoken character and word sounds are equivalent variants of sound 'symbols' that are mapped onto some auditory SOM surface.

An early experiment produced mixed results because of cross-talk. There were 24 words of mixed lengths, trained in the order shown in the following table, but tested in random order. The first column gives the actual word used, whereas the second column indicates the correct characters that resulted from the sequence reproduced when the word was used as a seed.

Word	Characters Correct	Word	Characters Correct
the	the	cognition	co
quick	quic	computational	co
brown	brow	man	man
fox	fox	woman	woman
jumped	jum	superman	su
over	over	cats	cat
lazy	la	catalog	cat
dogs	dog	catalogue	cat
this	th	category	cat
that	that	catwoman	cat
mat	mat		
sat	sat		
cat	cat		
hat	hat		

Some character sequences in the training set were learned reasonably well, but the length of correct regeneration of characters was not very good. There was too much crosstalk occurring between similar or identical sequences in different words, so in order to better separate the nodes in the learning, it was decided to use the 'exclusion' modification discussed in Section 2.3.6.

A new test set of 24 four character words was selected. These words were chosen because they were shorter (and so are able to be learned more readily), and included various common sub-strings—a good test of the ability of the network to handle cross-talk. The results of this experiment are shown in the following table.

Word	Characters Correct Run 1	Characters Correct Run 2	Word	Characters Correct Run 1	Characters Correct Run 2
able	able	able	near	near	near
born	born	born	over	over	ove
clay	clay	clay	prey	prey	prey
damp	damp	damp	quit	qui	quit
ergo	ergo	ergo	riot	riot	riot
five	five	five	sham	sham	sh
gosh	go	gosh	that	that	that
hats	hats	hats	user	us	us
junk	junk	junk	vein	vein	ve
knot	knot	knot	wren	wre	wren
lazy	laz	lazy	year	year	yea
mint	mint	mint	zone	zone	zone

The two columns show the results of two separate runs—in the first the words were tested in the same order in which they were learned—in the second they were tested in random order.

As can be seen, the system does extremely well, despite numerous sequence conflicts, some of which are shown in the following table.

nEAR	yEAR		dAMP	shAM	
HATs	SHAM	thAT	mINt	veIN	
CLAY	LAzy		erGO	GOSH	
fiVE	oVER	veIn	ERgo	oVER	usER
NEar	zONE		HATs	thAT	
pREy	wREn		knOT	riOT	
gOSH	SHam		thAT	HATS	

Another run with somewhat longer words produced the following results, thus indicating that extended sequences are possible.

Word	Characters Correct
handbook	han
special	special
behaviour	behaviour
principles	princi
emergency	emergen
university	univers
overseas	ov
generous	genero
cultural	cultural
advanced	advance

B.4.4 Computer Language

The network was given the task of learning a simple computer-like language.

```

stmts ;
if ( cond ) stmts ;
if ( cond ) stmts else stmts ;
while ( cond ) stmts ;
loop stmts until ( cond ) ;

```

The system was able to learn the language reasonably well, despite the problems caused by the ambiguities within the language. We will return to the problem of ambiguities in (computer) languages in a later section.

The network was obviously unable to accurately predict the first word of a new statement following a terminating ‘;’. However, once this initial word was recognised, the network was able to correctly predict the next several symbols.

Because of the ambiguity within the statements

```
if ( cond ) stmts ; and if ( cond ) stmts else stmts ;,
```

the network was unable to predict the `else` on every presentation of an `if` statement, but once the `else` has been recognised in sequence then the rest of the full statement (`stmts ;`) is predicted accurately.

The one difficulty occurred in predicting the final semi-colon in the statement

```
loop stmts until ( cond ) ;
```

The capability of the network to 'look-back' was overextended, resulting in confusion with the other sequence components of `(cond) stmts` found in the `if` and `while` statements. Although the network incorrectly predicted the component `stmts` at the end of a `loop` statement, the second choice was the correct `;`.

The network was able to predict the components `(cond)` and `until` to 100% accuracy.

In order to test the learned weights, an additional program was used to read in the weights, and to then generate the next n components in sequence following a seed word (see Section 2.4.3 for a discussion of sequence *generation*).

The table below summarises the generated components, indicating that the sequences have been learned.

Seed Word	Next n Generated Words					
if	(cond)	stmts	;	while cond
loop	stmts	until	(cond)	stmts ;
while	(cond)	stmts	;	loop ;
else	stmts	;				

B.4.5 Loop

An interesting test of temporal behaviour was used by Kangas (1989). Their test system consisted of a two dimensional space where an object moved on a figure-eight shaped track. The position of the object was measured in even intervals of time. Without temporal sequencing, the algorithm is not able to differentiate between the directions of the object when the track crosses over itself in the centre of the figure-eight.

This example is reproduced here as illustrated in Figure B.64. Because of the binary input requirement of the current version of the SOMA algorithm, the figure-eight is implemented as two adjacent squares.

This sequence presented no problems to the network, which was accurate to 100%. The selection for the centre point was always given as `point5` rather than `point13` on the return journey, but as both points were given an identical vector representation the names may be taken as pseudonyms.

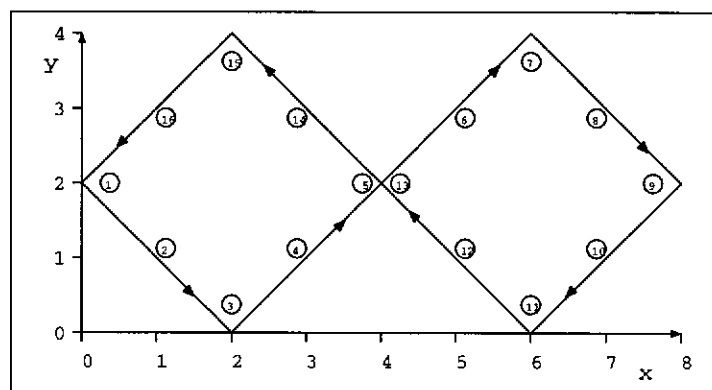


Figure B.64: Loop Trace.

Appendix C

Implementation of the Full Simulation Model

C.1 Introduction

One of the long-term goals of the research described in this thesis is to build an artificial vision system. With this in mind, a major project was begun in late 1995 to implement such a system. This appendix describes the ongoing development of the simulation, along with some of the background and rationale for this particular design.

The specific implementation undertaken came about as a result of a general challenge to cognitive scientists issued by Feldman et al. (1990). These authors felt that cognitive science was reverting to a disjoint collection of fields rather than maintaining an interdisciplinary perspective. They accordingly posed what they described as a *touchstone* problem for cognitive science—a compact, theory-free task that required an integrated solution. In essence, the task is to learn a subset of an arbitrary natural language from picture-sentence pairs. The task thus requires an integration of several areas of cognitive science including vision, language and learning.

The system that we have developed in response to this challenge is near completion. As well as answering the challenge, the simulator that is being implemented is also intended to be used as a general vehicle to evaluate various aspects of the ABC model.

We briefly describe the challenge in the following section, then move on to describe the implementation in subsequent sections.

C.2 The Challenge

The specific challenge posed by Feldman et al. (1990) is to build a system capable of performing a Miniature Language Acquisition (MLA) task, such that:

- the system is to be given examples of pictures paired with statements about those pictures in a natural language,
- the system is to learn the reduced language so that it can indicate whether a subsequent sentence is true of an accompanying picture.

Feldman et al. suggest that the problem has a number of attractive features. The task is strictly behavioural and theory-free, the proposed solution must be complete and must not resort to simply hand-waving about various sections of the task. While the MLA task is a reduction on the complexity and richness of the human language capability, it nevertheless requires that solutions be provided to deep questions in several areas of cognitive science, and that various currently separate disciplines be incorporated into one integrated whole. For example, the task requires a solution to the so-called *symbol-binding* problem.

The initial language specification (designated L_0 by Feldman et al.) is given by the following grammar specification:

```
S      = NP | NP VP
NP     = DET NP1 | DET NP1 and DET NP2
VP     = VI PP | VT NP
NP1    = OBJ | SHADE OBJ | SIZE OBJ | SIZE SHADE OBJ
PP     = REL NP
VI     = is | are
VT     = touches | touch
DET    = a
OBJ    = circle | square | triangle
SHADE = light | dark
```

SIZE = small | medium | large
 REL = REL1 | far REL1
 REL1 = above | below | to the left of | to the right of

Scenes are to consist of up to four objects drawn from a population of three shapes, (*circle*, *square* and *triangle*), with two shades (*light* and *dark*). The size and location of the objects is arbitrary. The original document specified that objects should not partially occlude each other, but our system allows for that possibility as occlusion is an important issue in vision.

During training, the system is to be presented with grammatically correct sentences and a corresponding picture for which the sentence is true. Isolated noun phrases are allowed in the initial stages.

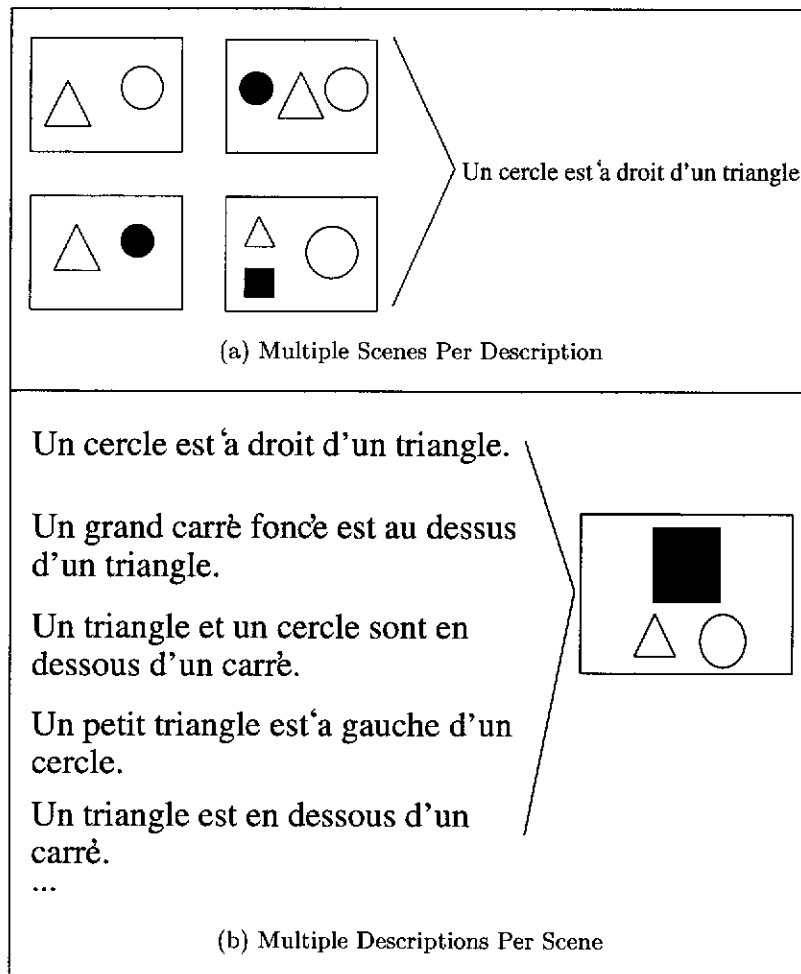


Figure C.1: Correspondence Between Images and Sentences

A given description may be consistent with a large (but finite) set of scenes, as shown in Figure C.1 (a). As well, a given scene may be consistent with a large (but also finite) set of descriptions, as shown in Figure C.1 (b).

Various other extensions to this initial specification are mentioned in the original article, such as relative sizes, quantifiers, abstraction and negation. As well, we have added other minor extensions to the task such as allowing for shapes of various colours and textures (striped, spotted, etc.), and object motion, expansion and contraction. The number of object types has also been extended to include other shapes (star, cross, and so on).

C.3 The Full Simulation Model

We specifically set about to implement the essence of ABC as a working model of the visual system, rather than as a general model of cognitive at this stage. For this reason, the input filters for the visual component are implemented in as realistic a manner as possible, whereas for the only other modality to be included—audition—the input filters are simply assigned on a random basis for each separate word in the language. A future development may incorporate more ‘naturalistic’ auditory inputs, but the nature of the vector processing suggests that in principle there will be no difference.

As per the challenge, the system is to be given paired images and sentences, and is required to indicate whether the sentence is true with respect to the picture. The simulation represents a single, unified architecture which in computer vision parlance, goes *from pixels to predicates*.

In the following sections, we indicate the ‘external world’ filters which provide input to the model, and the theory behind some of these.

C.3.1 Visual Model—Receptive Fields

The receptive fields used in the simulations are shown in Figure C.2. The simulator is flexible in that it is able to incorporate other divisions of the retinal field into receptive

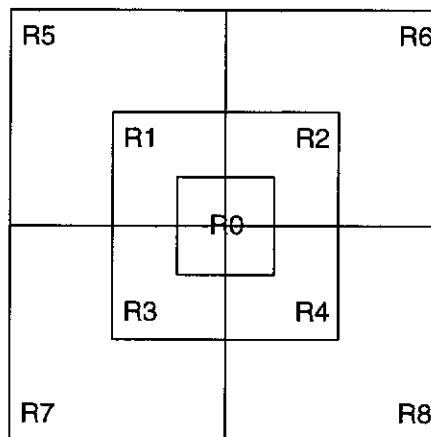


Figure C.2: Receptive Fields Used in Simulations

fields, but the scheme chosen was thought to provide simplicity (and hence a small number of vector elements to process) as well as being adequate for the task. Finer divisions will be attempted if necessary.

In Figure C.2, the central area **R0** represents the receptive field corresponding to a form of fovea, areas **R1** to **R4** represent directionally sensitive fields, and areas **R5** to **R8** also represent directionally sensitive fields of larger areas. All areas are rectangular. Note that any stimulation falling within the fovea will be included in all fields—in effect its influence will be greater than that of any stimulation falling outside **R0** but within **R1** to **R4**, and this in turn will be greater than any stimulation falling within **R5** to **R8** but outside of **R1** to **R4**.

Note also that this arrangement of receptive fields allows for the learning of attributes such as *above* and *to-the-left-of*.

C.3.2 Visual Model—Wilson Modified Line-Element Model

The Wilson model of visual processing (Wilson & Gelb 1984, Wilson 1986) is based on masking experiments which suggest that spatial pattern discrimination may be a “line-element” mechanism, similar to the line-element model of colour-discrimination. With chromatic discrimination, the three cone types have pigmentation sensitivities which respond to different wavelengths of light. The combination of responses of these

cones gives us the composite discrimination.

The spatial frequency masking experiments led Wilson to suggest that there are 6 spatial frequency tuning mechanisms used in spatial discrimination, with peak frequencies of 0.8, 1.7, 2.8, 4.0, 8.0 and 16.0 cycles per degree (cpd).[†]

The Wilson model is termed a ‘modified’ line-element model because it includes spatial separation mechanisms as well as spatial frequency mechanisms. The overall structure of the model incorporates a 2D linear filter ensemble which includes spatial and orientation selective mechanisms, as well as filter responses from nearest neighbours.

All filter mechanisms at each processor proceed in parallel. Thus the resultant values may be thought of as points in a multi-dimensional space with axes comprising the various spatial frequencies, orientations and nearest neighbours. At each time step, the spatial frequencies provide an appropriate vector for processing.

In the original Wilson model, the output of each filter is then passed through a non-linear contrast function which is compressive at high contrasts and accelerated at low contrasts. This is not done in our simulations.

To compare the responses to two stimuli, Wilson used the Euclidean distance between the two points in the multi-dimensional space. This value was then related to a psychometric function to allow comparison with the experimental results.

Wilson suggested that at each processing location of the receptive field, 2D filters are used. For a vertically aligned filter, the form of the filter equation is:[‡]

$$R_i(x, y) = A \left(e^{-\frac{x^2}{\sigma_1^2}} - B e^{-\frac{x^2}{\sigma_2^2}} + C e^{-\frac{x^2}{\sigma_3^2}} \right) e^{-\frac{y^2}{\sigma_y^2}}$$

$R_i(x, y)$ is the Response Function of the i th filter at point (x, y) , where x and y are the horizontal and vertical axes respectively.

[†]While some may object that Wilson’s results are not entirely correct, and that there is some evidence that the spatial frequency mechanism may be adaptive rather than fixed, the results obtained by Wilson will suffice for the simulation.

[‡]Note that Wilson was only concerned with the fovea, and did not consider any variation in the Response Function with retinal eccentricity.

Peak Frequency (cpd)	A	B	C	σ_1	σ_2	σ_3	σ_y	Orientation Difference
0.8	123.19	.267	-	.198	.593	-	.634	30.0°
1.7	123.19	.333	-	.098	.294	-	.314	30.0°
2.8	123.19	.894	.333	.084	.189	.253	.269	20.0°
4.0	123.19	.894	.333	.059	.132	.177	.189	20.0°
8.0	123.19	1.266	.500	.038	.060	.076	.122	15.0°
16.0	123.19	1.266	.500	.019	.030	.038	.061	15.0°

Table C.1: Values of the Constants Defining the Spatial and Spatial-Frequency Sensitivity for Each Mechanism as Determined from Masking Data. From Wilson & Gelb (1984).

The parameter values A , B , C , σ_1 , σ_2 and σ_3 (which have been determined from masking experiments) vary for each of the six spatial frequencies, and are shown in Table C.1. A plot of each of the filters, showing their relative sensitivity, is shown in Figure C.3. Figure C.4 shows a three-dimensional view of a typical filter.

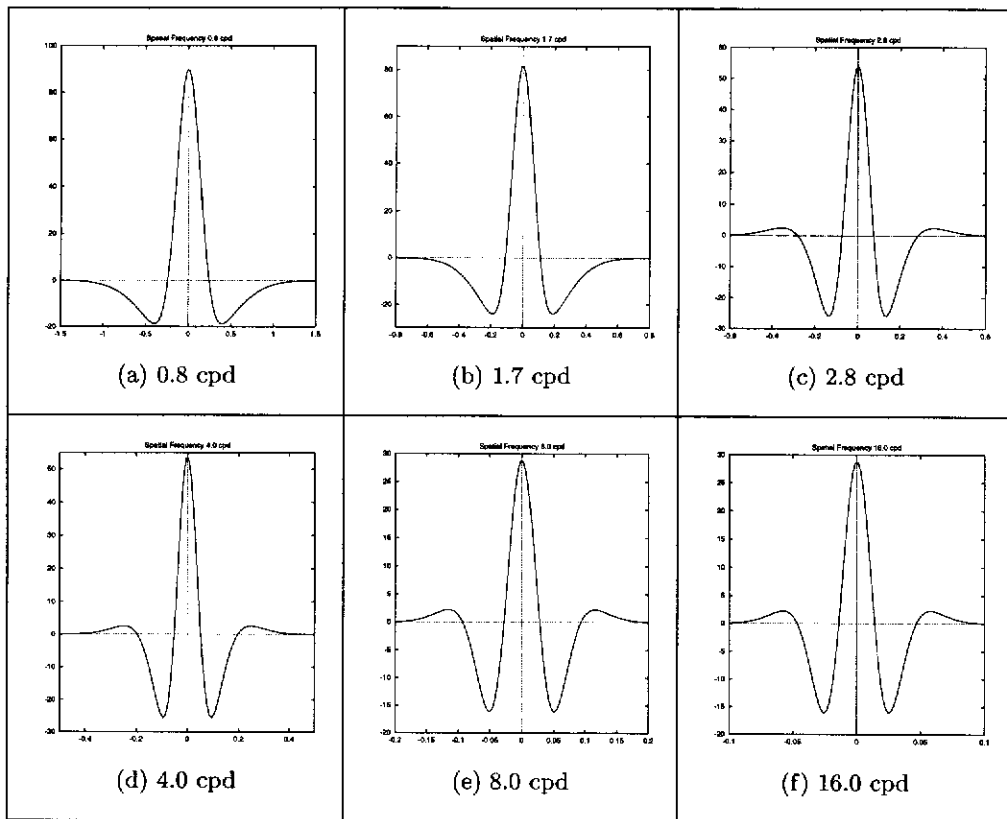


Figure C.3: Plots of the Spatial Frequencies (1D)

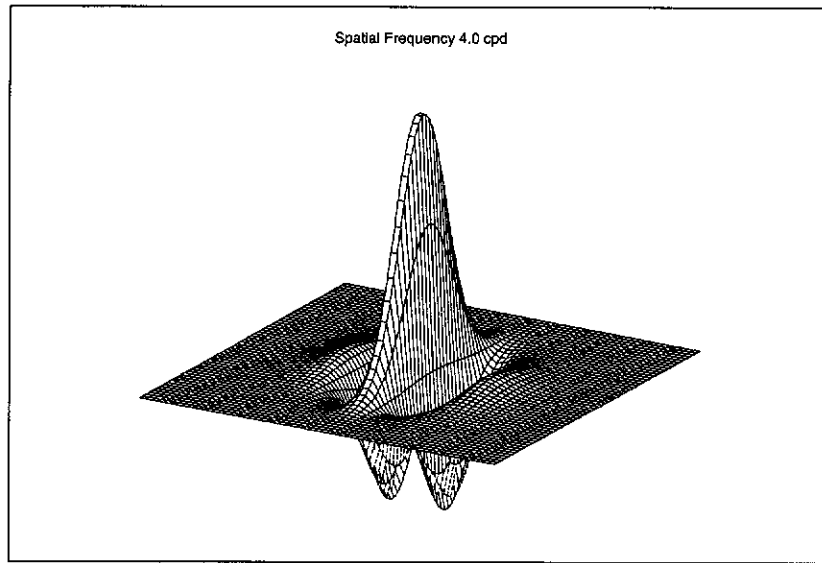


Figure C.4: Spatial Frequency 4.0 cpd

At each point in the receptive field, not only is a vertical filter active, but other filters, centred at the same point but with different orientations, are also active. The orientation differences depend on the spatial frequency of the filter, and are also shown in Table C.1. Thus, for example, the 0.8 and 1.7 cpd filters are separated from each other by 30 degrees, and so for these spatial frequencies there are 6 separate filters tuned to different orientations—at 0, 30, 60, 90, 120 and 150 degrees to the vertical. For the middle range spatial frequency filters, the orientation difference is 20 degrees, giving 9 orientation filters (0, 20, 40, 60, 80, 100, 120, 140, 160), while the higher spatial frequency filters support 12 orientations, each separated by 15 degrees (1, 15, 30, 45, 60, 75, 90, 105, 120, 135, 150, 165).[†]

Wilson found (Phillips & Wilson 1982) that for all spatial frequencies, the spread of the filter (σ_y) perpendicular to its primary axis was related to the spread of the primary

[†]The equation for filters oriented at an angle to the vertical is given by a rotation of the coordinates of the vertical filter equation:

$$\begin{aligned}x' &= x \cos \theta + y \sin \theta \\y' &= y \cos \theta - x \sin \theta\end{aligned}$$

excitatory Gaussian (σ_1). This relationship is:

$$\sigma_y = 3.2\sigma_1$$

The method then allows the calculation of the sensitivity of the i th filter $S_i(x, y)$ to a luminance pattern $P(x', y')$ via the convolution: †

$$S_i(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_i(x - x', y - y') P(x', y') dx' dy'$$

Wilson's initial attempt at developing a model included just these mechanisms—filters for spatial frequency and orientations, followed by the non-linear transfer function. However he found that this did not account for the experimental data at all well. He then incorporated nearest neighbour effects by adding (in parallel) neighbouring filters. Used as a free parameter, he found that the inclusion of two nearest neighbours best described the data. So for each mechanism, the Wilson model also includes units at other (neighbourhood) locations separated by a fixed fraction of the excitatory space constant σ_1 , that is:

$$0.56 \times \sigma_1$$

Thus at each processor location, three spatial filter responses are pooled. This implies that the visual system uses spatial information in addition to spatial frequency information to encode patterns. The use of spatial information is the reason the Wilson model is called a “modified” line-element model—the chromatic line-element has no counterpart to the spatial component.

Wilson justifies the spacing of the neighbouring filters by suggesting that it is consistent with optimal spatial sampling. He also suggests, (at least for one spatial frequency), that this choice is consistent with the latest sampling data on cone spacings in the human eye. For the highest spatial frequency of 16.0 cpd, $\sigma_1 = 0.019$, and so at this frequency the neighbouring filters are separated by

†The method also involves a non-linear contrast transfer function, but we ignore this component here.

Spatial Freq.	$\sigma_1 \times 0.56$	Equivalent Angle	Multiples of 38.5"
0.8	.1101°	6' 36.3"	10.3
1.7	.0549°	3' 17.6"	5.2
2.8	.0470°	2' 49.3"	4.4
4.0	.0330°	1' 58.9"	3.1
8.0	.0213°	1' 16.6"	2.0
16.0	.0106°	38.3"	1.0

Table C.2: Nearest Neighbour Filter Separations

$$0.56 \times 0.019 = 0.0106^\circ = 38.3'' \text{ of arc}$$

Cones at the fovea are separated by 24'', but this decreases to about 38.5'' within 0.25° of the fovea centre. Thus the spacing of neighbouring units for the highest spatial frequency mechanism is about the same as the cone separations. The values of the multiples of cone separations for other spatial frequencies is shown in Table C.2.

Wilson used the model to explore hyperacuity—the seemingly extraordinary accuracy with which the human eye can estimate the relative positions of lines or other features in the visual field (Westheimer & McKee 1977, Fahle & Poggio 1981, Klein & Levi 1985).[†] The expected responses to the various vernier shapes may be calculated and a discrimination threshold determined by computing the Euclidean distance between pattern representations in a multi-dimensional response space.

The Wilson model has been shown to give excellent results in hyperacuity and other aspects of early vision (Wilson & Regan 1984, Wilson 1986, Briscoe 1991).

In summary, the Wilson modified line-element model suggests that spatial discrimination is determined by vector differences in a multi-dimensional space. The axes in this

[†]For example, an acuity as small as 5'' may be achieved for the discrimination of a vernier (Westheimer & McKee 1977)—the equivalent of 0.02 mm at 1 m distance. This would seem impossible given a minimum cone spacing of 25'' at the fovea. The optical system is thus able to “interpolate the position of a feature in the visual field with an accuracy of better than one-fifth of the distance between two neighbouring photo-receptors and one-tenth of the smallest receptive field centre of the ganglion cells” (Fahle & Poggio 1981, page 452).

space are:

- 6 spatial frequencies,
- central filter plus 2 nearest neighbours,
- other preferred orientations centred at the stimulus point.

This is shown diagrammatically in Figure C.5.

Rather than use a multi-dimensional ‘feature’ space (the features being the spatial frequencies and angles), we construct a vector of the spatial frequencies and angles for mapping to a SOM surface. The elements of the vector are formed by finding the convolution of the input object intensity values and each spatial filter within each receptive field. The resultant values for each pixel are calculated using the sum of squares of the values within each receptive field. Thus there are $n_{SF} \times n_{RF}$ vector elements, where n_{SF} is the total number of distinct filters (including all rotations) and n_{RF} is the number of receptive fields.

C.3.3 Visual Model—Colour and Motion Filters

The human visual system has three different types of colour-sensitive photo-receptors in the retina. These cones show maximum absorption in the blue, green and red visible spectrum. At deeper levels of the retina and the visual cortex, psychophysical and neurophysiological experiments have shown that colour processing is done in terms of opponent colours; that is, there is evidence for two colour-opponent channels in addition

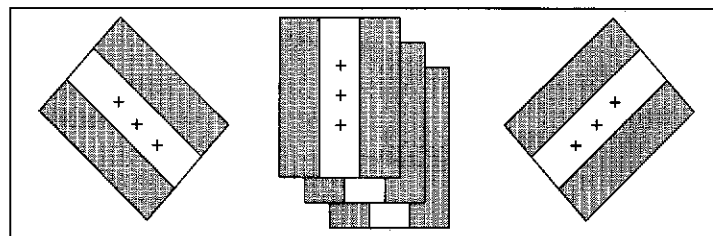


Figure C.5: Schematic Diagram of Model Visual Filters. Taken from (Wilson 1986). Note that the vertical displacement of the nearest neighbour filters is for diagrammatic purposes only.

to an achromatic channel (De Valois, Abramov & Jacobs 1966, Zeki 1980, Zeki 1983, Livingstone & Hubel 1984, Zeki 1993). The two opponent channels are red-green (r-g) and blue-yellow (b-y).

However, for the initial experiments using the simulator, we have taken a simpler route and implemented an *RGB* scheme. The colour of each pixel within each receptive field on the simulated retina is examined and separated into red, green and blue components. A (sum of squares) average is taken for each colour for each receptive field. As well, a monochrome (intensity) average is also calculated for each receptive field, thus forming a $4 \times n_{RF}$ vector, where n_{RF} is the number of receptive fields.

Later experimentation will incorporate opponent channel schemes.

Similarly, initial experiments will be performed with an artificial measure of object motion based simply on knowledge obtained from the internals of the simulator. Motion components for each pixel in each receptive field are separated into four separate elements—north, south, east and west. These values are averaged (sum of squares) for each receptive field and for each motion component. There are thus $4 \times n_{RF}$ motion vector elements, where n_{RF} is the number of receptive fields.

Again, later experimentation will incorporate more psychophysically realistic motion detector schemes, such as Reichardt-like detectors (van Santen & Sperling 1985).

C.4 Simulation Methods

The software development has so far consumed approximately one man-year of effort, and has produced some 40,000 lines of code. The system is written in the C++ language, and makes extensive use of template container classes. Mosaic and X-Windows are used to provide a graphical user interface (GUI).

The simulator was designed to be able to vary the architecture of the model and the connections between the components, as well as the sequence of events that made up the training of each experiment. In order to achieve these goals, there are several distinctive features in the design of the simulator;

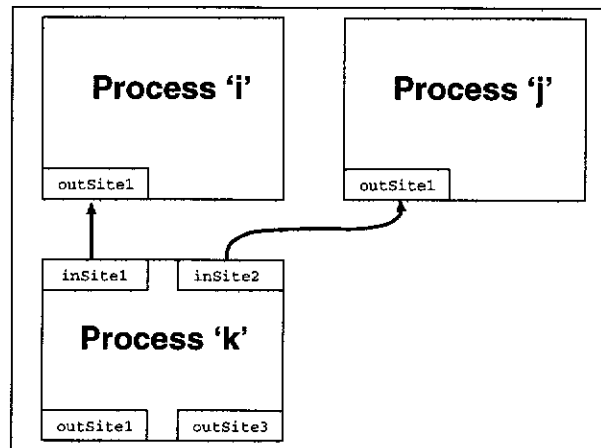


Figure C.6: Connecting *inSites* and *outSites* of Modules

data flow design

to ensure flexibility in the testing of various architectures and connectivities between components, a *data flow* design has been used,

action scripts

a form of scripting language was designed to ensure a convenient, easy and yet rich specification of the picture objects and their behaviours,

reusable weights

in order to be able to repeat, continue or vary standard experiments, the facility to save and restore previously learned weights (for a particular structure) is included.

The first two features are described in more detail in the following sections.

C.4.1 Data Flow Design

The framework of the simulator is centred around generalised building blocks called *modules*. Each module is able to be attached to other modules such that the outputs from one flow to the inputs of the following module. †

†The general design for the data flow nature of the simulator was modelled on that used in the *Sesame* neural network simulator tool (Linden & Tienz 1992, Linden & Tienz n.d., Linden, Dudbrak, Tietz & Weber 1992).

Each module has a number of *inSites* and *outSites* which provide communication between them. Each module links to its predecessor and successor site(s) via a linkage of *inSites* to *outSites*. This is shown diagrammatically in Figure C.6. To correctly handle the flow of data at each time step, and to ensure correct usage of recurrent data, the *inSites* and *outSites* are double buffered.

Flexibility of generating and connecting modules was a major design aim. Each module is initiated, and its connections specified, via a script language. For example, the following script segment specifies that a process of type *AssociationProcess* be initiated and added into the overall structure. It is given a title, and two (out of a possible three) of its *inSites* are connected to other processes. The *inSites* are specified by type (matrix or vector) and size (through the use of variable names such as <k113>) for overall consistency checking. The single *outSite* is given a size (<a113>), the process is allocated to a group and subgroup for later (diagrammatic) reporting, and the associated (standard) GUI window is turned off when the process is initiated.

```
process = theAssociationL1M1S3
{
  type      = AssociationProcess
  title     = "Association_L1_M3"
  nInSites  = 2
  inSite1   = matrix <k113> by <k113>
  inSite1WHO = theKohonenL1M1S3 at 1
  inSite2   = matrix <k112> by <k112>
  inSite2WHO = theKohonenL1M1S2 at 1
  nOutSites = 1
  outSite1  = vector <a113>
  group     = Vision
  subgroup  = Motion
  windowOpen = "no"
}
```

This method of connecting modules has proven to be easy and convenient to use, and modifications to the structure are able to be made quickly and safely.

C.4.2 Action Script

In order to facilitate a simple and convenient method of specifying the picture objects and their behaviours, a script language was developed. This allows for the specification

of each picture object, including their type, colour, size, position on the screen, and motion—for example:

```
actor = redtriangle {
    type = triangle
    color = red
    size = random
    position = random
    fill = solid
    motion = deltas - 10 , + 30
}
```

specifies a red triangle, of random size and position (within limits), and solid texture. The triangle is to be given a velocity of -10 in the x direction and 30 in the y direction.

Actor objects, once defined, may be used in multiple (picture) scenes; for example:

```
image = image1 {
    who = redtriangle
    who = hugebluecircle
}
```

```
sentence = utterance1 { a red triangle and a large blue circle . }
```

```
scene = scene1 {
    visual = image1
    dialog = utterance1
    ticks = 9
}
```

specifies that the objects (actors) *redtriangle* and *hugebluecircle* be incorporated into a scene, along with an accompanying sentence. The learning ‘cycle’ repeats after every 9 presentations (ticks), with the next word being learned at each tick, recycling to the beginning at the end of each cycle.

Finally, a number of scenes may be linked together into a composite learning regime, termed a *play*. As well, previously saved weights may be loaded from file, or the weights saved to file at any stage in the processing. Various reports are available to be printed, including an evaluation of whether the sentences match the pictures.

```
play = play1 {
    readfromfile "abc.save1"
    repeat scene1 10 times
```

```
repeat scene2 15 times
...
savetofile "abc.save2"
report "summary" "abc.summary"
}
```

C.4.3 GUI Windows

In this section, we describe some of the GUI windows which are available in the simulator.

Loader Window

The *Loader Window* allows for the loading and parsing of both the structure script and the action script, and is shown in Figure C.7. Other facilities allow for the iconification and restoration of the various windows associated with the application.

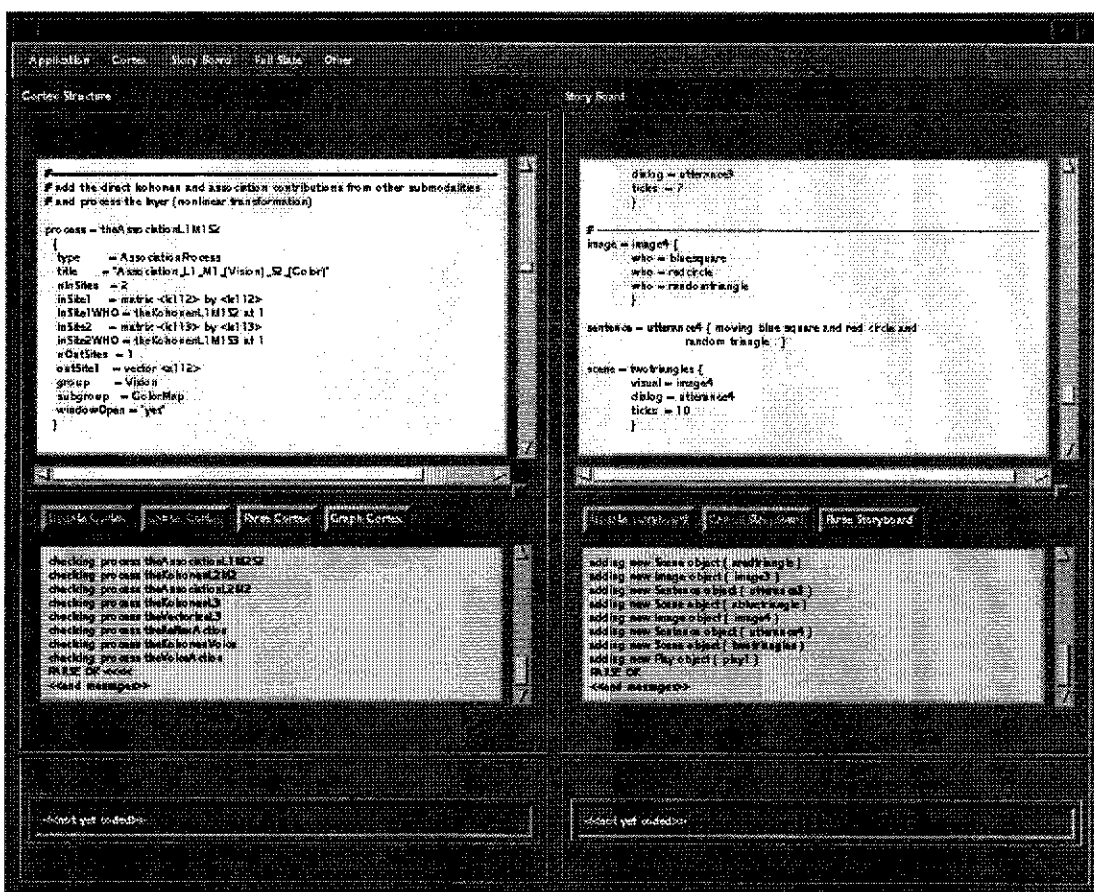


Figure C.7: Loader Window

Experiment Window

The *Experiment Window* controls the actual simulation. Figure C.8 shows the window prior to initiating an experiment, whereas Figure C.9 shows the upper portion of the window during an actual run.

At the top of the window are three canvases. The first represents the ‘world’ of objects that are ‘seen’ by the simulated entity. The second canvas indicates, via a colour-coding scheme, the speed and direction of each object. The third, smaller window is a sub-section of the world canvas, and represents the actual view as seen by the simulated retina.

During a run, various objects are placed on the world canvas, and the corresponding sentence displayed in the boxes underneath. The word currently being learned is

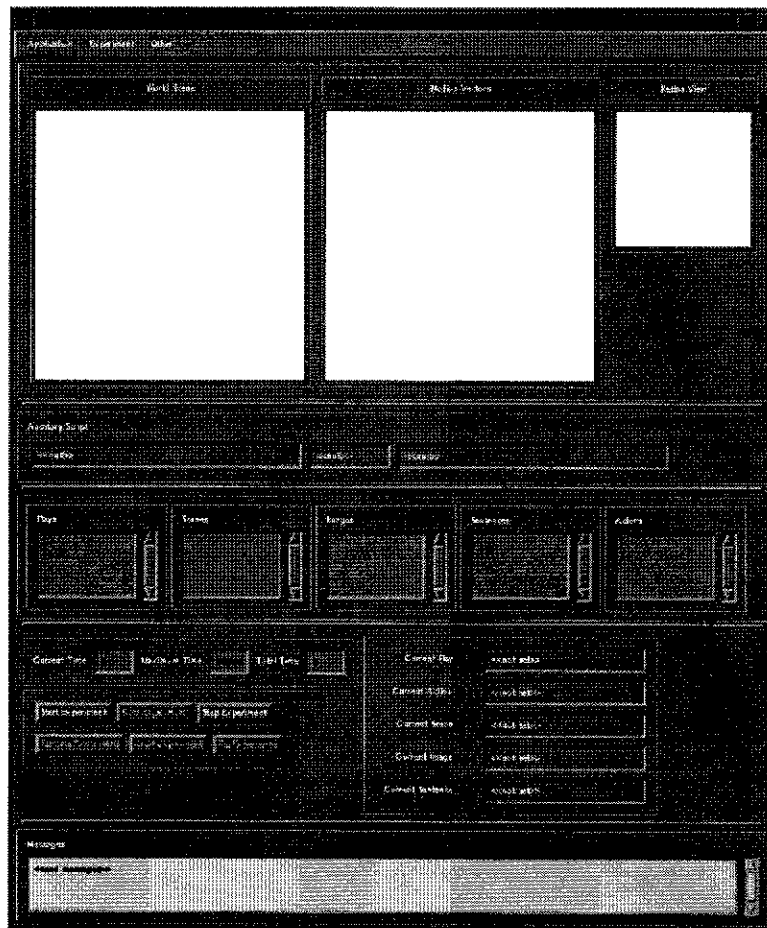


Figure C.8: Experiment Window

displayed in the centre box, whereas those previously examined are shown to the left, with those still to be processed on the right.

Other items on this window include lists of the actors, images, scenes and plays that are included in this run, as well as the times for various operations. Buttons allow the experiment to be halted, restarted or executed a step at a time.

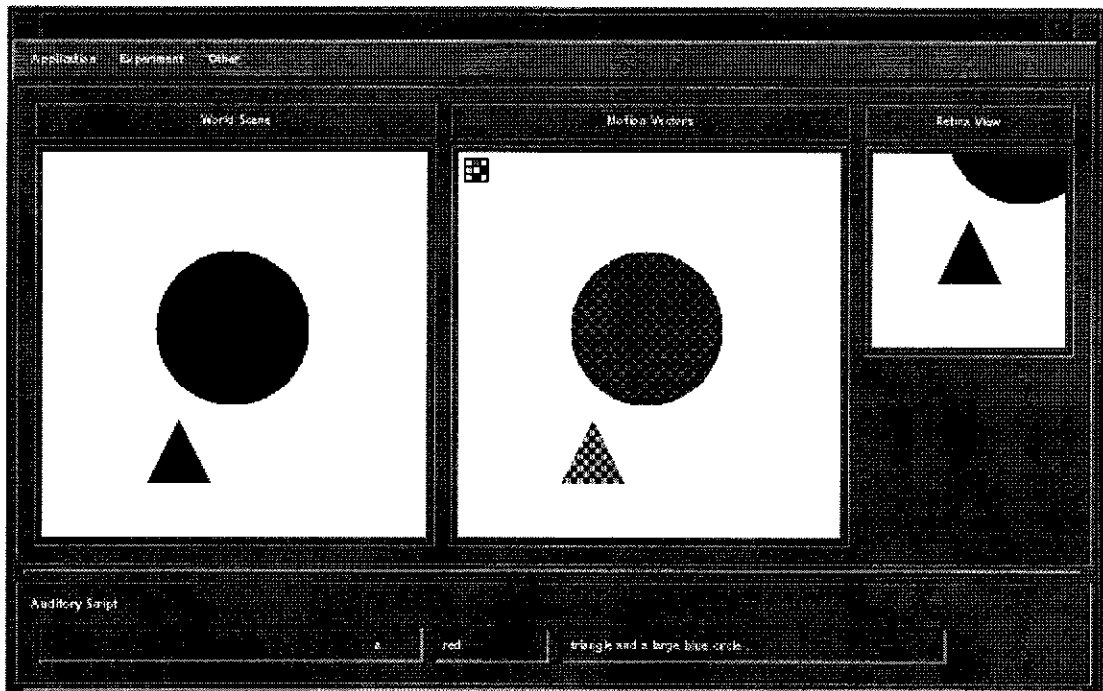


Figure C.9: Experiment Window During Run

Association Window

The *Association Window* is the window which accompanies the *association* process; that is, the process described in Section 3.2.4. Various canvases show the size of the neuron weights, and the process inputs and outputs via colour coding.

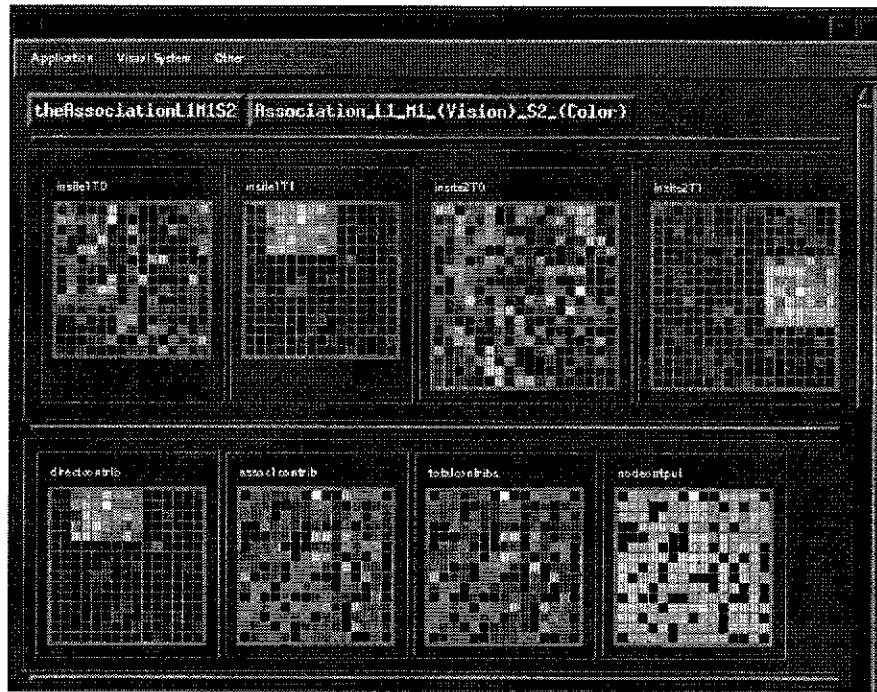


Figure C.10: Association Window

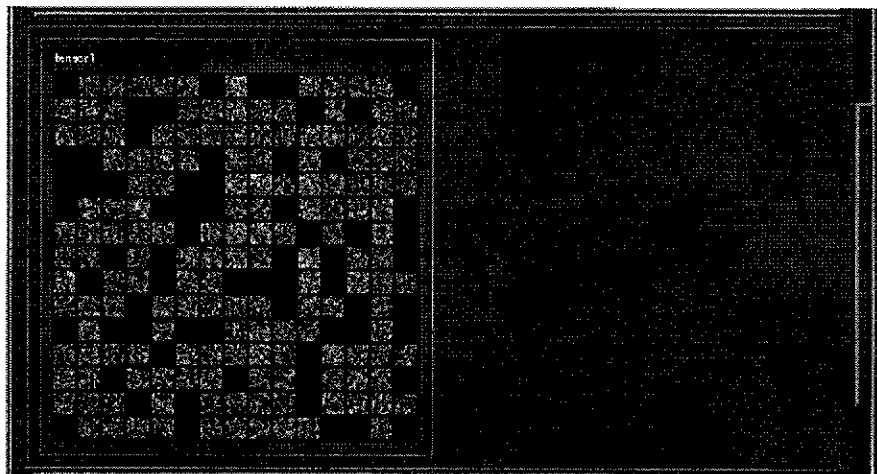


Figure C.11: Association Window (con't)

Kohonen Window

The *Kohonen Window* accompanies each SOM surface in the structure, and provides visual information on the various weights and input vectors found on that particular SOM surface.

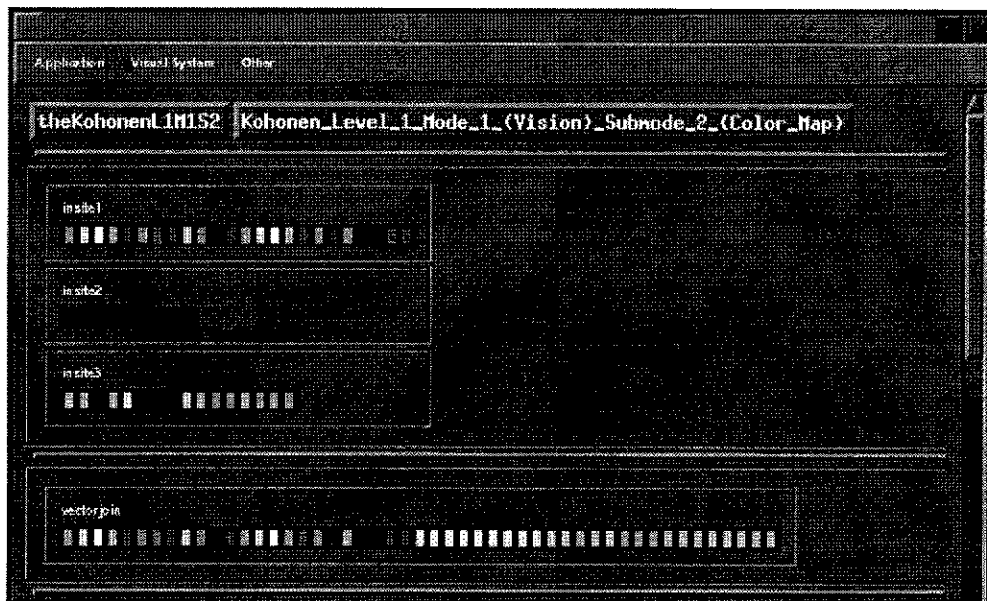


Figure C.12: Kohonen Window

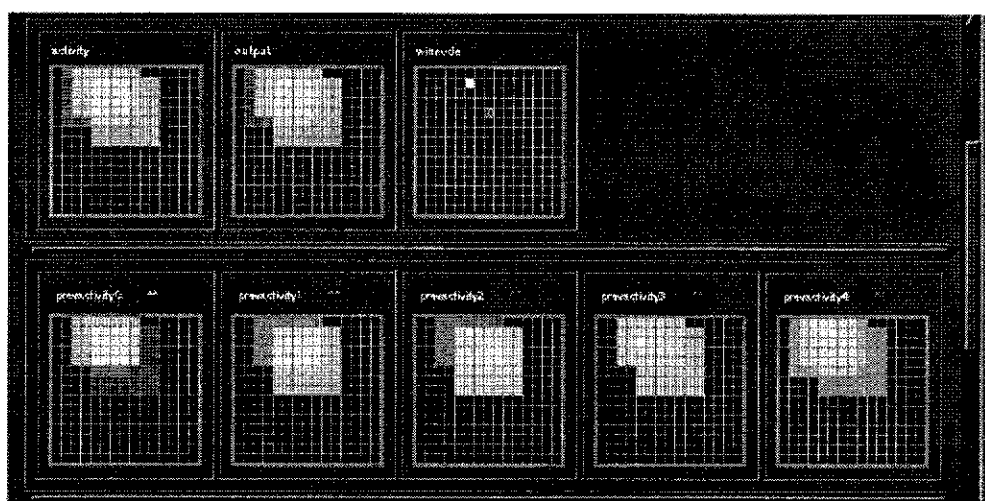


Figure C.13: Kohonen Window (con't)

Retina Window

The *Retina Window* provides visual information on the inputs that are received at the simulated retina. Canvases indicate the various colour and motion components of the original scene. The elements of the three input vectors are also shown.

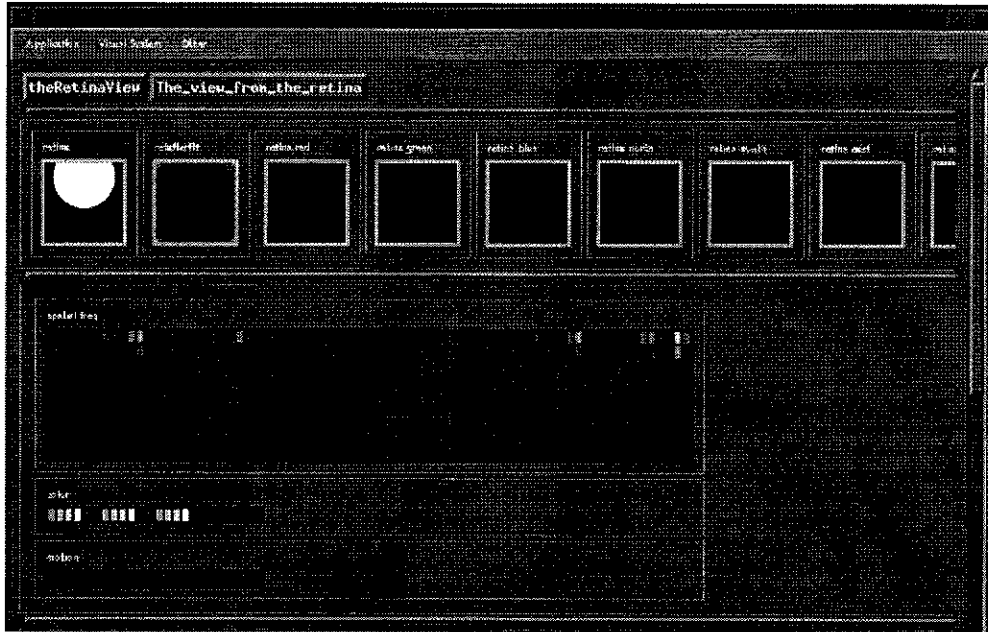


Figure C.14: Retina Window

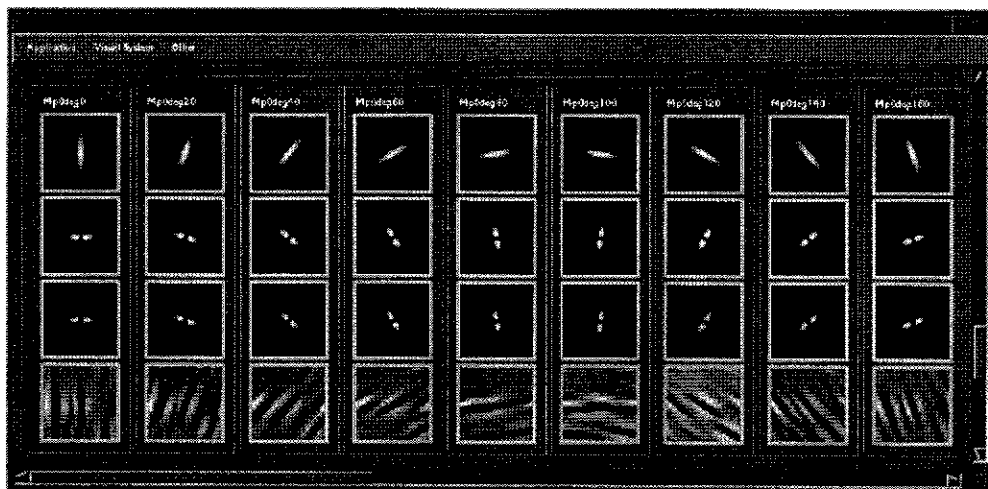


Figure C.15: Retina Window (con't)

Appendix D

Computers and Computation

One of the major claims of this thesis is that the computational theory of cognition (cognitivism) is not appropriate and must be replaced by a theory that does not involve representations and is a ‘hardware’-only model. One such model is the ABC model.

The full ABC model does no “calculations”. The model is more akin to a system of water flowing through variable-width pipes, or electricity passing through circuits with variable resistances, than it is to a computer. The process of cognition results from the learning and reproduction of temporal sequences, and subsequent learned behaviours. An appropriate metaphor might be a flow—a flow of behaviours (temporal sequencing) resulting from flows of vectors across self-organizing maps and the recurrent return of vectors to previous maps.

Serial (and current parallel) digital computers do not qualify as models of cognition for many reasons—reasons that we discuss in this appendix. They simply do not have the same overall properties as the brain.

In this appendix we examine a number of issues concerning computers, computation, representations, theories and cognition.

D.1 Binary (von Neuman) Computers

There is an obvious question to ask of the cognitivists—are the observed behavioural properties of the brain similar in any way to a computer with the *von Neuman architecture*, the architecture common to all binary computers in use today?

The question would seem to be answered in the negative. For example, Hertz et al. (1991, page 1) suggest that:

Anyone can see that the human brain is superior to a digital computer at many tasks. A good example is the processing of visual information: a one-year-old baby is much better and faster at recognizing objects, faces, and so on than even the most advanced AI system running on the fastest supercomputer.

The brain has many other features that would be desirable in artificial systems:

- it is robust and fault tolerant. Nerve cells in the brain die every day without affecting its performance significantly.
- It is flexible. It can easily adjust to a new environment by “learning”—it does not have to be programmed in Pascal, Fortran or C.
- It can deal with information that is fuzzy, probabilistic, noisy, or inconsistent.
- It is highly parallel.
- It is small, compact, and dissipates very little power.

Only in tasks based primarily on simple arithmetic does the computer outperform the brain!

The fact that a computer may be better at any one particular operation, such as finding a correlation between two signals, is irrelevant. In those behaviours that are considered to be important to humans, such as recognising faces or using language, the overall behaviour of humans is far superior to that of current computer systems.

Many differences between the two models have been listed in numerous publications, so we simply list a few without much comment.

single-point failures

The von Neuman model is vulnerable to single-point failures—anyone who has used a computer at all will realize that a single bit reversal can bring a computer to a halt. Unfortunately, this is an all too common occurrence with digital computers, and the computer model does not allow for any redundancy or recovery in this regard.

On the other hand, the brain (and the ABC model) does not suffer from single-point failures. Lesions, aging and even destruction of cells due to drugs and cosmic-rays may not effect performance in any noticeable way, or may elicit only a gradual decline in performance over time if continual destruction occurs. In general, lesions result in a degrading of memories rather than the loss of specific facts as would be expected with a computer. The effects of brain lesions on memory functions suggest that the principles of information storage in nervous systems are unlike those used by computers, and do not exhibit the brittleness of binary computers.

bottleneck at the CPU

The reliance on *control* mechanisms in the von Neuman architecture means that there is a bottleneck at the central control point—the central processing unit or CPU. All processing must pass through this one control point—all calculations, all retrieval of memory, all input and output. Although current computers divest the central CPU of much of the communication to peripheral devices, and so-called parallel computers provide multiple CPU devices to share the load, there is still a requirement for centralised control and synchronisation.

As well as the CPU being a bottleneck which slows down processing speed by enforcing strict serial processing at each CPU, there is absolutely no evidence that the brain has such a central point of processing. All evidence from brain scanning devices suggests that the flow of activity within the brain is much more diverse and holistic, spreading over large areas of the brain.

speed of operations

Although the modern computer has phenomenal speed of individual operations in comparison to the relatively slow pace of the separate neuron firings in the brain, the computer is still unable to match the brain in many cognitive tasks, such as face or voice recognition.

As we state elsewhere, the fact that recognising Granny take approximately 150-200 msecs, and involves some 100-200 synaptic joins, while a typical computer vision system will perform millions or perhaps billions of calculations and still fail to recognise a picture of Granny in any reasonable time (given realistic conditions), indicates that something fundamentally different is occurring in each case.

There are many other differences, but we need not go into further discussion here. Some of the differences are summarised in Table D.1.

D.2 Representations

[N]othing seems more possible to me than that people will some day come

Brain	von Neuman Computer
slow components (neuron firings)	fast components (registers)
short memory (7 items)	long memory
good at Gestalt	not good at Gestalt
relatively imprecise	precise, accurate
not brittle	fragile, brittle
massively parallel	serial processing (some multiple-serial)
self-organised	programmed
recognise Granny in 150 msecs	can't yet recognise Granny
recognise Granny in approximately 150 synaptic joins	can't recognise Granny after billions of calculations
context-based	search and representation
expertise entails shorter reaction times	expertise (more knowledge) entails longer reaction times (combinatorial explosion)

Table D.1: Differences Between the Brain and a von Neuman Computer

to the definite opinion that there is no copy in the . . . nervous system which corresponds to a *particular* thought, or a *particular* idea, or memory.

(Wittgenstein [1948] 1982)

Computers manipulate representations—things that *stand for* something else. For example, on an abacus, some of the wooden beads stand for units, some stand for tens of units, and so on. In a computer, binary patterns stand for objects, events or concepts. A register of a 32-bit computer could contain the bit pattern shown in Figure D.1. This binary pattern could stand for the number 0.5 if it is interpreted as a real number (1 sign bit, 8-bit biased exponent, 23-bit mantissa normalised), or it could be the integer 1,077,936,128. It could also be a randomly chosen code to represent an employee number, or the address of another computer—in fact anything. Computer codes only *mean* something through some semantic interpretation; that is, members of a particular society (those who program this type of computer) agree that the binary number in this situation means the real number 0.5.

Representations need an *interpreter* and in the case of an abacus it is obviously a human. But a human is also needed in the loop to interpret the binary bit patterns, so that suitable rules of calculation can be built into a computer. This person may not be you as the codes may be subsequently presented on a computer screen in the form of another representation, one more appropriate for general consumption, such as the patterns of light shown in Figure D.2.

This screen representation will allow most humans to interpret it as standing for the number ‘three’. But someone, perhaps the program’s developer or the compiler programmer, made the conversion from the representation external to the computer (perhaps a handwritten note leading to the typing at a keyboard of the characters ‘0’, ‘.’ and ‘5’) and the internal code shown in Figure D.1.



01000000010000000000000000000000

Figure D.1: Computer Binary Representation.

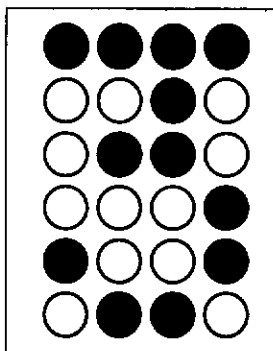


Figure D.2: Representation of the Digit ‘Three’ on a Computer Screen.

All of these forms of representations (binary computer codes, handwriting characters, abacus beads) and others (pictograms, signs) are *inventions* of the human brain, and the representations (in the form we recognise) are found external to the human brain. They are a product of a society.

But if the brain is a computer as suggested by the cognitivists, how were the ‘binary codes’ of the brain selected? Are we a 32-bit or a 64-bit computer? Do we use ones-complement or twos-complement real number representations? Are we all ASCII or are some of us EBCDIC?

On a more serious note, what is the form of the translation from the external representations to the internal ‘binary’ codes? Are the internal codes for humans all the same (and thus innate) or are they randomly chosen for each person. Surely though, the instruction codes for each person must be the same, regardless of the data codes.

To take the discussion further is simply folly. One needs to postulate innate computational structures and codes for humans, based on linguistic terms (which are not shared by other animals), in order to find any basis for the cognitivist model.

Representations in the sense of *standing for* are human inventions used to improve and extend communicative behaviours between members of a society. They are not internal mechanisms of the brain. In the same way that a sunflower does not need to have a representation of the sun within its physical extent in order to follow the sun across the sky, so humans do not need representations within their brain in order to behave.

D.3 Rules and Computers

“A distinction must be drawn between activities that are actually governed by explicit laws, and performance that is merely describable in terms of rules” (Mixon 1991, page 27).

Mixon’s comment suggests that there are two ways of *obeying* a rule: (a) the rule is actually inscribed inside the rule-governed device (for example, an IF/THEN/ELSE rule within a computer)—when needed, the rule is looked-up and applied; and (b) a *phenomenon* can *obey* a rule (a “law”) if we observe that its behaviour aligns with the rule.

These two cases represent the distinction between the *instrumentalist* and the *realist* positions in the philosophy of science. Instrumentalism has been defined by Popper ([1956] 1983, page 111, emphasis in original) as:

... the doctrine that a scientific theory, such as Newton’s, or Einstein’s, or Schrödinger’s, should be interpreted as an instrument, *and nothing but an instrument*, for the deduction of predictions of future events ... and more especially, that a scientific theory should not be interpreted as a genuine conjecture about the structure of the world, or as a genuine attempt to describe certain aspects of our world.

In contrast, the realist claims that theories must be taken as a true description of reality, while not denying that they also provide predictive and classification mechanisms.

There is a large difference between these two. In the first case, the rule constitutes a *component* of the entity; that is, the rule is an internal part of the structure or makeup of the entity. In the second case, the rule is only an *approximate* description of the behaviour of the entity expressed in some suitable language of convenience (usually in mathematical terms). The entity does not directly ‘obey’ the rule at all, but at a level of description (to a certain approximation) we can use the rule to predict the subsequent behaviour of the entity. In this case the rule is *external* to the entity and an ‘invention’ of humans.

The only entities that we know of with built-in rules are made by humans, such as binary computers and Jacquard looms, and these use some form of arbitrary codes to convert human instructions into behaviours.

This thesis takes the view that rules, computation, search and representations are not part of the process of cognition, but rather a result. They are learned behaviours. As humans we may use these tools in our evaluations and predictions of cognitive processes, and we use other tools such as pencils and computers to perform calculations using these rules.

Further, it is important that we not confuse the actual with a simulation. Clearly a simulation of the weather is very different from the behaviour of the many forces and geographical influences that come together to determine our weather patterns. People performing simulations invariably need to make simplifying assumptions, and must trust that the representations and rules that they employ are appropriate and accurate. The use of computation in a simulation only allows us to make *predictions* about natural phenomena.

A further issue is that simulating a process may not model the process very well—the weather may be modelled, but due to chaotic and other effects, the predictions achieved may not necessarily be accurate. And as we mention elsewhere, the simulation of inherently parallel processes on a serial digital computer is prone to combinatorial usage of time and computer resources.

Design engineers may become very good at predicting the behaviour of aeroplanes from simulations, but it is only when an actual physical plane is built and flown that we can be sure that the simulation is correct.

Computation is a human invention following on from counting. Computation is performed on representations—for example the representation of the ‘natural numbers’ 0, 1, 2, 3, . . . , or other squiggles on a page, or binary bits in a computer register, or the typed characters of a computer language, or wooden beads on an abacus, and so on.

Computation requires a semantic interpretation over and above the physical components of the device performing the computation. An abacus can be a computer if we

allow the wooden rings to represent numbers—some representing units, some representing tens of units, and so on.

D.4 Assumptions of Cognitivism

Dreyfus (1992, page 156) suggests that there are a number of pivotal assumptions that are made by cognitivists which must be questioned, and he goes on to demolish each of these assumptions. The highlights of his discussion are reproduced and extended here, and the criticisms examined in relation to the ABC model.

These assumptions, which form the basis of cognitivism, result from a powerful conjunction of the Platonic assumption that the formalism which enables us to *understand* behaviour is also involved in *producing* that behaviour, and the Kantian assumption that all orderly behaviour is governed by rules, both reinforced by the idea of a computer program (Dreyfus 1992, Winograd & Flores 1986).

biological assumption:

The first assumption Dreyfus calls the *biological assumption*. This is the assumption, required by the computer model of cognition, that the brain processes information in discrete (on/off) operations.

Dreyfus argues that the brain is in fact an analogue rather than a binary device, with graded synaptic potentials and impulses—the evidence is against the binary view. He suggests that the view of the brain as a general-purpose symbol-manipulation device operating like a digital computer is an empirical hypothesis which has had its day.

The ABC model requires no such assumption. Input vectors are not required to be binary, and the model proposed is indeed a dynamic analogue device.

psychological assumption

The *psychological assumption* posits that the brain (mind) can be viewed as a device operating on bits of information according to formal rules. This assumption justifies the use of the computer model in psychology.

This view derives from the cognitivist view that there is a level—the information-processing level (or symbol-processing level)—at which the *mind* uses computer pro-

cesses such as comparing, classifying, searching lists, and so on, to produce intelligent behaviour. The level is assumed to be between the neurological and the phenomenological levels.

The reasons for postulating this level were several, including:

- The success of the *levels hypothesis* in the hard sciences. This hypothesis held that nature and complex systems may be observed at a number of levels in a hierarchy of understanding. For example, the ‘laws’ of chemistry are all (in theory) able to be deduced from the laws of physics, but it is appropriate to treat them as two distinct disciplines. Likewise, computer scientists suggest that there are a number of levels in the understanding of computer systems.⁴³
- Intellectualism, in which the ‘pure’ sciences (such as physics), with their ‘precision’ and mathematical approach, were considered superior to the messy ‘soft’ sciences such as biology and the social sciences. This led to a process of physics-envy and an attempt to elevate the softer sciences to the hard variety.
- A ‘what else can we do?’ attitude exemplified by Fodor’s comment that cognitivism is ‘the only game in town.’

We are obviously not suggesting that one should not seek descriptive rules in any science, but that it is not appropriate to postulate a symbol-processing level unless one can provide a linking mechanism. As we will discuss in the following paragraphs, there is no real justification for the hypothesis, just a vague hope that eventually the symbol-processing level will be linked to a theory at the neural level.

The first thing to note is that experience is not a series of isolated, atomic alternative choices, but a dynamic kaleidoscope of events. As we have mentioned elsewhere, algorithms have completely the wrong characteristics to describe cognition.

But the main reason for rejecting the hypothesis is that it misrepresents a misunderstanding of the achievements of physics and other sciences in formulating ‘rules’. As Dreyfus (1992, page 167) states:

The argument gains its plausibility from the fact that if a psychologist were to take the first derivative of a texture gradient, he would compute it using a formalism (differential calculus) which can be manipulated in a series

of discrete operations on a digital computer. But to say that the brain is necessarily going through a series of operations when it takes the texture gradient is as absurd as to say that the planets are necessarily solving differential equations when they stay in their orbits around the sun, or that a slide rule (an analogue computer) goes through the same steps when computing a square root as does a digital computer when using the binary system to compute the same number.

Rules are *descriptions* of behaviour, not the *mechanism*. To some level of approximation, a rule allows us to measure and predict subsequent behaviours, but the rule *is not the same thing* as the behaviour. In stating a rule, we make no comment about the actual mechanisms of the behaviour—natural phenomena *do their own thing*. We observe the regularities of the behaviour of objects and events in the world, and to a certain level of approximation we are able to describe the behaviour in the form of ‘rules’ or ‘laws’. However, the rules are descriptions of behaviour and nothing else. Let us consider a few examples of natural phenomena—gravitation, the electrical force between two charged particles and the behaviour of an ideal gas.

The equation (rule) for the gravitational force between two mass particles (such as the earth and the moon) was discovered by Newton and published in his *Principia Mathematica* in 1687. The equation can be found in any undergraduate text book:

$$F = G \frac{m_1 m_2}{d^2}$$

where F is the force, m_1 and m_2 the two masses, d the distance between the masses, and G a universal constant having the same value for all pairs of particles.

Similarly, Coulomb’s law for the electrical force between two charged particles (such as electrons) was first expressed in 1785:

$$F = \frac{1}{4\pi\epsilon_0} \frac{e_1 e_2}{d^2}$$

where again F is the force between the particles, e_1 and e_2 the electrical charges on the two particles, d the distance between them and ϵ_0 a constant known as the permittivity constant.

Finally, the pressure, volume and temperature of an ideal gas are related by the formula :

$$PV = nRT$$

where P is the pressure of the gas, V the volume, T the temperature, n the number of moles (a measure of the mass of the gas), and R the universal gas constant. The Ideal Gas Law was discovered by Boyle in the 1600s.

But these rules are only approximate ‘measurement’ rules that have no explanatory value whatsoever. The rules are merely a description of the likely results of measurements, provided these measurements are made under similar conditions to the experimental data from which the formulae were derived.

Today, physicists have a very different picture—a very different description of these natural phenomena.

In the case of electromagnetism, it is thought that the exchange of a virtual photon between two electrons gives rise to the electro-magnetic force between them. This is illustrated in the Feynman diagram of Figure D.3. A photon is emitted by one of the electrons and absorbed by the other.

In the case of gravity, it is postulated that the underlying interaction between two mass particles comes about as a result of the interchange of a virtual *graviton*. In this case, the suspected particle of exchange has not yet been ‘discovered’ by physicists. This means of describing the interactions between elementary particles has continued with the discovery (and theoretical postulation) of many other elementary particles and their quantum particles of exchange, such as the various varieties of quark which interact via other particles called gluons. In fact, it is postulated that “any fundamental force must

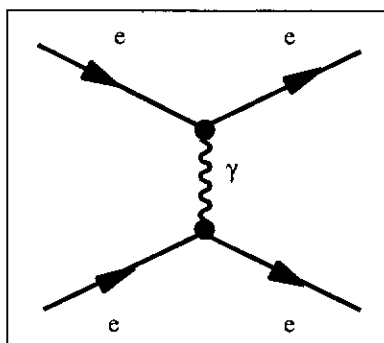


Figure D.3: Electromagnetic Interaction.

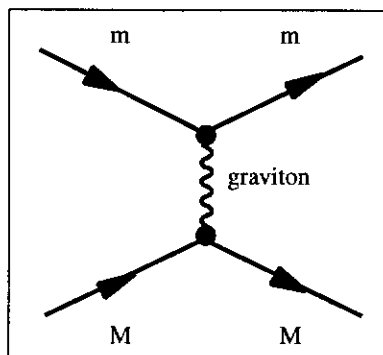


Figure D.4: Gravitational Interaction.

be associated with an elementary particle that is the quantum of the corresponding field. Sometimes the quantum is said to *carry* the corresponding force” (Gell-Mann 1994, page 124).

In the case of the ideal gas, the ‘rule’ only comes about as a result of the mean effect of interactions between the many particles within the gas.

It may be that we can delve deeper and deeper into natural phenomena and never get to the ‘final cause’. The universe may be fractal and infinitely divisible—inside molecules are protons, inside protons are quarks, inside quarks are

The point is that these laws are all *descriptions* of observed behaviours of phenomena, not the phenomena themselves. The rules are invented by humans to describe the phenomena and to be able to measure and predict the phenomena. The actual description may depend upon the degree of ‘accuracy’ involved in the measurements, or on the region of measurement (such as the need for high velocities before relativistic effects are noticed).

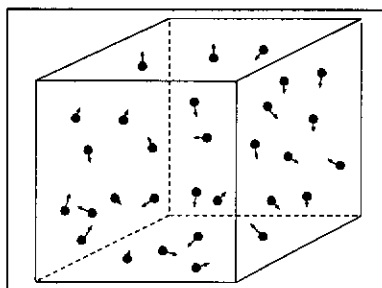


Figure D.5: Particles in a Gas.

We humans may measure and simulate these phenomena with computations, and we can estimate the likely outcome of some event using our rules and calculations; for example, we can compute the volume of a gas under certain conditions of temperature and pressure using the ideal gas law.

But at another level of approximation the particles just collide and bounce off each other—at another level of approximation their wave functions interact and interfere with each other—at another level the elementary particles that make up the molecules of the gas exchange quantum particles.⁴⁴

It is simply absurd to suggest that natural phenomena need to ‘compute’ their behaviour according to rules. In no way does the weather need to solve multiple differential equations in order to move to the next instant in time. Does a rock calculate when and how to erode—does an iron bar calculate how much to expand or contract with variations in its temperature?

The psychological assumption is that the rules used in the formalisation of behaviour are the *very same rules* which produce the behaviour.

We are able to simulate natural behaviours (to a certain level of approximation) on digital computers. However, “we need not conclude from the fact that all continuous physiochemical processes involved in human ‘information processing’ can in principle be formalized and calculated out discretely, that any discrete processes are actually taking place” (Dreyfus 1992, page 168).

Dreyfus provide empirical evidence that the psychological assumption is simply not tenable. The evidence includes the failure of the original research program by Newell and Simon to stand up to close inspection, and the failure of subsequent research in this area to live up to the exaggerated claims made by early practitioners.

The ABC model makes no such assumption. The model is a dynamical system in agreement with other natural phenomena. Although we are able to simulate the model on a digital computer (and do so as described elsewhere in this thesis), the model itself involves no calculations in accordance with any rule. The behaviour of the model is determined by the dynamic interaction of its components and its surroundings.

The ABC model uses statistical means to learn and reproduce the behaviours of the creature, not determinate rules.

The cognitivist approach holds sway in psychology today, but we feel that this has

been a backward step from the behaviourist approach. While most behaviourists were incorrect in not allowing for any discussion on the contribution of the brain to behaviour, the cognitivist approach is seriously flawed in its basic premises. The assumption that the brain is a computer running computer programs is unsustainable. Simply finding some formula to link inputs and outputs, and then to put that into a computer program as the mechanism of the behaviour, is neither adequate nor appropriate. The approach is just peace-meal approximation, and bears no relationship to the actual processes involved. Computers are a valuable tool for model simulations, but they are not the underlying mechanism of cognition.

epistemological assumption

The *epistemological assumption* suggests that all knowledge may be formalized; that is, whatever can be understood can be expressed in terms of logical relations (Boolean functions).

This is the assumption used by the cognitivist artificial intelligence (AI) community. Although they may concede that “human performance might not be *explainable* by supposing that people are actually following heuristic rules in a sequence of unconscious operations, intelligent behaviour may still be *formalizable* in terms of such rules and thus reproduced by machine” (Dreyfus 1992, page 189).

Thus AI might be able to use rules, even if human and animal cognition does not. After all, we can simulate other things successfully on a computer; perhaps we can simulate intelligent behaviour.

This is a weaker assumption than the psychological assumption, and may still have problems due to the characteristics of digital computers, but we may be forced to concede that given enough memory, computer power and time, a von Neuman computer could simulate (emulate) intelligent behaviour of any isolated kind (for example, the recent success of the IBM Deep Blue computer in defeating the world chess champion). But it is very difficult to accept that a digital computer could produce overall intelligent behaviour in a dynamic world to the extent that it could pass a general Turing test.

The epistemological assumption is an assumption of *competence* rather than actual *performance*. As pointed out by Dreyfus (1992, page 190): ⁴⁵

A man riding a bicycle may be keeping his balance just by shifting his weight

to compensate for his tendency to fall. The intelligent content of what he is doing, however, might be expressed according to the rule: wind along a series of curves, the curvature of which is inversely proportional to the square of the velocity [(Polanyi 1962)]. The bicycle rider is certainly not following this rule consciously, and there is no reason to suppose he is following it unconsciously. Yet this formalisation enable us to express or understand his *competence*, that is, what he can accomplish. It is, however, in no way an *explanation* of his *performance*. It tells us what it *is* to ride a bicycle successfully, but nothing of what is going on in his brain or in his mind when he performs the task.

Dreyfus (1992) counters the epistemological assumption by noting that it involves two claims: (a) that all behaviour can be formalized, and (b) that the formalisation can be used to reproduce the same behaviour. These claims are refuted in turn (Dreyfus 1992, pages 190–205).

One of the problems with the epistemological assumption is the issue of formulating a theory of human performance and of common sense. We examined this issue in Section 5.4.

Another problem was stated by Wittgenstein, who argued that it is impossible to supply normative rules which prescribe in advance the correct use of a word in all situations. This is the problem of context that we look at in the following section.

The ABC model again does not resort to ‘rules’ and so is not limited by the epistemological assumption.

ontological assumption

Dreyfus’s (1992) *ontological assumption* states that all relevant information about the world must be analysable as a set of situation-free determinate elements; that is, that all facts about the world are logically independent of each other. As stated by Dreyfus (1992, page 210):

Even a chair is not understandable in terms of any set of facts or “elements of knowledge.” To recognise an object as a chair, for example, means to understand its relation to other objects and to human beings. This involves a whole context of human activity of which the shape of our body, the institution of furniture, the inevitability of fatigue, constitute only a small part.

And all these factors in turn are no more isolable than is the chair. They all may get *their* meaning in the context of human activity of which they form a part.

The philosophers Heidegger, Wittgenstein and Merleau-Ponty bring this assumption into question. Merleau-Ponty calls the assumption that all that exists can be treated as determinate objects, the “presumption of common sense”. They show that we do not experience the world as a set of facts in our everyday activities. Nor is it self-evident that it is possible to carry through such an analysis. The context of the current situation must be taken into account.

As stated by Dreyfus, “the appeal to context would seem to be more fundamental than the appeal to facts, for the context determines the *significance* of the facts.” For humans, the present situation is a continuation or modification of the previous one, and so we carry over from the immediate past a set of anticipations based on what was relevant and important a moment ago. This carryover gives us a certain predisposition as to what is worth noticing. We do not need to re-cognise at each instant as implied by the context-free ontological assumption.

D.5 Computation by Humans

Does the human brain calculate? The obvious answer is yes, it does—one plus one equals two, and the square root of nine is three. A human brain (that of the author) has performed what we would all agree is a calculation—in fact two calculations.

But the real question is—were calculations of the sort performed by a computer (or a calculator, or an abacus) used in the *process* of that performance, that behaviour? For example, is there a register in the head, along with a program of some sort to *parse* the input symbols, adding the appropriate numbers into the register, examining the register and looking up the appropriate sound or written symbol to express the output, then finally producing the appropriate symbol as the result? Further, was a similar process required to determine that *two* calculations were in fact performed?

The ABC model proposes that the answer to this question is a resounding no. The so-called calculations as performed in the brain were instead re-performances of learned behaviours. During our school years, we learn to perform certain tasks such as calculating with numbers. The calculations are behaviours that are performed on symbols which are (if we ignore self-talk), external to the brain.

If we include self-talk, then multiple calculation-behaviours may be carried out within the brain, giving the illusion of a calculating device. But each process in the so-called calculation is simply a reproduction of a previously learned temporal sequence—for example, ‘one plus one equals two’.

In other words, the *result* is a ‘calculation’, but the *act* does not involve any ‘neural calculation’ to achieve that result.

What happens in a *hardware* version of the ABC model, and in the brain, is more like a table-lookup than a calculation—a table-lookup within a large, dynamic and recursive table. The process involves none of the things that a von Neuman computer would do; no stored program, no registers, no off-line memory.

People have a ‘look-up’ behaviours for square roots for those numbers that they have learned—like 4, 9, 16, and so on. Before the advent of computers and calculators people had a ‘square root’ behaviour that used pen and paper and an iterative formula. These days our ‘square root’ behaviour is to push a few buttons on a calculator, or a few keystrokes on a computer keyboard to run our favourite spreadsheet.

And the determination that a ‘calculation’ was performed twice is no more than a process of moving through a learned sequence (the counting numbers), matching each thing to be counted with the current term in the sequence, and moving to the next in the sequence for the next item—again a behaviour learned in childhood.

People who appear to have phenomenal calculation capabilities are simply using behaviours that they have learned in addition to the ‘calculation’ sequences used by the majority of people.

Although ABC has no explicit rules for performing calculations, it can learn behaviours which is similar—upon semantic interpretation—to that which would have been pro-

duced had the rules been explicitly incorporated.

In the same way that mass particles behave in a way that approximates the universal gravitational rule (this is the wrong way around—the reality is of course that the rule is an approximation to the actual behaviour), so too the ABC model behaves in a physical way that approximates the so-called rules implicit in the training examples.

D.6 Artificial Intelligence

The question of whether artificial intelligence is possible must be answered in the positive. There is no apparent ingredient in human brains and bodies that is not reproducible in some form in some material. The more interesting question is just what form this artificial brain will take. The various discussions in this thesis suggest that, except in limited domains (for example, chess, and other situations where explicit rules may be devised), the cognitivist approach will not succeed.

The ABC model represents a possible alternative approach. The need for adaptive, situated, context-dependent behaviour to enable any artificial creature to interact in a dynamic world requires features found in the ABC model. We suggest that any model of cognition will need to have similar properties and will need to have a similar architecture.

A computer simulation of the ABC model *as a working entity* is not appropriate, as the software model of ABC is slow on a serial computer. †

However, as a hardware implementation, the ABC model should be able to include the much larger numbers of neurons and interactions needed to potentially interact and behave intelligently in real-time. The confirmation of this is left to future research.

†For example, the model required some 3 days of processing on a reasonably fast Silicon Graphics computer to perform the necessary calculations for 800 neurons over 300 epochs. This is somewhat less than the 10^{11} neurons in the human brain, and hardly real-time processing!

D.7 ABC is Hardware-Only

The ABC model is a *hardware-only* theory of cognition. Although it is able to be emulated on a computer, the ultimate goal of this research is to build a hardware-only version of the model.

Many computer scientists mistakenly believe that hardware and software are in some way equivalent. This view possibly comes about as a result of their knowledge of micro-code (specialised software that can change the actual hardware characteristics of certain computers), as well as their knowledge that one can always write software to emulate some particular piece of hardware. However, the major premise of the equivalence of hardware and software is clearly false—hardware only devices do exist (clocks, automobiles, the ABC hardware model) whereas software-only systems are impossible. The equivalence is thus only in one direction—from hardware to software—the process of simulation (or emulation).

We look at a *simulation* of a proposed ABC *hardware* model of cognition in Appendix C. But this should not be confused with the actual ABC itself. The hardware model incorporates only neuronal processes and does not require any ‘computer’ capabilities.

The ABC is *not* a *computational model* which pre-supposes that the actual processes of cognition are computational. As we state elsewhere, the ABC model suggests that cognition is a process of incrementally learning appropriate temporal behaviours rather than computation. The ABC is rather a *dynamic* physical object, as is the brain, the heart, the solar system, a tree and an ant.

Appendix E

Vision Background

In this appendix, we briefly examine a number of issues relating to previous and current models of vision. The presentation relates mainly to views expressed in the computer vision literature.

E.1 Current Theories of Visual Recognition

Kuhn (1970), in his study of the history of scientific ideas, suggested that practitioners of an established scientific ‘paradigm’ adopt certain canons and beliefs that once established, are no longer questioned. These base-line beliefs tend to become preconditions that underpins the discipline. In this section, we look at some of the underlying views currently pervasive in vision research. A number of broadly shared assumptions pervade research into the study of the human and animal vision system, and in particular the modelling of that system in computer vision. In this section we provide a cursory examination of the current state of vision research, and in particular the assumptions that are made. More details may be obtained from review and introductory articles, such as Pinker (1986) or Grimson (1990).

The information processing paradigm is the basic intellectual scaffolding underpinning both vision research and most other disciplines within cognitive science and cognitive psychology. We deal with our main objections to this fundamental assumption in

another section, and so we will only discuss in this chapter particular objections to the computational approach that arise in relation to the visual system.

Current computer vision systems are surprisingly inept, and have yet to achieve a level of sophistication found in infants. People are able to perceive and move around in a dynamic world without any difficulty, but computer vision systems, in the main, require extremely limited and static environments, usually requiring large amounts of computer resources and background knowledge.

The issue of vision has traditionally been divided into two subtopics: visual recognition, including the representation of visual information, as well as retrieval and matching; and visual imagery, the remembering or reasoning about visual aspects in memory.

There is no doubt that the understanding of vision is a very complex problem. We touch upon some of the issues in Section 4.1.1. Foveation, saccades, the distribution of rods and cones in the retina, the motion of the body and other objects, the blurring in peripheral vision, variable focus of the lens, occlusion, variable lighting and shadows, reflections, surface markings and scratches all present difficult problems for a computer-based model to overcome.

Most theories of shape recognition postulate some form of internal representation (or a set of representations) for each object. The representation contains information about the shape and other properties of the object, including a label—a name for that object. The aim of the system is to be able to correctly retrieve that label during recognition. Representations are stored in long-term memory as separable, (usually symbolic) wholes, one (or a set) for each object or class of object. Some form of *learning* of the representations by inductive methods is currently being explored (see, for example, Dillon 1996).

Current models usually do not postulate a representation which is a direct replica of the retinal stimulation. Rather they introduce some form of representation which attempts to capture the supposed invariant properties of each object in various positions, sizes, rotations, and even under various lighting conditions. During recognition, the retinal image corresponding to the unknown object is converted to the same format, and the representation that provides the best match using some form of similarity measure is

taken to be the object recognised.

Each theory may make different assumptions regarding

- the type of representation used (e.g., feature space, geons, predicates, graph),
- the number of representations per object (e.g., store one composite 3D representation or multiple 2D representations obtained from different viewing positions),
- the breakup of objects into classes for mapping into representations,
- the inclusion of spatial relationships between objects and their component parts (relational or propositional representations),
- the amount and type of preprocessing given to the initial retinal image (e.g., various forms of edge detection, filtering, and contrast enhancement),
- how the matching is to be performed (e.g., sub-graph isomorphism, decision trees, feed-forward neural net, etc.),
- a suitable metric for similarity matching.

Most adherents to the computational view of visual cognition would claim that the primary issues in computer vision (as well as other artificial intelligence domains) are representation and search—how to form an appropriate representation for objects, and how to then efficiently search these representations for a match at recognition time. In the next few sections we briefly examine various representations used in the majority of traditional theories.

E.1.1 Template Matching

Template matching is the simplest form of representation in which a replica of the retinal stimulation pattern projected by a shape is stored in long-term memory. The recognition process simply compares all stored object templates with the input array, selecting the best match based on, say, the ratio of matching to non-matching points.

There are, however, many problems with this method, including:

- partial matches can give false results (e.g., an ‘O’ in a ‘Q’ template),
- any change in the distance, location or orientation of the input object in relation to the corresponding stored object will produce a different pattern, thus preventing recognition,
- any occlusion, shadow or other distortion of the input object may also produce inaccurate matching.

Some systems, for example, attempt to compensate for these problems by storing multiple templates, each recorded at various displacements, rotations and sizes. However, the combinatorics of the transformations usually prove to be unwieldy. The option of rotating, displacing or scaling of the input pattern to a canonical form before matching is also not feasible, as the required transformations cannot be known until the object is recognised.

Other problems with this representation include a difficulty in handling depth information, and the issue of interpolating between stored representations, especially when rotations cause changes and distortions in the object shape.

But by far the major problem with this representation (and in fact most of the representations that we will discuss) is that it is only appropriate for an object recorded in isolation. If multiple objects are present in a scene (and there are very few realistic situations in which this is not the case), then the template models fall down. The method is unable to determine which bits belong to which object—without, obviously, first recognising the objects.

It is also not reasonable to postulate some pre-process which is able to segregate figure and ground by employing depth information to separate objects. This is not only too complex, but also presupposes the identification and existence of separable objects. The differences in intensity supposedly due to depth (or a change in material) may in fact be due to differences in orientation, pigmentation, shadows, surface scratches or specular reflections. It is only by presupposing the existence of particular objects that these sources of intensity change can be separated. *Depth from shading* and other ‘depth from . . . ’ methods, while providing valuable cues, are insufficient in themselves

to separate figure and ground. † Most systems which purport to separate figure from ground are rather artificial and rely on flat lighting and relatively neutral backgrounds (ground).

E.1.2 Feature Models

Instead of storing templates for entire shapes, the feature model method utilises a series of mini-templates or *feature-detectors*. Generally the features included are of a geometric type, such as vertical and horizontal lines, curves and angles. Feature detectors may be located at every position in the input array, or more global feature detectors may be used.

In the case of multiple feature detectors, a calculation is made of the degree of match between the target feature and each section of the input array. The levels of activation for each feature may then be summed across the input array, or the number of occurrences of each feature counted, thus providing a set of numbers, one for each feature. This list of numbers, in the form of a vector of weights for the different features, or a list of the feature occurrences, is used as the stored representation of the object. The intention is to give each shape an invariant representation as each feature is independent of location.

Recognition again consists of finding the best match between the stored representations and the levels of activation of the feature detectors in the input. The comparison to find the best match may use the product of the two vectors, or simply count the number of matching features minus number of mismatched features. The object with the highest match is the shape recognised. ‡

†It is also reasonable to ask just what is meant by figure-ground. Most real-world scenes, which present no problem to human and animal observers, are made up of numerous components at various depths. Consider the case of observing a bird in a tree—what is the figure and what is the ground? The option of recognising objects that are isolated on a neutral background is simply not realistic.

‡Perhaps the original feature model was the “Pandemonium” system, an insightful and influential early neural model of shape recognition (Selfridge 1958). Low level “computational demons” attempted to detect boolean combinations of features. The outputs from these demons were in turn used as inputs to “cognitive demons” whose task it was to *weigh the evidence*. Each cognitive demon passed on a weight (a *shriek* in the words of Selfridge) to the final “decision demon” which determined a winner.

Some models first build up combinations of features into ‘parts’, and then these parts are combined in some form to achieve recognition. †

There are a number of problems with this model:

- the representation is not invariant to scale or rotation,
- in general, spatial relationships are not considered; that is, the relative locations of features are not recorded, only the presence or absence of the feature—for example in recognising a face, it may be important to determine the position of the eyes relative to the nose and mouth.
- the choice of features may be an issue in that natural shapes are not composed of simple lines or curves.

E.1.3 Fourier Models

In the Fourier representation model, a two-dimensional input intensity array is subjected to a spatial Fourier analysis in which the original array is decomposed into a set of spatial frequency components of various orientations and frequencies (sinusoidal gratings). Amplitude and phase are both recorded for the spectrum of spatial frequencies and angles.

The original input image is thus represented as the sum of the spatial frequency components, and this Fourier transform retains all of the information in the original image (given no restriction on angles and frequencies, and no computational problems such as *aliasing*).

Each shape is stored in memory in terms of its Fourier transform, and recognition proceeds by matching these stored representations with a similarly transformed input image.

In principal these models are similar to template models. When an exact match occurs, then there is no difference between the Fourier model and the template model as

†This begs the question of what constitutes a part—see Section E.3.

the Fourier model has the same (complete) information (in the frequency domain) as the template model (in the spatial domain). In theory no information is lost in the transformation. †

However, when an exact match is not made, the Fourier model is claimed to have better metrics of goodness of fit. For example, to a first approximation, the amplitude spectrum corresponding to a shape that is the same regardless of where in the visual field the object is located. If the recognition matching process is then restricted to amplitude only, ignoring the phase, then the shape recognition is translation independent. Further, any changes in orientation and scale of the object result in simple changes in orientation and scale in the transformed frequency space, allowing some to claim that these changes can be normalised, resulting in further invariance to rotation and scale. However, as mentioned previously, any such normalisation presupposes recognition and is thus in general not possible.

Some form of *pattern* and *texture* recognition may be possible, however, because of the particular peaks and other structures found in the transformed image that correspond to the period of repetition of the pattern. The transform process is also useful in that it separates information about sharp edges and small details from other information pertaining to gross overall shape. Techniques such as edge detector filters and convolutions may be used to tease apart these different details of the original image. The method also overcomes the problem of trying to match blurred edges, wiggly lines, and other slight distortions by means of edge enhancement and smoothing.

While initially giving promise, the Fourier model is not without significant problems.

- The so-called invariance only hold for entire scenes and isolated objects. If a number of objects are found in a scene, then any rearrangement of the objects will drastically alter the transformed image. The frequencies due to particular objects cannot be isolated in the transform, with the component frequencies of all objects simply added together. Thus the transform represents aggregated information for the whole scene, with no way to disambiguate the components.

†Computationally, however, some information is lost due to aliasing and the shift to finite, discrete transformations.

- The inability to separate components of the transformed image makes it difficult, if not impossible, to recognise familiar objects in novel scenes by attempting to match transforms of isolated objects in memory.
- Although the amplitude spectrum contains shape information, and the phase spectrum contains position information, there is no method of combining this information (short of performing a reverse transform to the spatial domain) in order to locate a particular objects at a particular location.
- While any rotation of an object in the plane of the transform will only produce a rotation of the spatial frequencies, any 3D rotation of the object will usually alter the transform significantly.
- The issue of how the original (isolated) object transforms might be initially recorded in biological systems is not addressed.

Psychological research suggests that the visual system does partition the information in the retinal image into a series of channels, each specific to a certain range of spatial frequencies (Campbell & Robson 1968).[†] This perhaps explains why the Fourier transform representation is so accepted, especially by many researchers in psychophysics and visual physiology. However, as pointed out by Pinker (1986, page 10) “... *filtering* the image according to its spatial frequency components is not the same as *transforming* the image into its spectra.”

The filtering alternative has the advantage that the original image may be split into a number of separate sub-images, each filtered at particular spatial frequencies. Each sub-image then allows an analysis of the original at at different scales. The processing is still performed in the spatial domain, but each sub-image array is the result of a bandpass filtering of the original image. This allows a segregation of gross shape from fine detail, but as the analysis is performed in the spatial domain it can still find correspondence between the parts of the representation and the parts of the scene.

While giving some respite, filtering only succeeds in reducing the problem a small

[†]The computer simulation of the ABC model actually uses the Wilson Modified Line-element theory of spatial frequencies and spatial angles (Wilson & Gelb 1984).

amount, and the overall problem of recognition within each array is still at issue. As well, the information obtained from each sub-image needs to be linked together into recognition of the whole object.

E.1.4 Structural Models

Structural models came about because early representations were unable to include *relational* information—information about the relative positions and relationships *between* parts of an object. The representation held in memory is a *structural description*, that is, a data structure which maintains a list (or tree, or graph) of predicates. A simple example might be:

```
eyeshape(X), eyeshape(Y), noseshape(Z),  
distance(X,Y,2.1), between(X,Y,Z), below(X,Z), . . .
```

The arguments of the clauses correspond to parts, while the predicates correspond to properties (features) of the parts and to spatial relationships among them. The representation is often depicted as a graph where nodes correspond to the parts or the properties, and the edges correspond to the spatial relations (Pearce 1996, Minsky 1975, Winston 1975).

The inclusion of explicit spatial relations differentiates the structural representation from feature-only models, allowing a possible solution to some of the issues raised by Minsky & Papert (1969). These issues were initially discussed in relation to the non-linear learning properties of the perceptron, but are generally applicable to any model. They concern the ability to discriminate between open and closed areas, and the inclusion or exclusion of a point in relation to a bounded area.

The structural representation is purported to be able to factor apart the information in a scene without necessarily losing information in it. This enables the representation to not only be able to supply a list of labels for the objects in a scene, but also how they are oriented and where they are with respect each other and to the human observer.

By specifying the shape of the object in one set of logical clauses, and its location,

orientation, size, and spatial relationships with other objects in another set, various spatial reasoning operations may be performed that selectively access the relevant information that pertains to a particular processing operation. Further, the recognition problem may be successively decomposed into simpler subprocesses. Statistical and logical operations may be included to further enhance the recognition process; for example, a designation of what *must* be found in an image to ensure the recognition of an object.

So far so good, but the appeal of this method is actually an illusion. Structural representations simply denote a theory of representation—it tells us nothing of the actual process of recognition. Structural modelling is only appropriate (if at all) *after* recognition has taken place using some other method.

Some may argue that some other process may be used for the recognition of parts, and then structural representations and analysis may be used to ‘assemble’ the parts into recognisable objects. However, this still leaves open the very real problem of finding the ‘parts’. It also leaves open the question of determining a suitable ‘assembly plan’, whether this is taken to be innate (or created by the programmer in the case of computer vision) or learned by some inductive method. Neither problem has been solved to any degree despite numerous attempts.

Other problems with this representation question its biological acceptability.

- The approach is based on a type of formalised semantics. But where do the semantic concepts originate? That is, what is the origin of the predicates? Are they innate? Where do the concepts ‘above’, ‘left’, ‘between’, and so on actually come from?
- Given that the concepts are semantic and hence culturally determined, how does this theory relate to animal vision? Presumably animals do not share these ‘linguistic’ descriptions. And infants?
- How might a new concept be incorporated into the scheme? Or how might existing terms be combined to form new semantic descriptions?

- The issue of *grading* of concepts is not adequately addressed. Human conceptualisation is not atomic, but rather fuzzy and polymorphic.

The use of structural descriptions seems to be driven purely by computational convenience, and has no biological support whatsoever. Further criticisms are based on the general objections to computationalism discussed elsewhere.

E.1.5 Neural Network Models

It would be reasonable to say that most vision models are based on the computationalist, symbol-processing paradigm. There have, however, been a number of neural models applied to various components of the visual system. These include:

- self-organisation of orientation sensitive cells in the striate cortex (von der Malsburg 1973),
- orientation specificity and binocular interaction in visual cortex (Bienenstock, Cooper & Munro 1982),
- stereo disparity (Marr & Poggio 1976),
- visual pattern recognition (Fukushima, Miyake & Ito 1983, Fukushima 1988, Barnard & Casasent 1990).

A number of books have been devoted to the examination of neural networks for vision, such as Carpenter & Grossberg (1992b) and Mammone (1993). However, the solutions proposed are generally for isolated components of vision, and do not consider the process as a whole.

E.2 The Computational Theory of Marr

The work of David Marr (Marr 1982, Marr & Nishihara 1978, Marr 1978) is perhaps one of the best (and most detailed) examples of the computational approach to the

recognition problem. It also probably remains the most influential contemporary model of three-dimensional shape recognition. Although many researchers may not accept the specifics of the Marr theory, there is a general tacit acceptance of his overarching claims.

The model attempts to separate early vision from recognition and visual cognition in general, and qualifies as an explicit theory of three-dimensional shape recognition by addressing most of the important problems facing this class of theory. Some feel that its shortcomings define many of the chief issues that researchers in shape recognition must face.

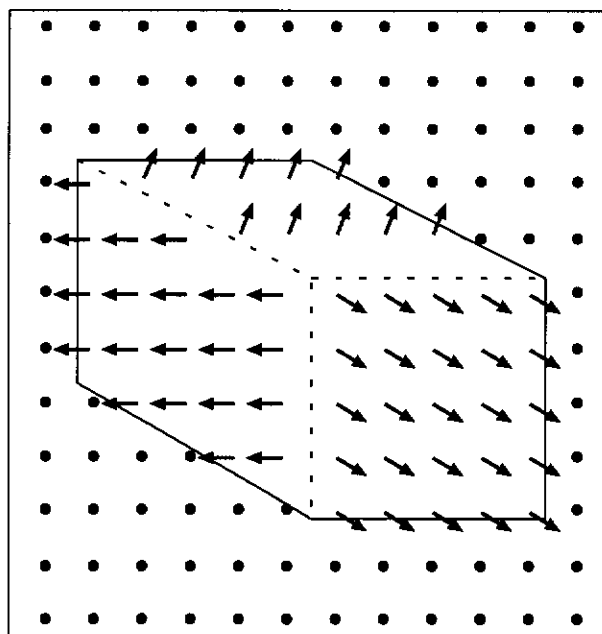
E.2.1 The $2\frac{1}{2}$ -D Sketch

Marr proposed that early vision culminates in the construction of a representation called the $2\frac{1}{2}$ -D sketch—essentially an array of cells, with each cell dedicated to a particular line of sight from the viewer's vantage point. Each cell in the array contains a number of symbols which indicate the depth of the local surface patch lying on that line of sight, the orientation of the patch relative to the viewer, and discontinuity information which indicates if an edge (a discontinuity in depth) or a ridge (a discontinuity in orientation) is present at that line of sight. An example of a $2\frac{1}{2}$ -D sketch is indicated in Figure E.1 (taken from Pinker (1986, page15)).

The arrows show the surface orientation of patches relative to the viewer, with the heavy dots representing foreshortened arrows directed at the viewer. Dotted lines represent a discontinuity in orientation (a ridge), and solid lines represent a discontinuity in depth (an edge). The actual depth of, say, the centroid of each patch is another piece of information not indicated on the figure.

Marr made the claim that the $2\frac{1}{2}$ -D sketch representation contains the richest information that the early visual processes can deliver. Another claim was that top-down processing took no part in the construction of the sketch, the associated biological claim being that human and animal low-level vision does not receive any inputs from 'higher cognitive functions'.

The representation only contains local depths and orientations, and does not contain

Figure E.1: $2\frac{1}{2}$ -D Sketch.

global information such as angles between lines, types of shapes, or part boundaries. This ‘early vision’ representation is thus available to later higher-order visual processes. See, for example, Marr (1982), Marr & Nishihara (1978), or Marr (1978) for full details on how the sketch is computed, or Pinker (1986) for an outline.

We note at this stage that the sketch is derived from a spatially-integrated full world image. This will present problems when we discuss the biological reality of the spatial fusion of saccades in Section 4.2.4.

As with other representations, the sketch is ill-suited for matching against stored shape representations. The same objections that applied to the template model are appropriate—only the visible surfaces of shapes are represented, and the third-dimension is only partially included. Further, the original sketch is viewpoint-specific with the result that a slight object rotation would result in a change in the representation.

To overcome this latter problem, Marr proposed that the object be described with respect to *some standard point* on the object, rather than viewer-centred. This would then enable the matching process to be made against stored representations with similar object-centred viewpoints. But once again this is putting the horse before the cart—the

object needs to be recognised before the standard point can be determined.

Real-world objects are not static but move about and change their shape. In an attempt to overcome this problem of variability of shape within objects, Marr proposed a higher-order representation comprising a hierarchy of models, each representing parts of different sizes and each with its own (local) coordinate system aligned with its axis of elongation. The proposal was for each object to be broken down into a series of connected cylinders—generalised cones.

An example of a human shape is illuminating. At a first approximation the human shape is composed of a cylinder for the *head*, another for the torso, two thin cylinders for the arms and a further two for the legs. At the next level of approximation, the arm cylinder may be subdivided into an upper arm cylinder and a lower arm cylinder. Next the lower arm cylinder is broken up into forearm and hand cylinders, and then at the next level down the hand cylinder is divided into palm, thumb, and four finger cylinders (see Pinker (1986, page 20)).

A criticism of this representation, other than it being an arbitrary scheme with no biological evidence, is that it is very cumbersome and unsuitable for irregular shapes. The real world is simply not composed of such simple geometric shapes, and even successive approximations based on simple shapes will not be able to produce the richness and variety of shapes found in reality.

This criticism is also appropriate to the initial $2\frac{1}{2}$ -D sketch. Other than man-made products, most objects simply do not have nice smooth patches even after some smoothing of the fine details.

A proposal for deriving 3-D descriptions from the $2\frac{1}{2}$ -D sketch has yet to be found.

Despite its initial appeal, the $2\frac{1}{2}$ -D sketch does not allow for a realistic or even efficient means of visual recognition.

E.2.2 Overarching Claims

Marr proposed a number of overarching claims that he suggested applied to the human visual system, and which are still accepted by most researchers.

world view & representation

The goal of the visual system is to provide to the rest of the brain with a *full representation* of the visual world as observed by the creature at that instant. The world view includes all of the panoramic scene (of 180° or more) that is perceived by the eyes.

hierarchical processing

There is a hierarchy of stages through which the visual signal processing must proceed—from the retina, through the LGN and then on to higher cortical stages. A typical scheme is shown in Figure 4.9. At each stage, the system builds up a representation of each object by extracting increasingly specific features, which are then combined into a fully elaborated representation specific to the object. It is only at this final stage that object recognition actually occurs—that is, visual learning occurs at later rather than earlier stages.

dependency relations

The flow of information in the system hierarchy is one-way. Higher levels in the hierarchy receive input from lower levels, but not, in general, vice versa. Some specific processing stages are considered to be part of the early (low level) visual system (for example, edge finding, stereo correspondence matching, shape from shading, motion), and these do not require any input from later stages such as segmentation or pattern recognition.

levels of description

Marr made the claim that any “algorithm” of perception could be understood at three different levels:

system level: an abstract statement of the problem,

algorithm level: the actual steps used in the process,

hardware level: the actual hardware used in the implementation.

Marr suggested that these levels were independent and that the solution to the perceptual problem could be discussed at any of these levels, without paying any attention

to the others. However, various experiments show that this cannot be the case, and that “our perceptual experience of the world is powerfully constrained by the actual neural machinery, the hardware that mediates perception.” (Ramachandran 1992, page 47).

Most would agree that these assumptions seem to concur with common sense, especially the view that we have available to us at all time a full percept of the “visual world”. At any given moment we seem to be able to see the full details of whatever visible features of the world are in front of our eyes. Further, all of these assumptions seem reasonable from an engineering point of view.

However, as we will show in the following sections, *all* of these underlying assumptions are not appropriate to primate visual processing.

E.3 Overall Problems with Existing Models as Related to Biological Processes

The first comment should be that there is nothing wrong with computer models that make no claim to biological reality, but propose some engineering solution to the issue of vision. After all, aeroplanes do not need to flap their wings for propulsion. We concern ourselves here with criticism of models that do make some claim as a viable model of biological vision, especially that of human and other vertebrates.

We first deal with some computational issues, then some objections based on epistemology, and then in the next section discuss some physiological and biological points which seem to question the viability of most current models.

the full task

A major criticism of many models of human and other vertebrate vision is that they do not address the whole problem. Many models which utilise structural descriptions, for example, take as their starting point a set of semantic primitives that are *assumed* to be supplied by some other, as yet unspecified process (but see Dillon 1996). To claim

that a primitive ‘feature’ such as ‘eye’ or ‘nose’ will simply fall out of a lower order process really begs the question, and places no importance on biological veracity.

On the other side, models which simply propose a ‘better edge-detector’ are of little value unless supported by psychological evidence, and are able to be positioned in an overall model of vision.

A related question concerns the required output of the system—what the recognition process must provide to the rest of cognitive facility. Is a label (of unknown origin), an adequate output from a vision system? A realistic model must be able to progress from biologically acceptable visual inputs through to some appropriate behaviour following recognition.

representations & shape primitives

There is much discussion, in the computer vision domain at least, about which representation is more appropriate. The discussion generally centres around computational considerations—which provides better search efficiency, which allows more compact descriptions, which can allow relational comparisons and so on—rather than on biological considerations.

Many of the proposed representations utilise shape primitives such as generalised cones, codons, or canonical volumetric shapes such as spheres, parallelepipeds, pyramids, and cones. But these representations are only appropriate to ‘regular’ shaped objects and are very limiting for more realistic general shapes. The concept of dividing objects up into smaller component parts of ‘standard’ shapes (such as cylinders) is inadequate.

A more appropriate approach might be to question the very notions of shape and representation. We deal with the notion of internal representations in a number of other sections, and especially in Section 4.2.5 in relation to visual representations. The issue discussed there is whether we really store a complete and separable internal representations for every conceivable object in the world?

The concept of ‘shape’ as a primary component of vision is suspect—most objects are much too dynamic and fractal to be covered by the simplistic notion of shape. What shape is a tree, a fence, a coastline, clouds? Yet humans and animals have no problem in recognising these ‘objects’, and discerning them in all their detail *if the need arises*.

The notion of 'object' is not quite so atomic as we sometimes believe.

frames of reference

Computational models need to assign a frame of reference to an object. Some postulate a viewer-centred frame, others object-centred, some even a global frame.

Some, which use the object-centred frame, put the cart before the horse in that the object must be (at least partially) recognised before the location of the object-centred reference point can be located. But in a biological visual system, the only viable frame of reference is with respect to the retina.

static images & re-cognition

Most visual systems (with the exception of the Active Vision paradigm) deal only with *static* images. They attempt to extract *as much* information as possible from stationary views of the world—at *each instant* of time. Most systems use only a single image—any additional images for a dynamic interpretation of the world would require a complete re-calculation of everything.

Not only is this computationally intractable, it is also not biologically realistic. The world is a dynamic, changing, real-time environment, with a continuous flow of imagery. Humans and animals do not *re-cognise* each object at each instant of time, over and over again. The process of vision is one of *confirmation* rather than repeated *recognition*.

computationally based

There have been a number of recent cutting criticisms of the computational approach to cognitive science (Port & van Gelder 1995, Dreyfus 1992, Winograd & Flores 1986, Costall & Still 1991, Irwin 1993b). While agreeing with this view that computationalism has no place in the study of the brain and cognition at any level other than as a simulation tool, we leave an in-depth discussion to other sections. However, a number of specific objections relating to vision may be put forward.

Current computer vision systems are essentially database solutions. The system builds up representations of world objects in the machine (either during manual construction of the program, or during learning), and then uses some search process to match inputs with representations at a subsequent recognition time. Each time the system is given the task of recognising an object in a scene, the input is put into

some representation, and a search performed through the entries in the database of stored representations seeking a match. This process often involves very complicated mathematics and numerical processing.

But there is one fundamental problem with this scheme from a computational point of view—the system will necessarily be slower at recognising when there are a larger number of objects in the database.

Even given the existence of a simple matching criteria (which is usually not the case), the very best search time that may be obtained for a database using a serial (von Neuman) computer is proportional to $\log(N)$ where N is the number of object representations in the database.

Many systems which use recognition by parts (for example, graph matching or sub-graph isomorphism) need to compare many parts with many equivalent components of the stored representations, resulting in a combinatorial explosion. It will take much longer to recognise an object in a larger set of representations than it will for a small set.

However, this is not the case for humans. Recall times for adults are not longer than those of children, and recall times for experts are not longer than those of novices. There is in fact the opposite finding—developmental trends in perceptual functioning show a progressive increase in specificity of discrimination, an optimisation of attention, and a progressive selectivity in information pickup with a correspondingly greater efficiency in ignoring irrelevant information (Gibson 1969, pages 450–472).

Another major problem with the current vision systems is that they are simply too slow. It takes many minutes of processing on high powered computers to ‘recognise’ isolated stationary objects in a ‘scene’ of only a small number of objects. A typical, current object recognition system might be able to recognise objects from a set of only, say, twenty pre-selected objects. Waiting for faster processors is hardly a satisfactory solution—more computing power simply begs the question.

Computer vision recognition time is usually nowhere near the approximately 150–300 msec recognition time (and approximately 100 synaptic joins) of human recognition (Feldman & Ballard 1982). Humans can recognise information about facial expressions as reliably from a 20 msec view as from much longer exposures (Simpson & Crandall 1972). Recognition of simple scenes and familiar faces may actually occur

the first glance (Eriksen & Eriksen 1971, Biederman, Glass & Stacy 1973, Biederman, Rabinowitz, Glass & Stacy 1974, Intraub 1980). See also Pollatsek et al. (1984) for recognition times for line drawings.

In all, then, the (serial von Neuman) computational approach simply does not have the right profile to be taken seriously as a model of human vision. The onus should be on the computationalists to show how these recognition time issues may be overcome, and how realistic recognition times can be achieved using their approach.

machine learning

Most current vision systems take no account of previous history, of learning, or of context. Some are now attempting to incorporate machine learning, but because of the nature of symbolic processing and learning by examples, this presents a number of problems.

Machine learning employs *induction* over examples as its principal mechanism of inference and generalisation. It thus has to contend with all of the open-ended problems associated with induction. How are the terms to be grouped for a more general rule? Should the *most* or *least* general generalisation be used? What *heuristic* generalisation technique should be used—minimum message length, entropy, Occam's razor, and so on? Each of these methods leads to different generalisations, and in principle there are an infinite number of generalisations that are possible.

Some minor use of *deductive* methods is made with explanation-based learning—but this requires a priori 'rules' which may then be put into a more operational form.

The problem is that of symbolisation. Machine learning requires the prior existence of symbols assigned to parts or features, and the process of machine learning is the forming of combinations of these symbols to describe an object.

Epistemologically this is again putting the cart before the horse. Objects have to be recognised (identified) to have parts isolated and labelled so that the part labels can be used to recognise the object. The origin of the labels is also not considered.

Further problems arise if new, as yet unseen objects are added to the database. There are difficulties in modifying the existing representations, and also in producing new symbolic terms.

segmentation

Most current models of vision use segmentation as the primary means of dividing the world into separate objects or parts. But segmentation is an inherently difficult task, especially when there are many objects in a scene partially occluding one another.

Imagine the difficulty of trying to dynamically segment a scene of an animal stalking through the undergrowth towards you. The various trees, bushes and grass which occlude the animal would provide confusing segmentation boundaries, as would the camouflage of the animal itself.

Segmentation typically tries to statically recognise an object in a scene by finding light intensity or depth discontinuities, so that the scene may be reduced to a series of lines and ultimately connected regions. The result (using natural scenes) is usually a confusing tangle of disconnected regions which is difficult to aggregate into separable real-world objects. Segmentation is invariably the bane of any artificial vision system when applied to anything other than simple, regular-shaped objects in limited numbers.

The problem for any segmentation system is that global information is required to make decisions at the local level concerning what goes with what. Some developers have attempted to supply such *background knowledge* as a built-in component, but without explaining where this knowledge originated.

In the hierarchical structure of conventional vision systems, segmentation is typically attempted prior to recognition. It is thus critical to correct recognition in this architecture—if the segmentation is poor, then the subsequent recognition is doomed. The whole process is simply too static for realistic recognition.

Segmentation also smacks of the humunculus problem—what is going to inspect the segmented image? Why should a line sketch of an object be more useful to a creature than the original object in the world?

object-centred vs. concept-centred

Most systems are *object-centred*, and consider that the object is the basic unit of recognition and identification. But the world is not composed of ready-made labelled objects waiting to be discovered. The task of the visual system (as well as the other senses) is to somehow separate the world into separable objects or events, so that appropriate behaviour may be performed upon their subsequent recognition.

Creatures with suitable inter-modal facilities are able to label certain objects or events based on the need for behavioural differentiation. Some animals will label objects with certain yaps and grunts (for example, the vervet monkey uses several different grunt signs to label the object/events 'snake-close-by', 'leopard-approaching' and 'eagle-overhead' (Cheney & Sayfarth 1990)). Humans use speech and sometimes sign languages to label.

But what constitutes an *object*? When, as infants, we start to recognise objects, do we have the same criteria to 'carve up the world' into objects that we do in later years?

Clearly the existence of objects is ill posed. Current vision systems, in pre-selecting and naming the 'objects' they want to recognise, are assuming the result they are trying to achieve.

The existence of objects is determined by a need to know. Birds are able to recognise and utilise trees, but other than their differential use of certain components of a tree, the bird would not be aware of the separate 'object' status that humans attribute to roots, leaves, boughs, bark and so on. The bird does not need to differentiate these objects.

All that is available to infants are the perceptual inputs as yet not separated into object correspondences. The task of the infants brain is to (self-)organise these inputs into suitable *concepts* that may then be associated with particular external objects and events.

what constitutes a "part"?

Many computer models break up objects into component parts in the belief that recognition by parts is an easier problem. Apart from the veracity of this belief (detecting these 'parts' may be no less difficult than recognising the whole), there is also an important epistemological issue here—exactly what constitutes a part?

For example, a division of a tree into parts such as 'trunk', 'root', 'leaf', 'bough' is not only culturally relative, but is arbitrary both in the linguistic terms used and in the "carving of nature at her joints". What constitutes a 'part' for an expert botanist when recognising a tree is very different from what may constitute a 'part' when the same tree is recognised by a non-expert at plant classification.

How is the botanist's representation different to that of the non-expert, and how do we move from one representation to the next with learning? Does the botanist have a different object representation for each of the different sub-species that he/she recognises? And how does recognition by parts apply to animals who do not (supposedly) make use of such arbitrary divisions in their visual recognition. Yet animals, such as birds, are quite able to recognise and utilise a tree.

Again it seems the method has little psychological backing and is introduced only for computational 'efficiency'.

In many artificial vision systems, the use of 'parts' presupposes recognition. Part labels are assumed to be provided by some lower-level process. Once given labels for parts in some symbolic form, the process of 'recognition' is reduced to a form of symbol matching—typically using computational devices such as decision-trees, graph matching, and so on.

But epistemologically, the concept 'part' also presupposes some form of prior recognition. If objects are defined in terms of their parts, how do the initial descriptions originate in children, who presumably start out without a representation of an object, without the knowledge of a particular break-up of objects into parts, and without the linguistic terms. One is left with a process of infinite regress, or recourse to some form of innate descriptions. The only 'parts' possible are the low-level vectors extracted from the various sensory filters to the world.

labels

The role of most computer vision systems is to produce a *label* to indicate recognition of an object. These labels are provided to the system a priori with no concern as to their origin. The view taken is that objects in the world are known a priori, each with an existing label.

But the production and assignment of labels to objects in the world is a dynamic process which occurs every day. Experts in various fields give labels to newly discovered or invented objects, new marketing and slang terms are invented, and new words are applied to different 'carvings' of existing objects.

While the world may have been initially divided into different concepts on the basis of simple behavioural needs—we learn to differentiate brightly coloured frogs

from dull ones because the brightly coloured ones are poisonous while the dull ones are edible—the use of labels extends our ability to divide the world into ever finer concepts. Whereas prior to labels each individual had to rediscover these linkages between different sensations and appropriate behaviour, the use of labels allows easier conceptualisation and differentiation through explicit training.

For example, two insects may appear the same to an untrained eye, yet a research entomologist, in associating new labels to specific sensory differences between the insects, allows us to separate the two species within the genus.

A model of cognition must include the provision for the creation and allocation of new labels, much less the problem of language.

control—a metaphor

An issue with most current theories of vision (and most of current computer-based AI) is reminiscent of the many-body problem within physics. Say we have n gas molecules in a box and we want to describe the behaviour of the molecules, (given a set of initial conditions and ignoring quantum effects). One way to do this is to attempt to predict the movement of each molecule given the initial position, velocity and acceleration, and all forces between the molecules and the boundary.

In principle this is possible, but in practice it is not. By the time you have calculated the changes in a small number of particles (by using *approximations* to their actual equations of motion) the other particles would have moved elsewhere and perhaps collided with other particles—the problem is intractable.

But the same system can be solved statistically, and results in the gas law

$$PV = nRT.$$

And so it is with computer vision. The use of symbol processing and the attempt to extract all information about all objects in a scene is, like the many-bodied problem, too much attention to detail, too much control.

Would a creature need to know everything about a static environment before moving on to the next static image—there may be no time for this luxury. To survive, a creature must dynamically and quickly associate perceptual inputs with appropriate behaviour.

Neural networks essentially work by collecting statistics, and self-organizing maps in particular are appropriate for this task.

The instant you make something atomic (e.g., symbolise it) you have to account for it from then on. The instant you *label* something you have to keep track of the label. Like the molecules within the gas in the many-body problem you have to account for them individually. The problem is even worse than the many-body problem as you have another problem of joining (aggregating) concepts or splitting (bifurcating) concepts with atomic labels.

Appendix F

Endnotes

¹A study of the history of views of cognition is important as it allows an appreciation as to why the current views prevail. A discussion of this history is given by Osherson & Lasnik (1990) and Gardner (1985). Here we simply add a very brief summary of the early ideas on cognition.

The early Greek philosophers set the stage for a debate that has raged ever since—the epistemology of human and animal cognition. Where do ideas or concepts originate—are they learned from the world around us, or are they innate? What are the processes of the brain that enable behaviour and thinking?

Plato believed that all knowledge was innate, and that the process of education was simply to bring this knowledge to conscious awareness. His opinion was that concepts came first, and that these *mental* ideas determined our sensory experiences. Plato prized logic, debate, and the ‘exact’ sciences such as mathematics, above perceptual considerations.

Aristotle, on the other hand, believed that concepts are formed from sensory experience. To him, there are no innate ideas, but rather an innate mental facility for organising all sensory impressions into categories and classes. Aristotle criticised Plato for confusing human imagination with the real world.

However, Aristotle did believe that humans possess an innate ability to reason, and thought that reasoning was man's most distinguishing characteristic. He is said to have founded the science of logic, and did much to categorise and establish the formation of taxonomies of objects in the world.

These two opposing views on the nature of cognition essentially continue to this day. The Platonic view is supported by the *rationalists* who hold that the brain has a set of inbuilt ideas that are the basis of what they see as a priori knowledge of various necessary truths. As ideas are innate, little emphasis is placed on learning. Their lineage passes from Plato, through Descartes, to Chomsky, and is the dominant model of cognitive processes in Cognitive Science today.

In opposition to this view, the *empiricists* denied that there are any innate ideas, nor any priori knowledge or necessary truths. For example, the English empiricist of the seventeenth century, Locke, held that the brain of an infant is like a blank sheet of paper, and that all of our ideas are imprinted on the brain by experience. He believed that the brain has certain inherent powers, such as remembering and imagining, but that the ideas of these powers are not innate.

The empiricists lineage follows from Aristotle, through Locke, to the *behaviourists* and to some extent, to connectionism.

Phrenology was in vogue from the middle of the eighteenth until the middle of the nineteenth centuries. This was the belief that a person's character was determined by the shape, and in particular, by the 'bumps' of the skull. Phrenology implied some kind of localisation of brain function, and innate facilities.

With the advent of physiological experimentation in the 1870s, it was soon realized that functions of the brain are localised, but that these "localised functions are not simply related to behavioural skills, or to mental attributes or abilities" (Gregory 1987, page 619).

Researchers at the turn of the century were interested in cognition: thinking, consciousness, language and culture, problem solving. Their preferred methodology was introspection—self-reflection by a trained observer about the nature and course of his

own thought patterns. This subjective method, however, had many critics, and introspectionism was eventually overthrown by the behaviourists.

²Cognitivism remains the dominant philosophical position, despite the immense difficulties it faces. Indeed, there are often strong objections to any alternative view (Winograd & Flores 1986, page 16):

“The rationalistic orientation not only underlies both pure and applied science but is also regarded, perhaps because of the prestige and success that modern science enjoys, as the very paradigm of what it means to think and be intelligent. In studies of thought, emphasis is placed on the form of the rules and on the nature of the processes by which they are logically applied. Areas of mathematics, such as symbolic logic and automata theory, are taken as the basis for formalising what goes on when a person perceives, thinks, and acts. For someone trained in science and technology it may seem self-evident that this is the right (or even the only) approach to serious thinking. Indeed, this is why many workers in artificial intelligence find critiques like that of Dreyfus (*What Computers Can't Do*, 1979) obviously wrong, since they challenge this deep-seated pre-understanding. In defence, they argue that the only conceivable alternative is some kind of mysticism, religion, or fuzzy thinking that is a throwback to earlier stages of civilisation.”

We reject this criticism, and propose an alternative that explains these higher cognitive functions not as the *basis* of cognition, but rather as a *product* of cognitive processes based on lower-order inputs and self-organisation.

³This thesis make a number of strong criticisms of existing paradigms. These criticisms apply to claims made in regard to animal and human cognition. These paradigms may be used in another sense however—as an engineering method to develop working tools

for, say, language interpretation, or perhaps recognition systems in computer vision, or even a planning systems for robots. The thesis makes no claims on these *engineering* solutions—if they work and provide useful behaviour in robots and other computer systems, then of course there can be no objection.

⁴Lashley (1951, page 113) stressed the importance of serial order;

“Temporal integration is not found exclusively in language; the co-ordination of leg movements in insects, the song of birds, the control of trotting and pacing in a gaited horse, the rat running the maze, the architect designing a house, and the carpenter sawing a board present a problem of sequences of action which cannot be explained in terms of successions of external stimuli.”

As quoted in Glencross (1995) “*Lashley regarded the serial order problem as central to the humans’ unique ability to learn new complex sequences of behaviour as in speaking, playing musical instruments, typing and the like.*”

⁵The current cognitivist (SPS) view on motor control—the movement and control of muscles—is concerned with the mechanical aspects of motion. This computational approach is divided into two broad areas; the consideration of the geometrical aspects of motion (kinematics), and the consideration of the force components (dynamics).

These systems are static, with little or no consideration given to learning. Key concepts are *planning, prediction* and *control*.

For example, consider the case of trajectory planning (see Hollerbach 1990a). Trajectory planning is considered to be the time evolution of kinematic variables. The idea is that the brain, through some sensory process, determines the endpoints of, say, a required hand movement. A computational process is then invoked to interpolate a path from the current hand position to the required final point, via the *joint interpolation*

equations:

$$\begin{aligned}\theta_1(t) &= (\theta_1(t_f) - \theta_1(t_0)) f(t) + \theta_1(t_0) \\ \theta_2(t) &= (\theta_2(t_f) - \theta_2(t_0)) f(t) + \theta_2(t_0)\end{aligned}\tag{F.1}$$

where t_0 and t_f are the initial and final times respectively, θ_1 and θ_2 the angles of the shoulder and elbow joints (in some Cartesian coordinate system), and $f(t)$ some parametric time function that satisfies the end conditions $f(t_0) = 0$, and $f(t_f) = 1$. A simple example of such a function giving a constant velocity from start to finish, is $f(t) = (t - t_0)/(t_f - t_0)$.

The *dynamics* of the limb movement needs to take into account the torques, positions, velocities and accelerations of the joints. Under the hypothesis that the motor control system will need to formulate a *plan* to pass to the muscles in order to carry out the required action, the *inverse dynamics* equations need to be solved. For the two-joint arm model, assuming that the centre of gravity of the upper and lower arm components is mid-way on the line joining their two ends, the dynamical equations that the brain is supposed to solve are:

$$\begin{aligned}\tau_1 &= \ddot{\theta}_1 \left(I_1 + I_2 + m_1 l_1 l_2 \cos \theta_2 + \frac{m_1 l_1^2 + m_2 l_2^2}{4} + m_2 l_1^2 \right) \\ &\quad + \ddot{\theta}_2 \left(I_2 + \frac{m_2 l_1 l_2}{2} \cos \theta_2 + \frac{m_2 l_2^2}{4} \right) \\ &\quad - \dot{\theta}_2^2 \frac{m_2 l_1 l_2}{2} \sin \theta_2 - \dot{\theta}_1 \dot{\theta}_2 m_2 l_1 l_2 \sin \theta_2 \\ &\quad + g \left(\frac{m_2 l_2}{2} \cos(\theta_1 + \theta_2) + l_1 \left(\frac{m_1}{2} + m_2 \right) \cos \theta_1 \right) \\ \tau_2 &= \ddot{\theta}_1 \left(I_2 + \frac{m_2 l_1 l_2}{2} \cos \theta_2 + \frac{m_2 l_2^2}{4} \right) + \ddot{\theta}_2 \left(I_2 + \frac{m_2 l_2^2}{4} \right) \\ &\quad + \dot{\theta}_2^2 \frac{m_2 l_1 l_2}{2} \sin \theta_2 + g \frac{m_2 l_2}{2} \cos(\theta_1 + \theta_2)\end{aligned}\tag{F.2}$$

where τ_i are the torques associated with each arm component, m_i their masses, l_i the arm link lengths, and I_i the rotary inertia about the centre of gravity for each link.

As before, the brain is expected to solve these equations in order to provide the muscles with appropriate instructions—and the brain must perform the calculations under

various conditions; for example with an additional mass at the end of the arm, (such as a tennis racquet), or under conditions in which the target position is changing.

Hollerbach (1990a) cites four difficulties for this approach:

- the dynamic model may be inadequate
- the initial conditions may differ from the original plan
- perturbations may deflect the movement from the intended path
- the inverse dynamics model may be too complex for the nervous system to compute quickly

To overcome these limitations, the SPS model of motor control looks to feedback and feedforward control. Other considerations are *motor programs* in which a chain of planned motions are strung together to form a co-ordinated action, and higher-level planning. We need not go into further details here (see Hollerbach 1990b, Bizzi & Mussa-Ivaldi 1990, Giliana 1990, Wright 1990, Bizzi & Mussa-Ivaldi 1989).

Suffice to say that the model is static, arbitrary, and is very dependent upon symbolic representations and computations. The motor control system is divided into a number of arbitrary modules thought to be responsible to various components of behaviour, and the theory lacks an overall coherence. Further, although the SPS motor control models purport to be dynamic, the dynamics is included in a very piecemeal fashion.

Other criticisms of the SPS model in general are discussed in Section 3.4.

⁶The current dominant psychological theory of (motor) learning is schema theory (Jordan & Rosenbaum 1989, page 753). Schemas are interacting functional units such as perceptual schemas (which undertake perceptual analysis), and motor schemas (which make up the control systems that are able to be coordinated, in order to effect a wide variety of movement) (Arbib 1995b, page 15). Schemas are seen as a means of bridging the gap between the functional descriptions of traditional AI and Cognitive Science, and the neural network view of the neurosciences.

Schemas are thus a complex set of distributed agents which are able to interact. Each has its own dedicated computational capability, and is able to process its own local information for dissemination to other agents.

Schemas are not required to map onto brain regions in a one-to-one manner (see page 739, Clark 1995), and the allocation of functions to schemas, and schemas to neural circuitry may be revised to better fit subsequent neurological data (for example, lesion studies). Each schema contains only a partial representation of the external world.

A simplified example of the schema model is given in Arbib (1995b, page 16) in relation to the visuomotor coordination in frogs and toads. Frogs and toads are known to snap at small moving objects, while avoiding large moving objects. In this simplified model of the frog brain, Arbib postulates four schemas; a perceptual schema to recognise small moving objects, which would activate a second *approach-and-snap* action schema; with another perceptual schema to recognise large moving objects, which, in turn, would activate an *avoidance* motor schema. This example is over-simplified, as Arbib goes on to explain, but it does give the flavour of the schema model. A full discussion of the model is given in Arbib (1995a). See also Arbib (1988, 1993). Arbib & Hill (1988) has a discussion of schema theory in relation to language acquisition.

Although there is much in common between the ABC model and schema theory in that both attempt to describe the distributed and 'cooperative' nature of cognition, and to build connections between perceptual inputs and motor controls in a way that does not require executive control, none-the-less a number of criticisms cause us to reject the schema model.

Schema theory is still based in the so-called higher-level descriptions of classical AI, with the concomitant problems associated with AI, such as problems of coordination and framing. The model is strongly rooted in computationalism and programming concepts (for example, the idea of instantiating multiple copies of a schema has parallels in object-oriented programming). A schema as both a store of knowledge and a description of a process for the application of that knowledge, is simply another name for a program, (or function, or object, or agent) in computer parlance (see Arbib 1993, page 278). As such, the model is open to many of the same criticisms as the cognitivist model.

The claim is that each schema will ultimately be able to be implemented in terms of neural networks, and that until the actual structure of the neural network is determined, it is appropriate to substitute a standard symbolic program with identical inputs and outputs. But this presupposes that the decomposition into functional schema units will also be found in the neural structures, and there is little evidence for this.

The schema model is not fully distributed, but is still somewhat atomic. As stated by Arbib (1993, page 277)

Analysis of animal behaviour or human cognition should yield a model of how that behaviour is achieved through the *cooperative computation* of concurrently active regions or schemas of the brain (the two analyses are not equivalent: regions are structural units; schemas are functional units), . . .

The evidence for cooperating regions is well founded, but there is no evidence for separate *functional* units in the brain. The ability to ‘carve up’ the brain into convenient functional units may not be possible.

While the model is adaptable (within limits) to changes, in say, the perceptual context, it is essentially static, with little attention given to learning. Schemas are considered to be pre-existing and are simply re-assembled in various combinations. The basic pattern of activity is reactive rather than adaptive, and the requirement for an essentially local knowledge representation is not justified.

⁷A possible justification for the use of these differential learning periods on each of the SOM surfaces is the likely *hardening* of the within-surface lateral inhibitory links brought about through their continued excitation over time. Hebbian learning on these lateral weights will cause the ‘Mexican-hat’ shape to become narrower over time.

Given that the initial weights for all maps and motor actions are random, and as there are more likely to be stronger self-organising correlations in the earlier SOM surfaces (the regularity primarily provided by the regularity of the external world), the lateral inhibitory weights between nodes on the earlier SOM surfaces (indicated by the grey

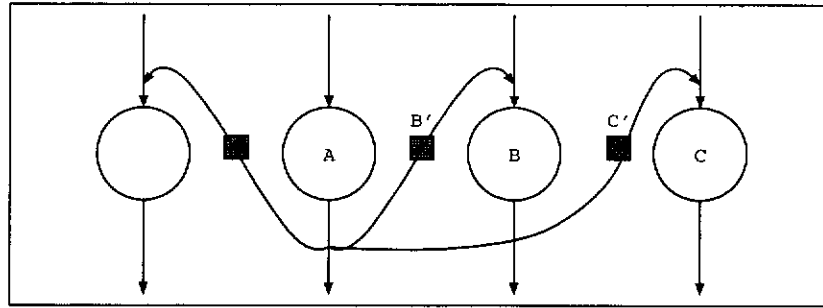


Figure F.1: Hebbian Learning of Lateral Inhibitory Links Within a SOM Layer.

boxes in Figure F.1) are more likely to ‘harden’ sooner (that is, become fully inhibitory on surrounding neurons).

For any particular inhibitory link, both the output from the winning neuron A and its nearest neighbours (B and C) will be firing, so weights B’ and C’ — which result from synaptic joins onto neurons B and C respectively—will be increased via Hebbian learning).

This will result in the ‘hardening’ of lateral inhibitions over time, but as there are more likely to be associative correlations in early maps, then one would expect hardening of these maps sooner. Hardening will produce an increasingly sharp and pointed ‘Mexican Hat’.

⁸A complimentary approach was used by Chappell & Taylor (1993) who considered the use of ‘leaky integrator’ neurons. This simple model of the rate of change of the neuron membrane potential m has

$$\tau_m \frac{dm(t)}{dt} = -m(t) + \sum_i w_i X_i(t) \quad (\text{F.3})$$

where the X_i are the inputs to the neuron, w_i the weights, and τ_m a time constant. The solution has an integrative component resulting from the inputs, but also a decay term.

Various other models of the behaviour of the neuron are found in the literature (see Softky & Koch 1995). Suffice to say that any model which retains some information about the *history* of the neuron will provide additional temporal information to the learning process.

⁹ *Alphabet*: system details of the alphabet learning task from page 52.

Program(s) Used	SOSCont/SOSTestCont		
Extensions	-		
Number of Words/Entities	28	SOM1 Initial Learning Rate	0.9
Dimension of Word Vector	35	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	12 × 12	SOM1 Decr Steps—Learn Rate	9
Number of Nodes on SOM2	12 × 12	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs—SOM1	100	SOM1 Initial Neighbour Radius	4
Number of Learn Epochs—SOM1	200	SOM1 Final Neighbour Radius	0
Number of Learn Epochs—Motor Wts	300	SOM1 Decr Steps—Radius	5
σ^2	0.3	SOM2 Initial Learning Rate	0.9
Gain	0.05	SOM2 Final Learning Rate	0.05
Cutoff	0.3	SOM2 Decr Steps—Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
		SOM2 Initial Neighbour Radius	4
		SOM2 Final Neighbour Radius	0
		SOM2 Decr Steps—Radius	5

¹⁰In this regard it is interesting to do a rough, corner-of-an-envelope examination of the number of neurons that may potentially be used in language. If there are 10^{11} neurons in total in the brain, and if we conservatively allow for say one thousandth of these to be used in language, and given that a well-educated person will know say 100,000 words (symbols) with say 10 connections/relationships per word, we are left with

$$\frac{10^{11} \times 10^{-3}}{10^5 \times 10^1} = 100$$

neurons per word—probably a conservative estimate. This calculation is probably too rough to be of any use, but the point is that the number of neurons per word is likely to be more than one.

¹¹The model as proposed could be altered slightly and yet still fit in with the overall requirements of temporal learning, self-organising structures, recurrent loops and associative connections that make up the major proposal of this thesis.

Various alternatives suggest themselves. For example, within the sensory modality structure, the recurrent loop might be taken following (o) rather than (k), which would

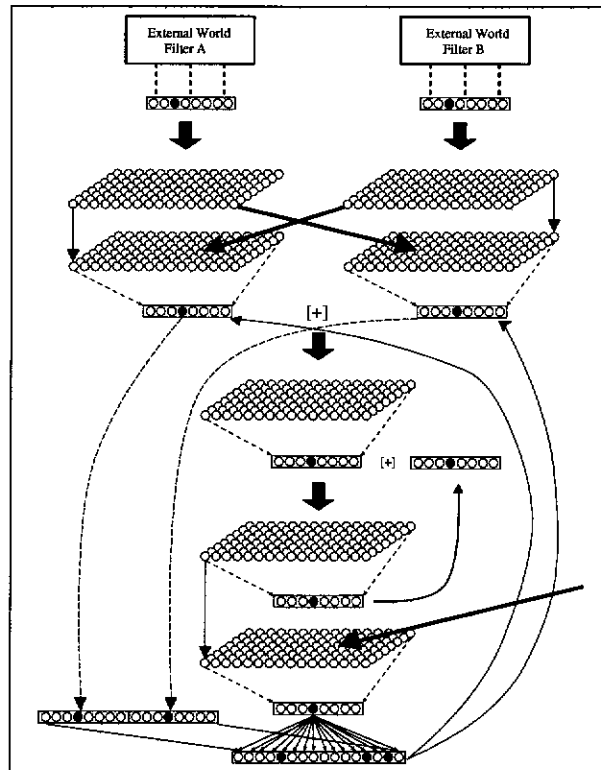


Figure F.2: Alternate Vision Structure.

include a component of the other modalities in the recurrent loop.

One such alternative is shown in Figure F.2. This, and other possible alternative arrangements will need to be examined at a later date in future research.

¹²By externalising the language of the problem, we may be forcing new paths to be taken through the concept attractors of the brain—that is, through those also linked by the overt speech mechanism. The process may be one of simply activating more paths from which to form associations and learn.

¹³There is also evidence that the ‘phoneme filter’ is self-organised during development. People raised in a Japanese speaking society are unable to subsequently differentiate the /l/ and /r/ sounds of an English speaker. These two English phonemes are in *con-*

trastive opposition in English (life vs. rife, belly vs. berry), but are in *complementary distribution* in Japanese, where the /l/ phoneme is sounded as both an [l] allophone at the beginning of a word (e.g., lan = flower) and an [r] allophone between vowels (e.g., nara = if).

In a similar way, English speakers have three distinct allophones for the phoneme /p/. These are [p^h] if 'p' starts a word, [p] which is unaspirated following an 's', and [p^o] which is unreleased if 'p' occurs at the end of a word. Although all three represent distinct sounds, they are recognised by English speakers as the one *sound meaning*. The same word spoken with all three variations would still be interpreted as the same word (for example, *map^h*, *map* and *map^o*). The same is true of the three allophone variations of the phoneme /t/ in English. Turkish speakers, on the other hand, do differentiate these three variations of 'p' and 't' into separate phonemes, and thus differentiable sound meanings.

Note that native Japanese speakers are able to later *learn* a distinction between the /l/ and /r/ phonemes of English through training and practice, but this *second language* learning needs to be much more explicit and may never result in the same skills as a native language user.

This phenomenon may also be explained through self-organisation and associative linkages. During development, the sounds of /l/ and /r/ for the Japanese child will self-organise to be close together on their 'phoneme filter', and will fuse into one attractor for sound meaning. Both sounds will be heard as the same meaning unit as both excite the same neural attractor on the phoneme map. However, for speakers of English, the two sounds will be separated as attractors as they lead to different subsequent behaviours. *

¹⁴As part of an ongoing discussion, the following is based on Winograd and Flores' *Understanding Computers and Cognition* (Winograd & Flores 1986). This discussion, in turn, was based on Dreyfus's *Being-in-the-world: A Commentary on Heidegger's Being and Time, Division I* (Dreyfus 1991). In the spirit of the discussion, this is one

interpretation of Winograd and Flores work, and in turn then, an interpretation of the discussion of Dreyfus.

Another extremely important and convincing critique of cognitivism, and the historical and philosophical assumptions leading to the cognitivist approach is Dreyfus (1992).

¹⁵The five central claims of Situated Cognition (SC) suggested by Clancey (1993) are:

1. **The representation storehouse view of memory confuses structures in the brain with physical forms that are created and used in speaking, drawing, writing etc.**
 - this reduces learning to syntactic modification of the modeller/teacher's pre-supplied ontology of standard notations,
 - representing, comprehending etc. are reified into acts of manipulating representations,
 - stored-schema models view *meaning* as a mapping between given information and stored conceptual primitives, facts and rules; *activity* as executing rules or scripts; *concepts* as stored descriptions;
2. **Schema models wrongly view learning as a secondary phenomenon, necessarily involving representations**
 - but learning occurs with every act of seeing or speaking,
 - categories are not stored things, but always adapted ways of talking, seeing, ... , and ways of coordinating behaviour,
 - comprehending is conceiving, not retrieving and matching,
 - that is, the process is NOT representation and search.
3. **Integration of perceiving and moving and higher order serial organisations is dialectic—coherent subprocesses arise together, not via linear causality or parallelism**

- perceiving, thinking, and moving always occur together, as coherent coordinations of activity,
- behaviour is not as a sequence of IF THEN decisions.

4. Practice cannot be reduced to theory

- there are not common ‘theories’ of the world in our heads (such as common laws, grammars, behaviour schemas, rules, and so on),
- concerning expert system knowledge transfer—if only we could *find the rules!*
- it is inappropriate to attempt to write a set of rules for a dynamic process such as driving a car,
- the rules are on the *outside* if we are forced to *describe* (verbalise) our actions,
- as we perceive patterns and articulate theories to explain them, we become increasingly alienated from the complexity of the activity itself.

5. Situated cognition relates to ideas in the philosophy of science concerning the nature of mechanisms and pattern descriptions

- rationalist philosophy of science has postulated that regularities in nature are caused by laws ($E = mc^2$), and this has had a profound influence on our thinking about human behaviour (modus ponens, stored facts and laws etc.),
- but even natural ‘laws’ are approximations with respect to third-party observers (e.g., Newton’s laws vs. relativity or quantum mechanics),

Clancey also states that SA is not:

- just about an agent “located in an environment” (situation), but a claim about the internal mechanisms that coordinates sensory and motor systems,
- a rejection of the value of planning and representations in everyday life, but seeking to explain how they are created and used in already coordinated action,
- claiming that *representing* does not occur internally, but rather trying to explain how perceiving and comprehending are co-organised,

- disputing descriptive value of schema models, but looking to explain how behaviour and pre-linguistic sensorimotor grounding is not captured by symbolic models.

According to Clancey, some of the implications of SA for any model of cognition are clear. We must:

- abandon the idea of stored first-person representational structures,
- abandon the CPU view of processing involving peripheral perception and motor systems, and replace it with a model of simultaneously coordinated perception-action,
- move deliberation out of the behaviour of the agent (as regards perception and action)
- view learning not as a process of storing new programs, but as part of an adaptive re-coordination that occurs with every behaviour, on top of which reflective representation (story telling, theorising) is based.

¹⁶ *StopWatch FSA*: system details of the Stopwatch FSA example from page 385.

Program(s) Used	FSA/FSA Test		
Extensions	-		
Number of States	3	Initial Learning Rate	0.9
Dimension of State Vector	7	Final Learning Rate	0.05
Number of Inputs	2	Decr Steps-Learn Rate	9
Dimension of Input Vector	3	Losing Learn Rate	0.001
Number of State Transitions	6	Initial Neighbour Radius	2
Number of Nodes on SOM	5 × 5	Final Neighbour Radius	0
Number of Learn Epochs-SOM	100	Decr Steps--Radius	3
Number of Learn Epochs-Motor Map	200	Max Initial Random Weight	0.5
σ^2	0.3	Gain	0.05
Cutoff	0.3		

¹⁷ *Binary Number FSA*: system details of the Binary Number FSA example from page 390.

Program(s) Used	FSA/FSA Test		
Extensions	-		
Number of States	11	Initial Learning Rate	0.9
Dimension of State Vector	10	Final Learning Rate	0.05
Number of Inputs	7	Decr Steps-Learn Rate	9
Dimension of Input Vector	6	Losing Learn Rate	0.001
Number of State Transitions	77	Initial Neighbour Radius	5
Number of Nodes on SOM	20 × 20	Final Neighbour Radius	0
Number of Learn Epochs-SOM	100	Decr Steps—Radius	6
Number of Learn Epochs-Motor Map	200	Max Initial Random Weight	0.5
σ^2	0.3	Gain	0.05
Cutoff	0.3		

¹⁸ *Decimal Number FSA*: system details of the Decimal Number FSA example from page 391.

Program(s) Used	FSAExcl/FSATest		
Extensions	exclusion		
Number of States	11	Initial Learning Rate	0.6
Dimension of State Vector	10	Final Learning Rate	0.05
Number of Inputs	15	Decr Steps-Learn Rate	9
Dimension of Input Vector	7	Losing Learn Rate	0.001
Number of State Transitions	165	Initial Neighbour Radius	14
Number of Nodes on SOM	30 × 30	Final Neighbour Radius	0
Number of Learn Epochs-SOM	100	Decr Steps--Radius	15
Number of Learn Epochs-Motor Map	200	Max Initial Random Weight	0.5
σ^2	0.3	Gain	0.05
Cutoff	0.3		

¹⁹ *Bidirectional Link FSA*: system details of the Bidirectional Link FSA example from page 392.

Program(s) Used	FSAExcl/FSATestRAG		
Extensions	exclusion/read-as-go		
Number of States	8	Initial Learning Rate	0.6
Dimension of State Vector	10	Final Learning Rate	0.05
Number of Inputs	6	Decr Steps-Learn Rate	9
Dimension of Input Vector	7	Losing Learn Rate	0.001
Number of State Transitions	48	Initial Neighbour Radius	9
Number of Nodes on SOM	20 × 20	Final Neighbour Radius	0
Number of Learn Epochs-SOM	100	Decr Steps--Radius	10
Number of Learn Epochs-Motor Map	200	Max Initial Random Weight	0.5
σ^2	0.3	Gain	0.05
Cutoff	0.3		

²⁰ *Reber FSA*: system details of the Reber FSA example from page 394.

Program(s) Used	FSAExcl/FSATestRAG		
Extensions	exclusion/read-as-go		
Number of States	8	Initial Learning Rate	0.6
Dimension of State Vector	10	Final Learning Rate	0.05
Number of Inputs	7	Decr Steps-Learn Rate	9
Dimension of Input Vector	7	Losing Learn Rate	0.001
Number of State Transitions	56	Initial Neighbour Radius	14
Number of Nodes on SOM	30 × 30	Final Neighbour Radius	0
Number of Learn Epochs-SOM	100	Decr Steps--Radius	15
Number of Learn Epochs-Motor Map	200	Max Initial Random Weight	0.5
σ^2	0.3	Gain	0.05
Cutoff	0.3		

²¹ An example of a minimal network discussed by Porat & Feldman (1991) is shown in Figure F.3. This network accepts strings over the alphabet {a, b} that contain an even number of a's—equivalent to the familiar *parity* problem.

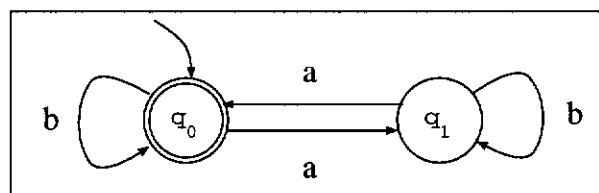


Figure F.3: Minimal Parity FSA.

- 22 *Temporal XOR*: system details of the Temporal XOR example (random sets of XOR strings) from page 397.

Program(s) Used	SOSCont/SOSTestCont		
Extensions	-		
Number of Words/Entities	2	SOM1 Initial Learning Rate	0.5
Dimension of Word Vector	2	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	3 × 3	SOM1 Decr Steps-Learn Rate	9
Number of Nodes on SOM2	3 × 3	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs-SOM1	100	SOM1 Initial Neighbour Radius	2
Number of Learn Epochs-SOM1	200	SOM1 Final Neighbour Radius	0
Number of Learn Epochs-Motor Wts	300	SOM1 Decr Steps—Radius	3
σ^2	0.3	SOM2 Initial Learning Rate	0.7
Gain	0.05	SOM2 Final Learning Rate	0.05
Cutoff	0.3	SOM2 Decr Steps-Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
		SOM2 Initial Neighbour Radius	2
		SOM2 Final Neighbour Radius	0
		SOM2 Decr Steps—Radius	3

- 23 *Temporal XOR*: system details of the Temporal XOR example (single set of example bits) from page 400.

Program(s) Used	SOSCont		
Extensions	-		
Number of Words/Entities	3	SOM1 Initial Learning Rate	0.9
Dimension of Word Vector	2	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	3 × 3	SOM1 Decr Steps-Learn Rate	9
Number of Nodes on SOM2	8 × 8	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs-SOM1	300	SOM1 Initial Neighbour Radius	4
Number of Learn Epochs-SOM1	600	SOM1 Final Neighbour Radius	0
Number of Learn Epochs-Motor Wts	900	SOM1 Decr Steps—Radius	5
σ^2	0.3	SOM2 Initial Learning Rate	0.9
Gain	0.05	SOM2 Final Learning Rate	0.05
Cutoff	0.3	SOM2 Decr Steps-Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
		SOM2 Initial Neighbour Radius	4
		SOM2 Final Neighbour Radius	0
		SOM2 Decr Steps—Radius	5

- 24 *Elman Consonant/Vowel Sequence*: system details of Elman’s consonant/vowel sequences from page 400.

Program(s) Used	SOSCont/SOSTestCont		
Extensions	-		
Number of Words/Entities	6	SOM1 Initial Learning Rate	0.5
Dimension of Word Vector	6	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	5 × 5	SOM1 Decr Steps-Learn Rate	9
Number of Nodes on SOM2	8 × 8	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs-SOM1	100	SOM1 Initial Neighbour Radius	3
Number of Learn Epochs-SOM1	200	SOM1 Final Neighbour Radius	0
Number of Learn Epochs-Motor Wts	300	SOM1 Decr Steps—Radius	4
σ^2	0.3	SOM2 Initial Learning Rate	0.7
Gain	0.05	SOM2 Final Learning Rate	0.05
Cutoff	0.3	SOM2 Decr Steps-Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
		SOM2 Initial Neighbour Radius	5
		SOM2 Final Neighbour Radius	0
		SOM2 Decr Steps—Radius	6
Data: 334 b, 333 d, 325 g			

- 25 *Elman Consonant/Vowel Sequence*: system details of Elman’s consonant/vowel sequences from page 403—modified vectors.

Program(s) Used	SOSSCont/SOSTestCont		
Extensions	-		
Number of Words/Entities	6	SOM1 Initial Learning Rate	0.5
Dimension of Word Vector	6	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	5 × 5	SOM1 Decr Steps-Learn Rate	9
Number of Nodes on SOM2	8 × 8	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs-SOM1	100	SOM1 Initial Neighbour Radius	3
Number of Learn Epochs-SOM1	200	SOM1 Final Neighbour Radius	0
Number of Learn Epochs-Motor Wts	300	SOM1 Decr Steps--Radius	4
σ^2	0.3	SOM2 Initial Learning Rate	0.7
Gain	0.05	SOM2 Final Learning Rate	0.05
Cutoff	0.3	SOM2 Decr Steps-Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
		SOM2 Initial Neighbour Radius	5
Data: 333 b, 333 d, 333 g		SOM2 Final Neighbour Radius	0
		SOM2 Decr Steps--Radius	6

26 *Bidirectional Link Sequence*: system details of the bidirectional link sequences from page 403.

Program(s) Used	SOSSupervised/SOSSupervisedTest		
Extensions	-		
Number of Concepts	6	SOM1 Initial Learning Rate	0.5
Dimension of Concept Vector	8	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	6 × 6	SOM1 Decr Steps-Learn Rate	9
Number of Nodes on SOM2	6 × 6	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs-SOM1	300	SOM1 Initial Neighbour Radius	3
Number of Learn Epochs-SOM1	600	SOM1 Final Neighbour Radius	0
Number of Learn Epochs-Motor Wts	900	SOM1 Decr Steps--Radius	4
σ^2	0.3	SOM2 Initial Learning Rate	0.6
High Gain	0.05	SOM2 Final Learning Rate	0.05
Low Gain	0.001	SOM2 Decr Steps-Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
		SOM2 Initial Neighbour Radius	3
Data: 100 strings		SOM2 Final Neighbour Radius	0
		SOM2 Decr Steps--Radius	4

27 *Bidirectional Link Sequence*: system details of the bidirectional link sequences from page 410 with extended learning.

Program(s) Used	SOSSupervised/SOSSupervisedTest		
Extensions	-		
Number of Concepts	6	SOM1 Initial Learning Rate	0.5
Dimension of Concept Vector	8	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	6 × 6	SOM1 Decr Steps-Learn Rate	9
Number of Nodes on SOM2	8 × 8	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs-SOM1	300	SOM1 Initial Neighbour Radius	3
Number of Learn Epochs-SOM1	600	SOM1 Final Neighbour Radius	0
Number of Learn Epochs-Motor Wts	900	SOM1 Decr Steps--Radius	4
σ^2	0.3	SOM2 Initial Learning Rate	0.6
High Gain	0.05	SOM2 Final Learning Rate	0.05
Low Gain	0.001	SOM2 Decr Steps-Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
		SOM2 Initial Neighbour Radius	3
Data: 500 strings		SOM2 Final Neighbour Radius	0
		SOM2 Decr Steps--Radius	4

28 *Bidirectional Link Generalisation*: system details of the bidirectional link generalisation from page 413.

Program(s) Used	SOSSupervised/SOSSupervisedTest		
Extensions	-		
Number of Concepts	6	SOM1 Initial Learning Rate	0.5
Dimension of Concept Vector	8	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	8 × 8	SOM1 Decr Steps-Learn Rate	9
Number of Nodes on SOM2	8 × 8	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs-SOM1	100	SOM1 Initial Neighbour Radius	2
Number of Learn Epochs-SOM1	200	SOM1 Final Neighbour Radius	0
Number of Learn Epochs-Motor Wts	300	SOM1 Decr Steps--Radius	3
σ^2	0.3	SOM2 Initial Learning Rate	0.6
High Gain	0.05	SOM2 Final Learning Rate	0.05
Low Gain	0.001	SOM2 Decr Steps-Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
		SOM2 Initial Neighbour Radius	2
Data: 500 strings		SOM2 Final Neighbour Radius	0
		SOM2 Decr Steps--Radius	3

29 *Reber Sequence Revisited: system details of the Reber sequences (supervised learning) from page 419.*

Program(s) Used	SOSSupervised/SOSSupervisedTest		
Extensions	-		
Number of Concepts	7	SOM1 Initial Learning Rate	0.5
Dimension of Concept Vector	8	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	8 × 8	SOM1 Decr Steps-Learn Rate	9
Number of Nodes on SOM2	8 × 8	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs-SOM1	100	SOM1 Initial Neighbour Radius	4
Number of Learn Epochs-SOM1	200	SOM1 Final Neighbour Radius	0
Number of Learn Epochs-Motor Wts	300	SOM1 Decr Steps—Radius	5
σ^2	0.3	SOM2 Initial Learning Rate	0.7
High Gain	0.05	SOM2 Final Learning Rate	0.05
Low Gain	0.002	SOM2 Decr Steps-Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
		SOM2 Initial Neighbour Radius	4
		SOM2 Final Neighbour Radius	0
		SOM2 Decr Steps—Radius	5

30 *Counting & Memory: system details of memory sequences from page 422.*

Program(s) Used	SOSContRP/SOSTestCont		
Extensions	Refractory period		
Number of Words/Entities	2	SOM1 Initial Learning Rate	0.5
Dimension of Word Vector	10	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	4 × 4	SOM1 Decr Steps-Learn Rate	9
Number of Nodes on SOM2	6 × 6	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs-SOM1	100	SOM1 Initial Neighbour Radius	2
Number of Learn Epochs-SOM1	200	SOM1 Final Neighbour Radius	0
Number of Learn Epochs-Motor Wts	300	SOM1 Decr Steps—Radius	3
σ^2	0.3	SOM2 Initial Learning Rate	0.5
Gain	0.05	SOM2 Final Learning Rate	0.05
Cutoff	0.3	SOM2 Decr Steps-Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
		SOM2 Initial Neighbour Radius	2
		SOM2 Final Neighbour Radius	0
		SOM2 Decr Steps—Radius	3

31 *Embedded Sequence—Common Components: system details of just the common components of the embedded sequences from page 428.*

Program(s) Used	SOSSupervisedMDataRP/SOSSupervisedTest		
Extensions	RP, multi data		
Number of Concepts	7	SOM1 Initial Learning Rate	0.5
Dimension of Concept Vector	10	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	8 × 8	SOM1 Decr Steps-Learn Rate	9
Number of Nodes on SOM2	10 × 10	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs-SOM1	100	SOM1 Initial Neighbour Radius	2
Number of Learn Epochs-SOM1	200	SOM1 Final Neighbour Radius	0
Number of Learn Epochs-Motor Wts	300	SOM1 Decr Steps—Radius	3
σ^2	0.3	SOM2 Initial Learning Rate	0.7
High Gain	0.05	SOM2 Final Learning Rate	0.05
Low Gain	0.001	SOM2 Decr Steps-Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
Data: 100 strings, 1 block		SOM2 Initial Neighbour Radius	1
RP = 1		SOM2 Final Neighbour Radius	0
		SOM2 Decr Steps—Radius	2

32 *Embedded Sequence—Context Bits: system details of the embedded sequences with context from page 431.*

Program(s) Used	SOSSupervizedContextMDataRP/SOSSupervizedTest		
Extensions	RP, multi data, context		
Number of Concepts	7	SOM1 Initial Learning Rate	0.5
Dimension of Concept Vector	8	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	10 × 10	SOM1 Decr Steps-Learn Rate	9
Number of Nodes on SOM2	10 × 10	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs-SOM1	100	SOM1 Initial Neighbour Radius	4
Number of Learn Epochs-SOM1	200	SOM1 Final Neighbour Radius	0
Number of Learn Epochs-Motor Wts	300	SOM1 Decr Steps—Radius	5
σ^2	0.3	SOM2 Initial Learning Rate	0.6
High Gain	0.05	SOM2 Final Learning Rate	0.05
Low Gain	0.001	SOM2 Decr Steps-Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
Data: 500 strings		SOM2 Initial Neighbour Radius	4
		SOM2 Final Neighbour Radius	0
		SOM2 Decr Steps—Radius	5

- ³³ *Simple English Sentences*: system details of the simple English sentences experiment from page 437.

Program(s) Used	SOSContX4/SOSTestCont		
Extensions	-		
Number of Words	38	SOM1 Initial Learning Rate	0.5
Dimension of Words Vector	10	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	14 × 14	SOM1 Decr Steps-Learn Rate	9
Number of Nodes on SOM2	22 × 22	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs-SOM1	100	SOM1 Initial Neighbour Radius	6
Number of Learn Epochs-SOM1	200	SOM1 Final Neighbour Radius	0
Number of Learn Epochs-Motor Wts	300	SOM1 Decr Steps—Radius	7
σ^2	0.3	SOM2 Initial Learning Rate	0.9
Gain	0.05	SOM2 Final Learning Rate	0.05
Cutoff	0.3	SOM2 Decr Steps-Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
		SOM2 Initial Neighbour Radius	8
		SOM2 Final Neighbour Radius	0
		SOM2 Decr Steps—Radius	9

- ³⁴ *Embedded English Sentences*: system details of the embedded English sentences experiment from page 445.

Program(s) Used	SOSSupervizedMDataRP/SOSSupervizedTest		
Extensions	RP=1		
Number of Concepts	22	SOM1 Initial Learning Rate	0.5
Dimension of Concept Vector	12	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	10 × 10	SOM1 Decr Steps-Learn Rate	9
Number of Nodes on SOM2	24 × 24	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs-SOM1	100	SOM1 Initial Neighbour Radius	4
Number of Learn Epochs-SOM1	200	SOM1 Final Neighbour Radius	0
Number of Learn Epochs-Motor Wts	300	SOM1 Decr Steps—Radius	5
σ^2	0.3	SOM2 Initial Learning Rate	0.7
High Gain	0.05	SOM2 Final Learning Rate	0.05
Low Gain	0.001	SOM2 Decr Steps-Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
Data: 0% 0% 100% (100)		SOM2 Initial Neighbour Radius	11
		SOM2 Final Neighbour Radius	0
		SOM2 Decr Steps—Radius	12

- ³⁵ *Embedded English Sentences*: system details of the embedded English sentences experiment from page 447.

Program(s) Used	SOSSupervizedMDataRP/SOSSupervizedTest		
Extensions	RP=1		
Number of Concepts	22	SOM1 Initial Learning Rate	0.5
Dimension of Concept Vector	12	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	7 × 7	SOM1 Decr Steps—Learn Rate	9
Number of Nodes on SOM2	20 × 20	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs—SOM1	100	SOM1 Initial Neighbour Radius	3
Number of Learn Epochs—SOM1	200	SOM1 Final Neighbour Radius	0
Number of Learn Epochs—Motor Wts	300	SOM1 Decr Steps—Radius	4
σ^2	0.3	SOM2 Initial Learning Rate	0.7
High Gain	0.05	SOM2 Final Learning Rate	0.05
Low Gain	0.001	SOM2 Decr Steps—Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
Data: 4 input periods		SOM2 Initial Neighbour Radius	9
(1) epoch 1 - 100% 0% 0% (50)		SOM2 Final Neighbour Radius	0
(2) epoch 50 - 75% 25% 0% (100)		SOM2 Decr Steps—Radius	10
(3) epoch 90 - 50% 25% 25% (150)			
(4) epoch 120 - 17% 50% 33% (300)			

36 *Embedded English Sentences*: system details of the embedded English sentences experiment using both singular and plural cases from page 451.

Program(s) Used	SOSSupervizedMDataRP/SOSSupervizedTest		
Extensions	RP=1		
Number of Concepts	22	SOM1 Initial Learning Rate	0.5
Dimension of Concept Vector	12	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	8 × 8	SOM1 Decr Steps—Learn Rate	9
Number of Nodes on SOM2	20 × 20	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs—SOM1	100	SOM1 Initial Neighbour Radius	3
Number of Learn Epochs—SOM1	200	SOM1 Final Neighbour Radius	0
Number of Learn Epochs—Motor Wts	300	SOM1 Decr Steps—Radius	4
σ^2	0.3	SOM2 Initial Learning Rate	0.7
High Gain	0.05	SOM2 Final Learning Rate	0.05
Low Gain	0.001	SOM2 Decr Steps—Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
Data: 4 input periods		SOM2 Initial Neighbour Radius	9
(1) epoch 1 - 100% 0% 0% (50)		SOM2 Final Neighbour Radius	0
(2) epoch 50 - 75% 25% 0% (100)		SOM2 Decr Steps—Radius	10
(3) epoch 90 - 50% 25% 25% (150)			
(4) epoch 120 - 17% 50% 33% (300)			

37 *Embedded English Sentences (4,4,2)*: system details of the embedded English sentences experiment using both singular and plural cases of 4 nouns, 4 transitive and 2 intransitive verbs from page 452.

Program(s) Used	SOSSupervizedMDataRP/SOSSupervizedTest		
Extensions	RP=0		
Number of Concepts	22	SOM1 Initial Learning Rate	0.5
Dimension of Concept Vector	12	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	10 × 10	SOM1 Decr Steps—Learn Rate	9
Number of Nodes on SOM2	20 × 20	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs—SOM1	100	SOM1 Initial Neighbour Radius	4
Number of Learn Epochs—SOM1	200	SOM1 Final Neighbour Radius	0
Number of Learn Epochs—Motor Wts	300	SOM1 Decr Steps—Radius	5
σ^2	0.3	SOM2 Initial Learning Rate	0.7
High Gain	0.05	SOM2 Final Learning Rate	0.05
Low Gain	0.001	SOM2 Decr Steps—Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
Data: 1 input period		SOM2 Initial Neighbour Radius	9
100% 0% 0% (200)		SOM2 Final Neighbour Radius	0
		SOM2 Decr Steps—Radius	10

38 *Embedded English Sentences*: system details of the embedded English sentences experiment using 4 nouns, 4 transitive and 2 intransitive verbs, up to 3 levels of embedding, from page 452.

Program(s) Used	SOSSupervisedMDataRP/SOSSupervisedTest		
Extensions	RP=1		
Number of Concepts	22	SOM1 Initial Learning Rate	0.5
Dimension of Concept Vector	12	SOM1 Final Learning Rate	0.05
Number of Nodes on SOM1	16 x 16	SOM1 Decr Steps—Learn Rate	9
Number of Nodes on SOM2	30 x 30	SOM1 Losing Learn Rate	0.001
Number of Learn Epochs—SOM1	100	SOM1 Initial Neighbour Radius	7
Number of Learn Epochs—SOM1	200	SOM1 Final Neighbour Radius	0
Number of Learn Epochs—Motor Wts	300	SOM1 Decr Steps—Radius	8
σ^2	0.3	SOM2 Initial Learning Rate	0.7
High Gain	0.05	SOM2 Final Learning Rate	0.05
Low Gain	0.001	SOM2 Decr Steps—Learn Rate	9
Max Initial Random Weight	0.5	SOM2 Losing Learn Rate	0.001
Data: 4 input periods		SOM2 Initial Neighbour Radius	14
(1) epoch 1 - 100% 0% 0% (50)		SOM2 Final Neighbour Radius	0
(2) epoch 50 - 75% 25% 0% (50)		SOM2 Decr Steps—Radius	15
(3) epoch 90 - 50% 25% 25% (100)			
(4) epoch 120 - 17% 50% 33% (200)			

³⁹Chomsky studied the mathematical properties of serially-processed string grammars, and instituted a hierarchy which divides these grammars into four categories (Chomsky 1959):

Unrestricted	Rules are unrestricted.
Context-sensitive	Contains only productions of the form $\alpha \rightarrow \beta$ where $ \alpha \leq \beta $, and where $ \alpha $ denotes the length of α . Rewriting the above in the form $\phi_1 A \phi_2 \rightarrow \phi_1 \psi \phi_2$ where $\alpha = \phi_1 A \phi_2$, $\beta = \phi_1 \psi \phi_2$, highlights the context-sensitivity— $A \rightarrow \psi$ in the context of ϕ_1 and ϕ_2 .
Context-free	Contains only productions of the form $\alpha \rightarrow \beta$ where $ \alpha \leq \beta $, and $\alpha \in V_n$ where V_n is a non-terminal symbol.
Regular	Contains only productions of the form $\alpha \rightarrow \beta$ where $ \alpha \leq \beta $, $\alpha \in V_n$, and β has the form aB or a where $A \in V_t$ and $B \in V_n$, and where V_t is a terminal symbol.

A grammar consists of a finite set of rules or *productions* which specify the syntax of the language, transforming one set of strings into another.

Introductory linguistics and AI textbooks incorrectly claim that at least context-dependent grammars are required for natural languages because of phenomena such as subject-verb agreement. But even finite-state languages can handle dependencies between symbols that are widely separated, as is shown in Section B.3. The dependency can be encoded in an extended state. The overwhelming majority of natural language structures are able to be parsed using finite-state or context-free techniques (Gazdar & Mellish 1989, page 133).

⁴⁰An interesting aside on so-called context-dependency in natural languages is discussed by Gazdar & Mellish (1989, page 132). They claim that a dialect of Swiss German spoken around Zürich exhibits a pattern that is context-dependent. An arbitrary number of noun phrases (NPs) may be followed by a corresponding number of verb phrases (VPs). The example they quote is

Claudia watched Helmut let Eva help Hans make Ulrike work.

which is ordered as

Claudia Helmut Eva Hans Ulrike watched let help make work

$NP_1 \quad NP_2 \quad NP_3 \quad NP_4 \quad NP_5 \quad V_1 \quad V_2 \quad V_3 \quad V_4 \quad V_5.$

This is of the form $NP^m V^n$ in which each NP is cross-linked to the appropriate V. But Swiss German also requires case markings of accusative and dative which must also match on the appropriate NPs and Vs. If the accusative and dative NPs and Vs are grouped, this will lead to a sentence structure of the form

$NP_a^m \quad NP_d^n \quad V_a^m \quad v_d^n$

or $a^m b^n c^m d^n$ ($n > 0$), which is context-dependent.

⁴¹ *Multiple FSAMaps—Context-Sensitive Grammar*: system details of the multiple FSAMaps—CSG example from page 469.

(a) Control FSAMap

Program(s) Used	FSAMulti/FSAMultiTest		
Extensions	-		
Number of States	5	Initial Learning Rate	0.5
Dimension of State Vector	12	Final Learning Rate	0.05
Number of Inputs	3	Decr Steps—Learn Rate	9
Input Vector A (no., no. bits)	3,6	Losing Learn Rate	0.001
Input Vector B (no., no. bits)	2,4	Initial Neighbour Radius	14
Input Vector C (no., no. bits)	2,4	Final Neighbour Radius	0
Number of Outputs	2	Decr Steps—Radius	15
Output Vector A (no. no. bits)	6,6	Max Initial Random Weight	0.5
Output Vector B (no. no. bits)	6,6	Gain	0.05
Number Nodes KM1	30 × 30	σ^2	0.3
Number Epochs KM1	200	Cutoff	0.3
Number Epochs Motor Weights	300		

(a) Counter FSAMap

Program(s) Used	FSAMulti/FSAMultiTest		
Extensions	Exclusion		
Number of States	7	Initial Learning Rate	0.6
Dimension of State Vector	6	Final Learning Rate	0.05
Number of Inputs	1	Decr Steps—Learn Rate	9
Input Vector A (no., no. bits)	6,6	Losing Learn Rate	0.001
Input Vector B (no., no. bits)	-	Initial Neighbour Radius	11
Input Vector C (no., no. bits)	-	Final Neighbour Radius	0
Number of Outputs	1	Decr Steps—Radius	12
Output Vector A (no. no. bits)	2,4	Max Initial Random Weight	0.5
Output Vector B (no. no. bits)	-	Gain	0.05
Number Nodes KM1	25 × 25	σ^2	0.3
Number Epochs KM1	200	Cutoff	0.3
Number Epochs Motor Weights	300		

⁴² *Multiple FSAMaps—c-like language*: system details of the multiple FSAMaps—c-like language example from page 474.

(a) Parser FSAMap

Program(s) Used	FSAMulti/FSAMultiTest		
Extensions	Exclusion		
Number of States	7	Initial Learning Rate	0.7
Dimension of State Vector	8	Final Learning Rate	0.05
Number of Inputs	2	Decr Steps—Learn Rate	9
Input Vector A (no., no. bits)	33,12	Losing Learn Rate	0.001
Input Vector B (no., no. bits)	3,4	Initial Neighbour Radius	14
Input Vector C (no., no. bits)	-	Final Neighbour Radius	0
Number of Outputs	1	Decr Steps—Radius	15
Output Vector A (no. no. bits)	7,8	Max Initial Random Weight	0.5
Output Vector B (no. no. bits)	-	Gain	0.05
Number Nodes KM1	30 × 30	σ^2	0.3
Number Epochs KM1	100	Cutoff	0.3
Number Epochs Motor Weights	500		

(a) Context FSAMap

Program(s) Used	FSAMulti/FSAMultiTest		
Extensions	Exclusion		
Number of States	8	Initial Learning Rate	0.7
Dimension of State Vector	8	Final Learning Rate	0.05
Number of Inputs	1	Decr Steps—Learn Rate	9
Input Vector A (no., no. bits)	7,8	Losing Learn Rate	0.001
Input Vector B (no., no. bits)	-	Initial Neighbour Radius	9
Input Vector C (no., no. bits)	-	Final Neighbour Radius	0
Number of Outputs	1	Decr Steps—Radius	10
Output Vector A (no. no. bits)	3,4	Max Initial Random Weight	0.5
Output Vector B (no. no. bits)	-	Gain	0.05
Number Nodes KM1	20 × 20	σ^2	0.3
Number Epochs KM1	100	Cutoff	0.3
Number Epochs Motor Weights	500		

⁴³ These computer system levels are often described as:

mechanism layer —the actual implementation medium, be it silicon chips, neurons, wooden rings of an abacus, registers and capacitors of analogue computers.

abstract layer —what the mechanism of the medium layer does. For example, a logic gate performs an XOR of its inputs. The processes performed at this layer are supposed to be independent of the mechanism layer—it doesn't matter what material is used to perform the XOR.

interpretation layer —what humans think the process has done in semantic terms.

For example, the computer has taken two numbers and multiplied them together.

⁴⁴Even the exchange of particles in current electro-magnetic and gravitational ‘explanations’ of this phenomena are approximations. The developers of these ideas proposed this ‘metaphor’ because the quantum mechanical mathematics (calculations) they used produced some results which came to suggested particle exchange as a mechanism, rather than action at a distance as for the traditional Newtonian rule.

But as particles can exhibit properties of waves under certain conditions, this formulation is not as neat and tidy as the metaphor suggests. In the fullness of time physicists will no doubt find a *better fit* to the data as anomalies in the current theory are found. The new ‘theory’ (string theory perhaps) will be just another better approximation to fitting the observations to some representational mathematical system of calculations. A new form of manipulating representations (a new form of mathematics) may even be ‘invented’ to better represent the theory. But there is nothing to say that a new theory is anything more than an approximation itself.

The view that the universe is made up of (simple) fundamental equations (laws) that are just waiting to be discovered by humans is just too simplistic. The universe is very complicated, non-linear, dynamic and probably fractal.

⁴⁵The passage from Polanyi (1962, page 49) is an excellent example of ‘rule learning’:

From my interrogations of physicists, engineers, and bicycle manufacturers, I have come to the conclusion that the principle by which the cyclist keeps his balance is not generally known. The rule observed by the cyclist is this. When he starts falling to the right he turns the handlebars to the right, so that the course of the bicycle is deflected along a curve towards the right. This results in a centrifugal force pushing the cyclist to the left and offsets the gravitational force dragging him down to the right. This manoeuvre presently throws the cyclist out of balance to the left, which he counteracts

by turning the handlebars to the left; and so he continues to keep himself in balance by winding along a series of appropriate curvatures. A simple analysis shows that for a given angle of unbalance the curvature of each winding is inversely proportional to the square of the speed at which the cyclist is proceeding.

But does this tell us exactly how to ride a bicycle? No. You obviously cannot adjust the curvature of your bicycle's path in proportion to the ratio of your unbalance over the square of your speed; and if you could you would fall off the machine, for there are a number of other factors to be taken account of in practice which are left out in the formulation of this rule.

As stated by Dreyfus (1992), this is an important insight in that the formalism cannot account for the performance.

Bibliography

- Agre, P. E. (1993). The symbolic worldview: Reply to Vera and Simon, *Cognitive Science* 17: 61–69.
- Ahmad, S. & Omohundro, S. (1991). Efficient visual search: A connectionist solution, *Proceedings of the 13th Annual Conference of the Cognitive Science Society, Chicago*, Hillsdale, NJ: Lawrence Erlbaum.
- Aitchison, J. (1987). Other keyholes: Language universals from a pidgin-creole viewpoint, in S. Modgil & C. Modgil (eds), *Noam Chomsky: Concensus and Controversy*, Falmer International Master-Minds Challenged: 3. The Falmer Press: NY, chapter 8.
- Allott, R. (1991). The motor theory of language, in W. von Raffler-Engel, J. Wind & A. Janker (eds), *Studies in Language Origins: Volume 2*, John Benjamins Publ. Co., pp. 123–157.
- Altman, G. T. M., Dienes, Z. & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21(4): 899–912.
- Amari, S.-I. (1990). Mathematical foundations of neurocomputing, *Proceedings of the IEEE: Special Issue on Neural Networks I* 78(9): 1443–1463.
- Anderson, J. A. (1995). Associative networks, in M. A. Arbib (ed.), *The Handbook of Brain Theory and Neural Networks*, The MIT Press: A Bradford Book: Cambridge, MA, pp. 102–107.
- Anderson, J. A. & Rosenfeld, E. (1988). *Neurocomputing: Foundations of Research*, MIT Press, Cambridge, MA.
- Angluin, D. (1988). Queries and concept learning, *Machine Learning* 2: 319–342.
- Angluin, D. & Smith, C. H. (1987). Formal inductive inference, in S. C. Shapiro (ed.), *Encyclopedia of Artificial Intelligence*, Vol. 1, New York: Wiley-Interscience Publication.
- Anstis, S. (1992). Visual adaption to a negative, brightness-reversed world: Some preliminary observations, in G. A. Carpenter & S. Grossberg (eds), *Neural Networks for Vision and Image Processing*, A Bradford Book: The MIT Press, pp. 1–14.

- Aparicio IV, M. & Levine, D. S. (1994). Why are neural networks relevant to higher cognitive function?, in D. S. Levine & M. Aparicio IV (eds), *Neural networks for knowledge representation and inference*, Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 1–26.
- Arbib, M. A. (1988). Schemas and neurons: Two levels of neural computing, in R. Eckmiller & C. v.d. Malsberg (eds), *Neural Computers*, Springer-Verlag: NATO ASI Series, Series F, Vol. 41, pp. 311–320.
- Arbib, M. A. (1993). Review of: Allen Newell, “Unified Theories of Cognition”, *Artificial Intelligence* 59: 265–283.
- Arbib, M. A. (1995a). Schema theory, in M. A. Arbib (ed.), *The Handbook of Brain Theory and Neural Networks*, The MIT Press: A Bradford Book: Cambridge, MA, pp. 830–834.
- Arbib, M. A. (ed.) (1995b). *The Handbook of Brain Theory and Neural Networks*, The MIT Press: A Bradford Book: Cambridge, MA.
- Arbib, M. A. & Hill, J. C. (1988). Language acquisition: Schemas replace universal grammar, in J. A. Hawkins (ed.), *Explaining Language Universals*, Basil Blackwell, chapter 3. Reprinted 1989.
- Åström, K. J. (1995). Adaptive control: General methodology, in M. A. Arbib (ed.), *The Handbook of Brain Theory and Neural Networks*, The MIT Press: A Bradford Book: Cambridge, MA, pp. 66–69.
- Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes, in K. W. Spence & J. T. Spence (eds), *Advances in the psychology of learning and motivation*, Vol. 2, New York: Academic.
- Atkinson, R. C. & Shiffrin, R. M. (1971). The control of short-term memory, *Scientific American* 224: 82–90.
- Baars, B. J. (ed.) (1988). *A Cognitive Theory of Consciousness*, Cambridge, UK: Cambridge University Press.
- Baber, R. N. (1983). The impending demise of the icon: A critique of the concept of iconic storage in visual information processing, *Behavioral and Brain Sciences* 6: 1–54.
- Bach, E., Brown, C. & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study, *Language and Cognitive Processes* 1: 249–262.
- Baddeley, A. D. & Hitch, G. J. (1974). Working memory, in G. H. Bower (ed.), *The psychology of learning and motivation: Advances in research and theory*, Vol. 8, New York: Academic Press, pp. 742–775.
- Bajcsy, R. (1986). Passive perception vs. active perception, *Proc. IEEE Workshop on Computer Vision*, pp. 333–356.

- Bajcsy, R. (1988). Active perception, *Proc. of the IEEE* **76**(8): 996–1005.
- Baker, L. & Cantwell, D. P. (1982). Language Acquisition, Cognitive Development, and Emotional Disorder in Childhood, in K. E. Nelson (ed.), *Children's Language: Volume 3*, Lawrence Erlbaum Associates: Hillsdale, NJ, chapter 6, pp. 286–321.
- Ballard, D. H. (1991). Animate Vision, *Artificial Intelligence* **48**: 57–86.
- Barbas, H. & Mesulam, M.-M. (1981). Organization of afferent input to subdivisions of area 8 in the rhesus monkey, *J. Comp. Neurol.* **200**: 407–432.
- Barnard, E. & Casasent, D. (1990). Shift invariance and the neocognitron, *Neural Networks* **3**: 403–410.
- Barsalou, L. W. (1992). *Cognitive Psychology: An Overview for Cognitive Scientists*, Hillsdale, NJ: Lawrence Erlbaum.
- Barwise, J. (1989). The situation in logic, *Technical report*, Stanford University.
- Barwise, J. & Etchemendy, J. (1987). *The liar: An essay on truth and circularity*, New York: Oxford University Press.
- Barwise, J. & Perry, J. (1983). *Situations and attitudes*, Cambridge, MA: MIT Press.
- Bates, E. & Elman, J. (1996). Learning rediscovered, *Science* **274**: 1849–1850.
- Bechtel, W. (1994). Consciousness: Perspectives from symbolic and connectionist AI, *Archive of Philosophy/Neuroscience/Psychology technical reports: Washington University*. <ftp://thalamus.wustl.edu/pub/pnp/papers/94-15/bechtel-consciousness.ps>.
- Benedict, H. (1979). Early lexical development: comprehension and production, *Journal of Child Language* **6**: 183–200.
- Benevento, L. A. & Fallon, J. H. (1975). The projection of occipital cortex to orbital cortex in the rhesus monkey (*Macaca mulatta*), *Exp. Neurol.* **46**: 402–408.
- Berk, L. E. (1994). Why Children Talk to Themselves, *Scientific American* pp. 60–65.
- Berk, L. E. & Garvin, R. A. (1984). Development of private speech among low-income Appalachian children, *Developmental Psychology* **20**(2): 271–286.
- Berk, L. E. & Landau, S. (1993). Private speech of learning disabled and normally achieving children in classroom academic and laboratory contexts, *Child Development* **64**(2): 556–571.
- Berkeley, B. (1709). *An essay towards a new theory of vision*.
- Berns, G. S., Dayan, P. & Sejnowski, T. J. (1993). A correlational model for the development of disparity selectivity in visual cortex that depends on prenatal and postnatal phases, *Proc. Natl. Acad. Sci. USA* **17**: 8277–8280.

- Bernstein, B. (1961a). Social class and linguistic development: A theory of social learning, in A. H. Halsey, J. Floud & E. Anderson (eds), *Education, Economy and Society*, New York: Free Press.
- Bernstein, B. (1961b). Social structure, language and learning, *Educational Research* 3: 163–176.
- Berry, D. C. (1994). Implicit learning: Twenty-five years on: A tutorial, *Attention and Performance*, Vol. 15, Amsterdam : North-Holland Pub. Co., chapter 30, pp. 755–782.
- Berry, D. C. & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalisable knowledge, *Quarterly Journal of Experimental Psychology* 36: 209–231.
- Berry, D. C. & Broadbent, D. E. (1987). The combination of explicit and explicit learning processes, *Psychological Research* 49: 7–15.
- Berry, D. C. & Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction, *British Journal of Psychology* 79: 251–272.
- Bickhard, M. H. & Richie, D. M. (1983). *On the Nature of Representation*, New York: Praeger.
- Biederman, I., Glass, A. L. & Stacy, E. W. (1973). Searching for objects in real world scenes, *J. Exp. Psychol.* 97: 22–27.
- Biederman, I., Rabinowitz, J. C., Glass, A. L. & Stacy, E. W. (1974). On the information extracted from a glance at a scene, *J. Exp. Psychol.* 103: 597–600.
- Bienenstock, E. & Geman, S. (1995). Compositionality in neural systems, in M. A. Arbib (ed.), *The Handbook of Brain Theory and Neural Networks*, The MIT Press: A Bradford Book: Cambridge, MA, pp. 223–226.
- Bienenstock, E. L., Cooper, L. N. & Munro, P. W. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex, *Journal of Neuroscience* 2: 32–48. Reproduced in *Neurocomputing: Foundations of Research*, James A. Anderson and Edward Rosenfeld, eds., Cambridge, MA: The MIT Press, pp. 439–455.
- Bignall, D. E. & Imbert, M. (1969). Polysensory and cortico-cortical projections to frontal lobe of squirrel and rhesus monkey, *Electroencephalogr. Clin. Neurophysiol.* 26: 206–215.
- Bisiach, E. (1988). Language without thought, *Thought Without Language*, Clarendon Press: Oxford, pp. 464–484.
- Bisiach, E., Berti, A. & Vallar, G. (1985). Analogical and logical disorders underlying unilateral neglect of space, in M. I. Posner & O. S. M. Martin (eds), *Attention and Performance, XI*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 239–246.

- Bizzi, E. & Mussa-Ivaldi, F. A. (1989). Geometrical and mechanical issues in movement planning and control, in M. I. Posner (ed.), *Foundations of Cognitive Science*, The MIT Press: A Bradford Book: Cambridge, MA, chapter 19, pp. 769–792.
- Bizzi, E. & Mussa-Ivaldi, F. A. (1990). Muscle properties and the control of arm movements, in D. N. Osherson, S. M. Kosslyn & J. M. Hollerbach (eds), *An Invitation to Cognitive Science*, Vol. 2—Visual Cognition and Action, The MIT Press: A Bradford Book, chapter 3, pp. 213–242.
- Bjorklund, D. F. (1989). *Children's Thinking: Developmental Function and Individual Differences*, Brooks/Cole Publishing Company: CA.
- Blake, A. & Yuille, A. (eds) (1992). *Active Vision*, Cambridge, MA: MIT Press.
- Blakemore, C. (1970). Development of the mammalian visual system, *Brit. Med. Bull.* **30**: 152–157.
- Blakemore, C. & Cooper, G. F. (1970). Development of the brain depends on the visual environment, *Nature* **228**: 477–478.
- Bloom, F. E. & Lazerson, A. (1988). *Brain, Mind, and Behavior (Second Edition)*, New York: W. H. Freeman and Company.
- Bloom, L. (1973). *One Word at a Time: The Use of Single Word Utterances before Syntax*, The Hague: Mouton.
- Blumer, D. & Benson, D. F. (1975). Personality changes with frontal and temporal lobe lesions, in D. F. Benson & D. Blumer (eds), *Psychiatric Aspects of Neurologic Disease*, Grune & Stratton, New York, pp. 151–170.
- Bornstein, M. H., Kessen, W. & Weiskopf, S. (1976). Color vision and hue categorization in young human infants, *Journal of Experimental Psychology: Human Perception and Performance* **2**: 115–129.
- Bourne, Lyle E., J., Dominowski, R. L. & Loftus, E. F. (1979). *Cognitive Processes*, Englewood Cliffs, NJ: Prentice-Hall.
- Bowerman, M. (1978). The acquisition of word meaning: An investigation of some current conflicts, in N. Waterson & C. Snow (eds), *Development in Communication: Social and pragmatic factors in language acquisition*, Chichester, UK: Wiley.
- Brandt, S. A. & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene, *Journal of Cognitive Neuroscience* **9**(1): 27–38.
- Brandt, T., Dichgans, J. & Koenig, E. (1973). Differential effects of central versus peripheral vision on egocentric and exocentric motion perception, *Exp. Brain Res.* **16**: 476–491.
- Braunwald, S. R. (1978). Context, word and meaning: Toward a communicational analysis of lexical acquisition, in A. Lock (ed.), *Action, Gesture and Symbol: The Emergence of Language*, Academic Press: London, chapter 18.

- Brickner, R. (1934). An interpretation of frontal lobe function based upon the study of a case of parial bilateral frontal lobectomy, *Res. Publ. Assoc. Res. Nerv. Ment. Dis.* **13**: 259–351.
- Bridgeman, B., der Heijden, A. H. C. V. & Velichkovsky, B. M. (1994). A theory of visual stability across saccadic eye movements, *Behavioral and Brain Sciences* **17**: 247–292.
- Brief, J.-C. (1983). *Beyond Piaget: A Philosophical Psychology*, Teachers College Press.
- Brinkman, C. & Porter, R. (1983). Supplementary motor area and premotor area of monkey cerebral cortex: Functional organization and activities of single neurons during performance of learned movement, in J. E. Desmedt (ed.), *Motor Control Mechanisms in Health and Disease*, Raven Press, New York, pp. 393–420.
- Briscoe, G. (1991). A computer simulation of the Wilson Modified Line-element Model of visual processing and hyperacuity. Unpublished Report, M.Cog.Sci., University of Melbourne.
- Briscoe, G. & Caelli, T. (1996). *A Compendium of Machine Learning*, Vol. 1: Symbolic Machine Learning, Norwood, NJ: Ablex Publishing Corporation.
- Brislin, R. & Keating, C. (1976). Cultural differences in the perception of a three-dimensional Ponzo illusion, *Journal of Cross-Cultural Psychology* **7**: 397–411.
- Broca, P. (1861). Perte de la parole. Ramollissement chronique et destruction partielle du lobe anterieur gauche du cerveau, *Bulletin de la Société d'Anthropologie* **2**: 235–238.
- Brodmann, K. (1909). *Vergleichende Lokalisationlehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*, Barth, Leipzig.
- Brown, C. M. (1990). Gaze control with interactions and delays, *Biological Cybernetics* **63**: 61–70.
- Brugge, J. F. & Reale, R. A. (1985). Auditory cortex, in A. Peters & E. G. Jones (eds), *Cerebral Cortex*, Vol. 4 Association and Auditory Cortices, Plenum Press: NY, chapter 6, pp. 229–271.
- Bruner, J. (1987). Prologue to the English edition, in R. W. Reiber & A. S. Carton (eds), *The Collected Works of L. S. Vygotsky: Volume 1—Problems of General Psychology*, Plenum Press: NY, pp. 1–16.
- Bruner, J. S. (1983). *Child's Talk: Learning to Use the Language*, Oxford University Press.
- Bruner, J. S. (1990). *Acts of meaning*, Harvard University Press.
- Bruyer, R. e. (1983). A case of prosopagnosia with some preserved covert rememberence of familiar faces, *Brain and Cognition* **2**: 257–284.
- Buckingham Jr., H. W. & Kertesz, A. (1974). A linguistic analysis of fluent aphasics, *Brain and Language* **1**: 29–42.

- Bunke, H. & Messmer, B. T. (1996). Recent advances in graph matching, *Technical report*, Institut für Informatik und Angewandte Mathematik, University of Bern, Bern, Switzerland.
- Burnham, D. K. (1987). The development of the categorical identification of speech, in B. E. McKenzie & R. H. Day (eds), *Perceptual Development in Early Infancy: Problems and Issues*, Lawrence Erlbaum, chapter 11, pp. 237–275.
- Butters, N., Pandya, D., Stein, D. & Rosen, J. (1972). A search for the spatial engram within the frontal lobes of monkeys, *Acta Neurobiol. Exp.* **32**: 305–329.
- Butterworth, G. & Grover, L. (1988). The origins of referential communication in human infancy, *Thought Without Language*, Clarendon Press: Oxford, pp. 5–24.
- Büttner, U. & Lang, W. (1979). The vestibulocortical pathway: Neurophysiological and anatomical studies in the monkey, in R. Granit & O. Pompeiano (eds), *Reflex Control of Posture and Movement*, Elsevier, Amsterdam, pp. 581–588.
- Caelli, T. & Reye, D. (1993). On the Classification of Image Regions by Colour, Texture and Shape, *Pattern Recognition* **26**(4): 461–470.
- Campbell, F. W. & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings, *J. Physiol.* **197**: 551–566.
- Campion, J., Latto, R. & Smith, Y. M. (1983). Is blindsight an effect of scattered light, spared cortex, and near-threshold vision?, *The Behavioral and Brain Sciences* **3**: 423–486.
- Caplan, D. (1987). Discrimination of normal and aphasic subjects on a test of syntactic comprehension, *Neuropsychologia* **25**: 173–184.
- Caplan, D. (1992). *Language: Structure, Processing and Disorders*, Cambridge, MA: MIT Press, Bradford Books.
- Caplan, D. (1995). The cognitive neuroscience of syntactic processing, in M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, A Bradford Book: The MIT Press: Cambridge, MA, pp. 871–879.
- Caramazza, A. (1988). Some aspects of language processing revealed through the analysis of acquired aphasia: The lexical system, *Ann. Rev. Neurosci.* **11**: 395–421.
- Carmody, D. P., Nodine, C. F. & Kundel, H. L. (1980). Global and segmented search for lung nodules of different edge gradients, *Invest. Radiol.* **15**: 224–233.
- Carpenter, G. A. & Grossberg, S. (1988). The Art of Adaptive Pattern Recognition by a Self-organizing Neural Network, *IEEE Computer* **21**(3): 77–88.
- Carpenter, G. A. & Grossberg, S. (1992a). A Self-Organizing Neural Network for Supervised Learning, Recognition, and Prediction, *IEEE Communications Magazine* pp. 38–49.

- Carpenter, G. A. & Grossberg, S. (eds) (1991). *Pattern Recognition by Self-Organizing Neural Networks*, MIT Press, Cambridge, MA: A Bradford Book.
- Carpenter, G. A. & Grossberg, S. (eds) (1992b). *Neural Networks for Vision and Image Processing*, A Bradford Book: The MIT Press.
- Carpenter, G. A., Grossberg, S. & Reynolds, J. H. (1991). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network, *Neural Networks* 4: 565–588.
- Carpenter, M. B. (1991). *Core Text of Neuroanatomy (Fourth Edition)*, Williams & Wilkins; Baltimore, ML.
- Cazden, C. B. (1966). Subcultural differences in child language: An interdisciplinary review, *Merrill-Palmer Quarterly* 12: 185–219.
- Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press.
- Chappell, G. J. & Taylor, J. G. (1993). The temporal Kohonen map, *Neural Networks* 6: 441–445.
- Chavis, D. A. & Pandya, D. N. (1976). Further observations on corticofrontal connections in the rhesus monkey, *Brain Res.* 117: 369–386.
- Chen, L. & Aihara, K. (1997). Chaos and asymptotic stability in discrete-time neural networks, *Non-Linear Electronic Journal*: <http://xyz.lanl.gov/chao-dyn/9701020>. To appear in *Physica D*. (in press).
- Cheney, D. L. & Sayfarth, R. M. (1990). *How Monkeys See the World: Inside the Mind of Another Species*, University of Chicago Press.
- Cheng, P. W. (1985). Restructuring versus automaticity: Alternative accounts of skill acquisition, *Psychological Review* 92(3): 414–423.
- Chino, Y. M., Kaas, J. H., Smith III, E. L., Langston, A. L. & Cheng, H. (1992). Rapid reorganization of cortical maps in adult cats following restricted deafferentation in retina, *Vision Research* 32(5): 789–796.
- Chomsky, N. (1957). *Syntactic Structures*, The Hague: Mouton.
- Chomsky, N. (1959). On certain formal properties of grammars, *Information and Control* 2(2): 137–167.
- Chomsky, N. (1964). A Review of B. F. Skinner's *Verbal Behavior*, in J. A. Fodor & J. J. Katz (eds), *The Structure of Language: Readings in the Philosophy of Language*, Prentice-Hall, Inc: NJ, chapter 21, pp. 547–578. Originally published: Chomsky, N., 1957, A review of Skinner's *Verbal Behavior*, in *Language*, 35:26–58.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*, MIT Press, Cambridge, MA.
- Chomsky, N. (1967). The general properties of language, in C. H. Millikan & F. L. Darley (eds), *Brain mechanisms underlying speech and language*, New York: Grune and Stratton.

- Chomsky, N. (1972). *Language and Mind (Enlarged Edition)*, New York: Harcourt Brace Jovanovich.
- Chomsky, N. (1975a). *The Logical Structure of Linguistic Theory*, New York: Plenum Press.
- Chomsky, N. (1975b). *Reflections on Language*, Pantheon Books, New York.
- Chomsky, N. (1980). *Rules and Representations*, Columbia University Press, New York.
- Chomsky, N. (1988). *Language and the Problems of Knowledge: The Managua Lectures*, Cambridge, MA: MIT Press.
- Christiansen, M. H. (1994). Language Learning in the Full or, Why the stimulus might not be so poor after all, *Archive of Philosophy/Neuroscience/Psychology technical reports: Washington University*.
- Christiansen, M. H. & Chater, N. (1994). Natural Language Recursion and Recurrent Neural Nets, *Archive of Philosophy/Neuroscience/Psychology technical reports: Washington University*.
- Churcher, J. & Scaife, M. (1982). How infants see the point, in G. E. Butterworth & P. Light (eds), *Social cognition: Studies of the development of understanding*, Harvester Press, Brighton, Sussex, pp. 110–136.
- Churchland, P. M. (1995). *The Engine of Reason, the Seat of the Soul*, Cambridge, MA: MIT Press, A Bradford Book.
- Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind-Brain*, MIT Press, Cambridge, MA.
- Churchland, P. S., Ramachandran, V. S. & Sejnowski, T. J. (1994). A critique of pure vision, in C. Koch & J. L. Davis (eds), *Large-Scale Neuronal Theories of the Brain*, A Bradford Book: The MIT Press, pp. 23–60.
- Cicerone, C. M. & Nerger, J. L. (1989). The relative number of long-wavelength-sensitive to middle-wavelength-sensitive cones in the human fovea centralis, *Vision Research* **29**: 115–128.
- Clancey, W. J. (1993). Situated action: A neuropsychological interpretation response to Vera and Simon, *Cognitive Science* **17**: 87–116.
- Clark, A. (1995). Philosophical issues in brain theory and connectionism, in M. A. Arbib (ed.), *The Handbook of Brain Theory and Neural Networks*, The MIT Press: A Bradford Book: Cambridge, MA, pp. 738–741.
- Clark, E. (1973a). Non-linguistic strategies and the acquisition of word meanings, *Cognition* **2**: 161–182.
- Clark, E. V. (1973b). What's in a word? On the child's acquisition of semantics in his first language, *Cognitive Development and the Acquisition of Language*, Academic Press, pp. 65–110.

- Cleeremans, A. (1993). *Mechanisms of Implicit Learning: Connectionist Models of Sequence Processing*, The MIT Press, A Bradford Book; Cambridge, MA.
- Cleeremans, A. (1994). The representation of structure in sequence prediction tasks, *Attention and Performance*, Vol. 15, Amsterdam : North-Holland Pub. Co., chapter 31, pp. 783–809.
- Cleeremans, A. & McClelland, J. L. (1991). Learning the structure of event sequences, *Journal of Experimental Psychology: General* **120**(3): 235–253.
- Cleeremans, A., Servan-Schreiber, D. & McClelland, J. L. (1989). Finite state automata and simple recurrent networks, *Neural Computation* **1**: 372–381.
- Clippingdale, S. & Wilson, R. (1996). Self-similar neural networks based on a Kohonen learning rule, *Neural Networks* **9**(5): 747–763.
- Colby, C. (1991). The neuroanatomy and neurophysiology of attention, *J. Child Neurol. Suppl.* **6**: S90–S118.
- Cole, M. (1990). Cultural psychology: A once and future discipline?, in J. J. Berman (ed.), *Nebraska Symposium on Motivation, 1989*, Vol. 37, Lincoln, NE: University of Nebraska Press, pp. 279–336.
- Cole, M., John-Steiner, V., Scribner, S. & Souberman, E. (eds) (1978). *L. S. Vygotsky: Mind in Society*, Harvard University Press: Cambridge, MA.
- Colombo, J. (1986). Recent studies in early auditory development, in G. J. Whitehurst (ed.), *Annals of child development*, Vol. 3, Greenwich, CN: JAI Press.
- Coltheart, M., Curtis, B. & Haller, M. (1993). Models of Reading Aloud: Dual-Route and Parallel-Distributed-Processing Approaches, *Psychological Review* **100**(4): 589–608.
- Costall, A. & Still, A. (1991). Introduction: Cognitivism as an approach to cognition, in A. Still & A. Costall (eds), *Against Cognitivism: Alternate Foundations for Cognitive Psychology*, Harvester Press, pp. 1–6.
- Cottrell, M., Fort, J.-C. & Pagès, G. (1994). Two or three things that we know about the Kohonen algorithm, *Technical Report 31*, Université Paris 1, Paris, France. (Reproduced on *neuroprose* electronic site. <ftp://cheops.cis.ohio-state.edu/pub/neuroprose/cottrell.things.ps>).
- Coulter, J. (1983). *Rethinking Cognitive Theory*, London: Macmillan.
- Coulter, J. (1989). *Mind in Action*, Cambridge: Polity Press.
- Cowan, P. A. (1978). *Piaget with Feeling: Cognitive, Social, and Emotional Dimensions*, Holt, Rinehart and Winston: NY.
- Cowey, A. & Stoerig, P. (1991). Reflections on blindsight, in A. D. Milner & M. D. Rugg (eds), *The Neuropsychology of Consciousness*, London, UK: Academic Press, pp. 11–37.

- Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research, *Journal of Verbal Learning and Verbal Behavior* **11**: 671–684.
- Crick, F. (1984). Function of the thalamic reticular complex: The searchlight hypothesis, *Proc. Natl. Acad. Sci. USA* **81**: 4586–4590. Reproduced in *Biology and Computation: A Physicist's Choice, Reprint Volume*, edited by H. Gutfreund and G. Toulouse. Advanced Series in Neuroscience – Vol. 3; World Scientific, 1994.
- Crick, F. & Asanuma, C. (1986). Certain aspects of the anatomy and physiology of the cerebral cortex, in J. L. McClelland, D. E. Rumelhart & the PDP Research Group (eds), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 2: Psychological and Biological Models, MIT Press, Cambridge, MA: A Bradford Book, chapter 20.
- Crick, F. & Koch, C. (1991). Towards a neurobiological theory of consciousness, *Neuroprose Electronic Site: ftp:// archive.cis.ohio-state.edu /pub/neuroprose/koch.awareness*.ps*. As published in *Seminars in The Neurosciences*, Volume 2, pp. 263–275, 1990.
- Crystal, D. (1992). *The Cambridge Encyclopedia of Language*, Cambridge University Press.
- Csabai, I. & Geszti, T. (1992). Criticality in the one-dimensional Kohonen neural map, *Physical Review A* **46**(10): 6181–6184.
- Curtiss, S. (1977). *GENIE: A Psycholinguistic Study of a Modern-Day "Wild Child"*, Academic Press: New York.
- Damasio, A. (1979). The frontal lobes, in K. M. Heilman & E. Valenstein (eds), *Clinical Neuropsychology*, Oxford University Press, London, pp. 360–412.
- Damasio, A. R. (1985). Disorders of complex visual processing: Agnosias, achromatopsia, Balint's syndrome, and related difficulties of orientation and construction, in M. M. Mesulam (ed.), *Principles of behavioral neurology*, Philadelphia, PA: F. A. Davis, pp. 259–288.
- Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergent zones, *Neural Computation* **1**: 123–132. Reproduced in *Biology and Computation: A Physicist's Choice, Reprint Volume*, edited by H. Gutfreund and G. Toulouse. Advanced Series in Neuroscience – Vol. 3; World Scientific, 1994.
- Darwin, C. ([1872] 1965). *The expression of the emotions in man and animals*, Chicago : University of Chicago Press.
- Das, S., Giles, C. L. & Sun, G.-Z. (1992). Learning context-free grammars: Capabilities and limitations of a recurrent neural network with an external stack memory, *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society, Indiana University, 1992*. (Neuroprose electronic site: ftp:// archive. cis. ohio-state.edu:/ pub/ neuroprose/ das.cfg-induction.ps).

- de Gelder, B. (1985). The cognitivist conjuring trick or how development vanished, in C. J. Bailey & R. Harris (eds), *Developmental Mechanisms of Language*, London: Pergamon Press, pp. 149–166.
- De Groot, A. D. (1978). *Thought and Choice in Chess*, Mouton, The Hague.
- De Valois, R. L., Abramov, I. & Jacobs, G. H. (1966). Analysis of response patterns of LGN cells, *J. Opt. Soc. Am.* **56**: 966–977.
- De Vito, J. L. & Smith, O. A. (1964). Subcortical projections of the prefrontal lobe of the monkey, *J. Comp. Neurol.* **123**: 413–424.
- Deacon, T. W., Rosenberg, P., Eckert, M. K. & Shank, C. E. (1982). Afferent connections of the primate inferior arcuate cortex, *Anat. Rec.* **8**: 933.
- DeCasper, A. J., Lecanuet, J.-P., Busnel, M.-C., Granier-Deferre, C. & Maugeais, R. (1994). Fetal reactions to recurrent maternal speech, *Infant Behavior and Development* **17**: 159–164.
- Dennett, D. (1989). Cognitive ethology: Hunting for bargains or a wild goose chase, in A. Montefiore & D. Noble (eds), *Goals, No-Goals and Own Goals: A Debate on Goal-Directed And Intentional Behaviour*, London: Unwin Hyman.
- Dennett, D. C. (1990). Cognitive wheels: The frame problem of AI, in M. A. Boden (ed.), *The Philosophy of Artificial Intelligence*, Cambridge University Press, pp. 147–170. Reprinted from: Daniel C. Dennett, 'Cognitive Wheels: The Frame Problem of AI', in C. Hookway (ed.), *Minds, Machines, and Evolution: Philosophical Studies* (1984), pp. 129–51, Cambridge University Press.
- Dennett, D. C. (1991). *Consciousness Explained*, Allen Lane: The Penguin Press.
- Dennett, D. & Kinsbourne, M. (1993). Time and the observer: The where and when of consciousness in the brain, *Center for Cognitive Studies, Tufts University* ([ftp.santafe.edu/pub/users/djp](ftp:santafe.edu/pub/users/djp)).
- Deuchar, M. (1987). Sign language as creoles and Chomsky's notion of universal grammar, in S. Modgil & C. Modgil (eds), *Noam Chomsky: Consensus and Controversy*, Falmer International Master-Minds Challenged: 3. The Falmer Press: NY, chapter 7.
- Devlin, K. (1991). *Logic and information*, Cambridge, UK: Cambridge University Press.
- Diamond, M. C., Scheibel, A. B. & Elson, L. M. (1985). *The Human Brain Coloring Book*, Barnes & Noble Books.
- Diaz, R. M. & Berk, L. E. (1992). *Private Speech: From Social Interaction to Self-Regulation*, Lawrence Erlbaum Associates.
- Dietrich, E. (1990). Programs in the search for intelligent machines: The mistaken foundation of AI, in D. Partridge & Y. Wilks (eds), *The Foundations of Artificial Intelligence: A Sourcebook*, Cambridge: Cambridge University Press, pp. 223–229.

- Dillon, C. (1996). *A Theory of Scene Understanding and Object Recognition*, PhD thesis, School of Computing, Curtin University of Technology, Perth, WA.
- Domasio, A. R. & Domasio, H. (1992). Brain and language, *Scientific American* pp. 89–95.
- Doya, K. (1994). Bifurcations of recurrent neural networks in gradient descent learning, *Neuroprose electronic site: ftp://archive.cis.ohio-state.edu:/pub/neuroprose/doya.bifurcation2.ps*.
- Doya, K. (1995). Recurrent networks: Supervised learning, in M. A. Arbib (ed.), *The Handbook of Brain Theory and Neural Networks*, The MIT Press: A Bradford Book: Cambridge, MA, pp. 796–800.
- Dreyfus, H. L. (1991). *Being-in-the-world: A Commentary on Heidegger's Being and Time, Division I*, Cambridge, MA: MIT Press.
- Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*, Cambridge, MA: MIT Press.
- Dreyfus, H. L. (1996). The current relevance of Merleau-Ponty's phenomenology of embodiment, *The Electronic Journal of Analytic Philosophy* 4. (<http://www.phil.indiana.edu/ejap/ejap.html>).
- Dreyfus, H. L. & Dreyfus, S. E. (1988). Making a mind versus modelling the brain: Artificial intelligence back at a branch point, *Artificial Intelligence* 117(1). Reproduced in "The Philosophy of Artificial Intelligence", ed. Margaret A. Boden, OUP, 1990.
- Duffy, F. H. & Burchfield, J. L. (1971). Somatosensory system: Organizational hierarchy from single units in monkey area 5, *Science* 172: 273–275.
- Duffy, F. H. & Burchfield, J. L. (1975). Eye movement-related inhibition of primate visual neurons, *Brain Res.* 89: 121–132.
- Edelman, G. M. (1978). Group selection and phasic reentrant signaling: A theory of higher brain function, in G. M. Edelman & V. B. Mountcastle (eds), *The Mindful Brain: Cortical Organization and the Group-Selective Theory of Higher Brain Function*, The MIT Press, Cambridge, MA, pp. 51–100.
- Edelman, G. M. (1985). Neural Darwinism: Population thinking and higher brain function, in M. Shafto (ed.), *How We Know: Proceedings of the Nobel Conference XX*, Harper & Row, San Francisco, pp. 1–30.
- Edelman, G. M. (1989a). *Neural Darwinism—The Theory of Neuronal Group Selection*, Oxford University Press.
- Edelman, G. M. (1989b). *The Remembered Present*, Basic Books, New York.
- Edelman, G. M. & Finkel, L. H. (1984). Neuronal group selection in the cerebral cortex, in G. M. Edelman, W. E. Gall & W. M. Cowan (eds), *Dynamical Aspects of Neocortical Function*, Wiley, chapter 22, pp. 653–693.

- Edelman, G. M. & Reeke Jr., G. N. (1982). Selective networks capable of representative transformations, limited generalizations and associative memory, *Proc. Natl. Acad. Sci. USA* **79**: 2091–2095.
- Edelman, G. M., Reeke Jr., G. N., Gall, W. E., Tononi, G., Williams, D. & Sporns, O. (1992). Synthetic neural modeling applied to a real-world artifact, *Proc. Natl. Acad. Sci. USA* **89**: 7267–7271.
- Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the (r-l) distinction by young infants, *Perception and Psychophysics* **18**: 341–347.
- Elman, J. L. (1990). Finding structure in time, *Cognitive Science* **14**: 179–211.
- Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure, *Machine Learning* **7**(2/3): 195–225. Reproduced in: “Connectionist Approaches to Language Learning”, David Touretzky (ed.), Kluwer Academic Press, 1991.
- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small, *Cognition* **48**: 71–99.
- Eriksen, C. W. & Eriksen, B. A. (1971). Visual perception processing rates and backward and forward masking, *J. Exp. Psychol.* **89**: 306–313.
- Erlich, S. F. & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading, *Journal of Verbal Learning and Verbal Behavior* **20**: 641–655.
- Ervin-Tripp, S. (1966). Language development, *Review of Child Development Research*, Ann Arbor: University of Michigan Press.
- Evarts, E. V. (1981). Motor cortex and voluntary movement, in J. M. Brookhart, V. B. Mountcastle, V. B. Brooks & S. R. Geiger (eds), *Handbook of Physiology*, Vol. 2, Sec. 1, (Part 2), Bethesda, MD: American Psychological Society, pp. 1083–1120.
- Fahle, M. & Poggio, T. (1981). Visual hyperacuity: Spatiotemporal interpretation in human vision, *Proc. R. Soc. Lond.* **B 213**: 451–477.
- Fahlman, S. E. (1991a). The recurrent cascade-correlation architecture, *Technical Report CMU-CS-91-100*, School of Computer Science, Carnegie Mellon University.
- Fahlman, S. E. (1991b). The recurrent cascade-correlation architecture, in R. P. Lippmann, J. E. Moody & D. S. Touretzky (eds), *Advances in Neural Information Processing Systems*, San Mateo, CA: Morgan Kaufmann, pp. 190–196.
- Fahlman, S. E. & Lebiere, C. (1990). The cascade-correlation learning architecture, *Technical Report CMU-CS-90-100*, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.
- Farah, M. J., Peronnet, F., Gonon, M. A. & Giard, M. H. (1988). Electrophysiological evidence for a shared representational medium for visual images and visual percepts, *Journal of Experimental Psychology: General* **117**: 248–257.

- Farber, I. B. & Churchland, P. S. (1995). Consciousness and the neurosciences: Philosophical and theoretical issues, in M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, A Bradford Book: The MIT Press: Cambridge, MA, pp. 1295–1306.
- Farthing, G. W. (ed.) (1973). *The Psychology of Consciousness*, Englewood Cliffs, NJ: Prentice Hall.
- Feldman, J. A. & Ballard, D. H. (1982). Connectionist models and their properties, *Cognitive Psychology* 6: 205–254.
- Feldman, J. A., Lakoff, G., Stolcke, A. & Weber, S. H. (1990). Miniature language acquisition: A touchstone for cognitive science, *Technical Report TR-90-009*, International Computer Science Institute, Berkeley, CA. Neuroprose electronic site: <ftp://archive.cis.ohio-state.edu/pub/neuroprose/feldman.tr90-9.ps>.
- Fischer, B. & Weber, H. (1993). Express saccades and visual attention, *Brain and Behavioral Sciences* 16(3): 553–567.
- Fischler, I. & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition, *Journal of Verbal Learning and Verbal Behavior* 18: 1–20.
- Fitzpatrick, K. A. & Imig, T. J. (1980). Auditory cortico-cortical connections in the owl monkey, *J. Comp. Neurol.* 192: 589–610.
- Flanagan, J. A. (1996). Self-organization in Kohonen's SOM, *Neural Networks* 9(7): 1185–1197.
- Florence, S. L. & Kaas, J. H. (1995). Somatotopy: Plasticity of sensory maps, in M. A. Arbib (ed.), *The Handbook of Brain Theory and Neural Networks*, The MIT Press: A Bradford Book: Cambridge, MA, pp. 888–891.
- Fodor, J. A. (1976). *The language of thought*, Sussex: Harvester Press (Harvard University Press paperback).
- Fodor, J. A. (1981). *Representations*, MIT Press, Cambridge, MA.
- Fodor, J. A. (1988). Précis of *The Modularity of Mind*, in Collins, Allan and Edward E. Smith (ed.), *Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence*, Morgan Kaufmann, pp. 73–77.
- Fodor, J. A. & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis, in S. Pinker & J. Mehler (eds), *Connections And Symbols: A COGNITION Special Issue*, The MIT Press: Cambridge, MA, pp. 3–71.
- Forgas, J. (1996). How emotion affects our thinking, *ABC Radio: The Health Report: Transcript*. <http://www.abc.net.au/rn/talks/8.30/helthrpt/healthrp.htm>.
- Fort, J.-C. & Pagès, G. (1996). About the Kohonen algorithm: Strong or weak self-organization?, *Neural Networks* 9(5): 773–785.

- Frazier, L. (1987a). Exploring the architecture of the language-processing system, in G. T. Altmann (ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, Cambridge, MA: MIT Press: A Bradford Book, pp. 409–433.
- Frazier, L. (1987b). Sentence processing: A tutorial review, in M. Coltheart (ed.), *Attention and Performance XII: The Psychology of Reading*, London: Erlbaum, pp. 559–586.
- Frazier, L. (1987c). Theories of sentence processing, in J. Garfield (ed.), *Modularity in Knowledge Representation and Natural-Language Processing*, Cambridge, MA: MIT Press, pp. 291–307.
- Fredrickson, J. M., Figge, U., Scheid, P. & Kornhuber, H. H. (1966). Vestibular nerve projection to the cerebral cortex of the rhesus monkey, *Exp. Brain Res.* **2**: 318–327.
- Fukushima, K. (1988). A neural network for visual pattern recognition, *IEEE Computer* pp. 65–75.
- Fukushima, K., Miyake, S. & Ito, T. (1983). Neocognitron: a neural network model for a mechanism of visual pattern recognition, *IEEE Transactions on Systems, Man, and Cybernetics SMC-13*: 826–834. Reproduced in *Neurocomputing: Foundations of Research*, James A. Anderson and Edward Rosenfeld, eds., Cambridge, MA: The MIT Press, pp. 526–534.
- Fuster, J. M. (1973). Unit activity in prefrontal cortex during delayed-response performance, *J. Neurophysiol.* **36**: 61–78.
- Fuster, J. M. (1980). *The Prefrontal Cortex*, Raven Press, New York.
- Fuster, J. M. (1981). Prefrontal cortex in motor control, in V. B. Brooks (ed.), *Handbook of Physiology*, Vol. II, American Physiological Society, Washington, D.C., pp. 1149–1178.
- Fuster, J. M. (1985). The prefrontal cortex and temporal integration, in A. Peters & E. G. Jones (eds), *Cerebral Cortex*, Vol. 4 Association and Auditory Cortices, Plenum Press: NY, chapter 4, pp. 151–177.
- Fuster, J. M. (1990). Inferotemporal units in selective visual attention and short-term memory, *J. Neurophysiol.* **64**: 681–697.
- Fuster, J. M., Bauer, R. H. & Jervey, J. P. (1982). Cellular discharge in the dorsolateral prefrontal cortex of the monkey in cognitive tasks, *Exp. Neurol.* **77**: 679–694.
- Gadamer, H.-G. (1976). *Philosophical Hermeneutics*, Berkeley, CA: University of California Press. (Trans. David E. Linge).
- Galaburda, A. M. & Pandya, D. N. (1983). The intrinsic architectonic and connectional organization of the superior temporal region of the rhesus monkey, *J. Comp. Neurol.* **221**: 169–184.
- Galletti, C. & Battaglini, P. P. (1989). Gaze-dependent visual neurons in area V3A of monkey prestriate cortex, *J. Neurosci.* **9**: 1112–1125.

- Gallistel, C. R. (1990). *The Organization of Learning*, Cambridge, MA: The MIT Press.
- Gardner, B. T. & Gardner, R. A. (1969). Teaching sign language to a chimpanzee, *Science* **162**: 664–672.
- Gardner, B. T. & Gardner, R. A. (1975). Early signs of language in child and chimpanzee, *Science* **187**: 752–753.
- Gardner, H. (1985). *The Mind's New Science: A History of the Cognitive Revolution*, Basic Books: NY.
- Gazdar, G. & Mellish, C. (1989). *Natural Language Processing in PROLOG: An Introduction to Computational Linguistics*, Addison-Wesley Publishing Company.
- Gazzaniga, M. (1985). *The Social Brain: Discovering the Networks of the Mind*, New York: Basic Books.
- Gazzaniga, M. S. (1992). *Nature's Mind: The Biological Roots of Thinking, Emotions, Sexuality, Language, and Intelligence*, Basic Books.
- Gazzaniga, M. S. (1995a). Consciousness and the cerebral hemispheres, in M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, A Bradford Book: The MIT Press: Cambridge, MA, pp. 1391–1400.
- Gazzaniga, M. S. (ed.) (1995b). *The Cognitive Neurosciences*, A Bradford Book: The MIT Press: Cambridge, MA.
- Gazzaniga, M. S. & LeDoux, J. E. (1978). *The Integrated Mind*, New York: Plenum.
- Geldard, F. A. & Sherrick, C. E. (1972). The cutaneous 'rabbit': A perceptual illusion, *Science* **178**: 178–179.
- Geldard, F. A. & Sherrick, C. E. (1983). The cutaneous saltatory area and its presumed neural base, *Perception and Psychophysics* **33**: 299–304.
- Geldard, F. A. & Sherrick, C. E. (1986). Space, time and touch, *Scientific American* **254**: 90–95.
- Gell-Mann, M. (1994). *The Quark and the Jaguar: Adventures in the simple and the complex*, Little, Brown and Company.
- Georgopoulos, A. P. (1990). Neurophysiology and reaching, in M. Jeannerod (ed.), *Attention and Performance*, Vol. 13, Hillsdale, NJ: Erlbaum, pp. 227–263.
- Georgopoulos, A. P. (1995). Motor cortex and cognitive processing, in M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, A Bradford Book: The MIT Press: Cambridge, MA, pp. 507–517.
- Georgopoulos, A. P., Kalaska, J. F. & Caminiti, R. (1985). Relations between two-dimensional arm movements and single cell discharge in motor cortex and area 5: Movement direction versus movement endpoint, *Exp. Brain Res. Suppl.* **10**: 176–183.

- Georgopoulos, A. P., Kalaska, J. F., Caminiti, R. & Massey, J. T. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex, *J. Neurosci.* **2**: 1527–1537.
- Geschwind, N. (1970). The organization of language and the brain, *Science* **170**: 940–944.
- Geschwind, N. & Galaburda, A. M. (eds) (1986). *Cerebral Lateralization*, Cambridge, MA: MIT Press.
- Ghiselin, B. (1952). *The Creative Process*, New York: Mentor.
- Gibson, E. J. (1969). *Principles of Perceptual Learning and Development*, Appleton-Century-Crofts, New York.
- Gibson, J. J. (1950). *The perception of the visual world*, Boston: Houghton Mifflin.
- Gibson, J. J. (1968). *The Senses Considered as Perceptual Systems*, George Allen & Unwin Ltd.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*, Boston, MA: Houghton Mifflin.
- Giefing, G.-J., Janßen, H. & Mallot, H. (1992). Saccadic Object Recognition with an Active Vision System, *Proceedings of the 11th IAPR International Conference on Pattern Recognition*, Vol. I, IEEE Computer Society Press, pp. 664–667.
- Gilbert, C. D. (1995). Dynamic properties of adult visual cortex, in M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, A Bradford Book: The MIT Press: Cambridge, MA, pp. 73–90.
- Gilbert, C. & Wiesel, T. N. (1991). Short and long-term changes in receptive field size and position following focal retinal lesions, *Society for Neuroscience Abstracts* p. 1090.
- Giles, C. L., Horne, B. G. & Lin, T. (1995). Learning a class of large finite state machines with a recurrent neural network, *Neural Networks* **8**(9): 1359–1365.
- Giles, C. L., Miller, C. B., Chen, D., Chen, H. H., Sun, G. Z. & Lee, Y. C. (1992). Learning and extracting finite state automata with second-order recurrent neural networks, *Neural Computation* **4**(3): 393–405.
- Giliana, H. L. (1990). Oculomotor control, in D. N. Osherson, S. M. Kosslyn & J. M. Hollerbach (eds), *An Invitation to Cognitive Science*, Vol. 2—Visual Cognition and Action, The MIT Press: A Bradford Book, chapter 4, pp. 243–283.
- Glass, A. L. & Holyoak, K. J. (1986). *Cognition (2nd ed.)*, New York: Random House.
- Gleason, J. B., Hay, D. & Cain, L. (1989). Social and affective determinants of language acquisition, in M. L. Rice & R. L. Schiefelbusch (eds), *The Teachability of Language*, Paul H. Brookes Publ. Co.: Baltimore, chapter 5, pp. 171–186.

- Glencross, D. J. (1995). Motor control and sensory-motor integration, in D. J. Glencross & J. P. Piek (eds), *Motor Control and Sensory Motor Integration: Issues and Directions*, Elsevier Science B. V., pp. 3–7.
- Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud, *Journal of Experimental Psychology: Human Perception and Performance* **5**: 674–691.
- Gold, E. (1967). Language identification in the limit, *Information and Control* **10**: 447–474.
- Goldenberg, G., Podreka, I., Steiner, M., Willmes, K., Suess, E. & Deecke, L. (1989). Regional cerebral blood flow patterns in visual imagery, *Neuropsychologia* **27**: 641–664.
- Goldstein, K. (1948). *Language and Language Disturbances*, Grune & Stratton, New York.
- Goltz, F. (1888). On the functions of the hemispheres, *Pflügers Archiv für die gesamte Physiologie* **42**: 419–467. (Trans. by G. von Bonin in *Some papers on the cerebral cortex*, Springfield, IL: Thomas, 1960.).
- Goodglass, H. & Quadfasel, F. A. (1954). Language laterality in left-handed aphasics, *Brain* **77**: 521–548.
- Goodglass, H., Wingfield, A., Hyde, M. R. & Theurkauf, J. C. (1986). Category specific dissociations in naming and recognition by aphasic patients, *Cortex* **22**: 87–102.
- Götz, K. G. (1988). Cortical templates for the self-organization of orientation-specific *d*- and *l*-hypercolumns in monkeys and cats, *Biological Cybernetics* **58**: 213–223.
- Graham, R. B. (1990). *Physiological Psychology*, Belmont, CA: Wadsworth Publishing Company.
- Greenfield, P. M. & Smith, J. (1976). *Constructions and the Beginnings of Language: The Development of Semantic Structure in One Word Speech*, New York and London: Academic Press.
- Greeno, J. G. (1989). Situations, mental models and generative knowledge, in D. Klahk & K. Kotovsky (eds), *Complex Information Processing: The Impact of Herbert A. Simon*, Hillsdale, N.J.: Lawrence Erlbaum.
- Greeno, J. G. & Moore, J. L. (1993). Situativity and symbols: Response to Vera and Simon, *Cognitive Science* **17**: 49–59.
- Greenwald, A. G. (1992). New Look 3: Unconscious cognition reclaimed, *The American Psychologist* **47**: 766–779.
- Greenwood, D. (1991). An Overview of Neural Networks, *Behavioural Science* **36**: 1–33.
- Gregory, R. L. (1987). *The Oxford Companion to the Mind*, Oxford University Press.

- Grey Walter, W. (1963). Presentation to the Ostler Society, Oxford University. Noted in Dennett & Kinsbourne (1993).
- Grimson, W. E. L. (1990). *Object Recognition By Computer: The Role of Geometric Constraints*, The MIT Press.
- Groner, R. & Menz, C. (1985). The effect of stimulus characteristics, task requirements and individual differences on scanning patterns, in R. Groner, G. W. McConkie & C. Menz (eds), *Eye Movement and Human Information Processing*, Elsevier Science Publishers B.V. (North-Holland), pp. 239–250.
- Groner, R., Walder, F. & Groner, M. (1984). Looking at faces: Local and global aspects of scanpaths, in A. G. Gale & F. Johnson (eds), *Theoretical and Applied Aspects of Eye Movement Research*, Elsevier Science Publishers B.V. (North-Holland), pp. 523–533.
- Grosf, D. H., Shapley, R. M. & Hawken, M. J. (1992). Macaque striate responses to anomalous contours?, *Invest. Ophthalmol. Vis. Sci.* **33**: 1257.
- Gross, C. G., Desimone, C. J. B. R., Fleming, J. & Gattass, R. (1981). Cortical visual areas of the temporal lobe: Three areas in the macaque, in C. N. Woolsey (ed.), *Cortical Sensory Organization*, Vol. 2, Humana Press, Clifton, N.J., pp. 187–216.
- Gross, L. (1974). Modes of communication and the acquisition of symbolic competence, *Seventy-third Yearbook of the National Society for the Study of Education*, Chicago: University of Chicago Press.
- Guckenheimer, J. & Holmes, P. (1983). *Nonlinear oscillation, dynamical systems, and bifurcations of vector fields*, Springer Verlag.
- Hacaen, H. (1964). Mental symptoms associated with tumors of the frontal lobe, in J. M. Warren & K. Akert (eds), *The Frontal Granular Cortex and Behavior*, McGraw-Hill, New York, pp. 335–352.
- Hacisalihzade, S. S., Stark, L. W. & Allen, J. S. (1992). Visual perception and sequences of eye movement fixations: A stochastic modeling approach, *IEEE Transactions on Systems, Man, and Cybernetics* **22**(3): 474–481.
- Haines, R. (1991). A breakdown in simultaneous information processing, in L. Stark & G. Obrecht (eds), *IVth International Symposium on Presbyopia*, New York: Plenum, pp. 171–176.
- Hallet, P. E. & Adams, B. D. (1980). The predictability of saccadic latency in a novel voluntary oculomotor task, *Vision Research* **20**: 329–339.
- Hamlyn, D. W. (1990). *In and Out of the Black Box: On the philosophy of cognition*, Oxford: Blackwell.
- Hampton, J. A. (1982). A demonstration of intransitivity in natural concepts, *Cognition* **12**: 151–164.
- Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions, *Memory & Cognition* **15**: 55–71.

- Hampton, J. A. (1988). Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14: 12–32.
- Harnad, S. (1982). Consciousness: An Afterthought, *Cognition and Brain Theory* 5: 29–47.
- Harnad, S. (1990). The symbol grounding problem, *Physica D* 42: 335–346. (Neuroprose electronic site: <ftp://archive.cis.ohio-state.edu/pub/neuroprose/> and <ftp://princeton.edu/pub/harnad/Harnad/harnad90.sgproblem>).
- Harnad, S. (1992). Connecting object to symbol in modeling cognition, in A. Clark & R. Lutz (eds), *Connectionism in Context*, Springer-Verlag. (Available from: Neuroprose electronic site: <ftp://archive.cis.ohio-state.edu/pub/neuroprose/harnad.symbobj.asc> and <ftp://princeton.edu/pub/harnad/Harnad/harnad92.symbol.object>).
- Harnad, S. (1993a). Grounding symbolic capacity in robotic capacity, in L. Steels & R. Brooks (eds), *The “artificial life” route to “artificial intelligence.” Building Situated Embodied Agents*, New Haven: Lawrence Erlbaum. (Available from: <ftp://princeton.edu/pub/harnad/Harnad/harnad93.robot>).
- Harnad, S. (1993b). Grounding symbols in the analog world with neural nets: A hybrid model, *Think (Special Issue on “Connectionism versus Symbolism”)* 2: 12–78. (Available from <ftp://princeton.edu/pub/harnad/Harnad/harnad93.symb.anal.net>).
- Harnad, S. (1993c). Symbol grounding is an empirical problem: Neural nets are just a candidate component, *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society*, NJ: Erlbaum. (Available from <ftp://princeton.edu/pub/harnad/Harnad/harnad93.cogsci>).
- Harnad, S. (1994). Does the mind piggyback on robotic and symbolic capacity?, in H. Morowitz & J. Singer (eds), *The Mind, the Brain, and Complex Adaptive Systems*, Santa Fe Institute/Addison Wesley. (Available from <ftp://princeton.edu/pub/harnad/Harnad/harnad94.mind.robot>).
- Harnad, S., Hanson, S. J. & Lubin, J. (1991). Categorical perception and the evolution of supervised learning in neural nets, *Neuroprose electronic site: ftp://archive.cis.ohio-state.edu/pub/neuroprose/*.
- Harnad, S., Hanson, S. J. & Lubin, J. (1994). Learned categorical perception in neural nets: Implications for symbol grounding, in V. Honavar & L. Uhr (eds), *Symbol Processing and Connectionist Network Models in Artificial Intelligence and Cognitive Modelling: Steps Toward Principled Integration*. (Available from <ftp://princeton.edu/pub/harnad/Harnad/harnad94.cpnet>).
- Hart, J., Berndt, R. S. & Caramazza, A. (1985). Category-specific naming deficit following cerebral infarction, *Nature* 316: 439–440.

- Harting, J. K., Updyke, B. V. & Lieshout, D. P. V. (1992). Corticotectal projections in the cat: Anterograde transport studies of twenty-five cortical areas, *J. Comp. Neurol.* **324**: 379–414.
- Hasher, L. & Zacks, R. T. (1979). Automatic and effortful processing in memory, *Journal of Experimental Psychology: General* **108**(3): 365–388.
- Hayes, N. A. & Broadbent, D. E. (1988). Two modes of learning for interactive tasks, *Cognition* **28**: 249–276.
- Hebb, D. O. (1949). *The organization of behavior*, New York: Wiley.
- Hecht-Nielsen, R. (1987). *Neurocomputing*, Addison-Wesley.
- Heidegger, M. (1978). *Being and time*, Oxford: Blackwell.
- Heil, J. (1981). Does cognitive psychology rest on a mistake?, *Mind* **90**: 321–342.
- Helmholtz, H. v. ([1910] 1962). *Physiological optics*, New York: Dover. (Ed. and trans. J. P. C. Southall).
- Helmholtz, H. v. (1925). *Physiological optics*, Vol. 3 (J.P.C. Southall, Trans.), Rochester, NY: Optical Society of America. (Originally published 1909).
- Henton, W. W. & Iverson, I. H. (1978). *Classical conditioning and operant conditioning: A response pattern analysis*, New York: Springer-Verlag.
- Heritage, J. (1984). *Garfinkel and Ethnomethodology*, Cambridge: Polity Press.
- Hertz, J. A., Krogh, A. S. & Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation*, Addison-Wesley, Redwood City, CA.
- Hilgard, E. R. & Hilgard, J. R. (eds) (1983). *Hypnosis in the Relief of Pain*, Kaufman, Los Altos, CA.
- Hinton, G. E. (1989). Connectionist learning procedures, *Artificial Intelligence* **40**: 185–234.
- Hirst, W. (1975). The psychology of attention, in J. E. LeDoux & W. Hirst (eds), *Mind and Brain: Dialogues in Cognitive Neuroscience*, Hillsdale, N. J.: Erlbaum.
- Hirst, W. (1995). Cognitive aspects of consciousness, in M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, A Bradford Book: The MIT Press: Cambridge, MA, pp. 1307–1319.
- Hochberg, J. (1984). Form perception: Experience and explanations, in P. C. Dodwell & T. M. Caelli (eds), *Figural synthesis*, Hillsdale, NJ: Erlbaum, pp. 1–30.
- Hockett, C. (1954). Two models of grammatical description, *Word* **10**: 210–231.
- Hollerbach, J. M. (1990a). Fundamentals of motor behavior, in D. N. Osherson, S. M. Kosslyn & J. M. Hollerbach (eds), *An Invitation to Cognitive Science*, Vol. 2—Visual Cognition and Action, The MIT Press: A Bradford Book, chapter 1, pp. 153–182.

- Hollerbach, J. M. (1990b). Planning of arm movements, in D. N. Osherson, S. M. Kosslyn & J. M. Hollerbach (eds), *An Invitation to Cognitive Science*, Vol. 2—Visual Cognition and Action, The MIT Press: A Bradford Book, chapter 2, pp. 183–211.
- Holyoak, K. J. & Glass, A. L. (1975). The role of contradictions and counterexamples in the rejection of false sentences, *Journal of Verbal Learning and Verbal Behavior* **14**: 215–239.
- Honderich, T. (1995). *The Oxford Companion to Philosophy*, Oxford University Press.
- Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks, *Neural Networks* **4**: 251–257.
- Howard, I. P. (1982). *Human Visual Orientation*, John Wiley & Sons.
- Hubel, D. H. (1995). *Eye, Brain, and Vision*, Scientific American Library.
- Hubel, D. H. & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *Journal of Physiology (London)* **160**: 106–154.
- Hubel, D. H. & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex, *J. Physiol (London)* **195**: 215–243.
- Hubel, D. H. & Wiesel, T. N. (1970). The period of susceptibility to the physiological effects of unilateral eye closure in kittens, *Journal of Physiology (London)* **206**: 419–436.
- Hubel, D. H. & Wiesel, T. N. (1977). Functional architecture of macaque monkey visual cortex, *Proceedings of the Royal Society of London* **B198**: 1–59.
- Hubel, D. H., Wiesel, T. N. & LeVay, S. (1977). Plasticity of ocular dominance columns in monkey striate cortex, *Philosophical Transactions of the Royal Society of London* **B278**: 377–409.
- Hubel, D. H., Wiesel, T. N. & Stryker, M. P. (1978). Anatomical demonstration of orientation columns in macaque monkey, *Journal of Comp. Neurology* **177**: 361–380.
- Hudson, W. (1960). Pictorial depth perception in sub-cultural groups in Africa, *J. Soc. Psychol.* **52**: 183–208.
- Humphrey, N. K. (1974). Species and individuals in the perceptual world of monkeys, *Perception* **3**: 105–114.
- Husserl, E. (1970). *Logical Investigations*, London: Routledge & Kegan Paul.
- Hutchins, E. (1990). The technology of team navigation, in J. Galegher, R. E. Kraut & C. Egidio (eds), *Intellectual Teamwork*, Hillsdale, NJ: Erlbaum.
- Hutchins, E. (1991). How a cockpit remembers its speeds, *Technical report*, University of California, San Diego.

- Huttenlocher, J. (1975). The origin of language comprehension, in R. L. Solso (ed.), *Theories in Cognitive Psychology*, New Jersey: Lawrence Erlbaum Associates.
- Hymes, D. (1971). Competence and performance in linguistic theory, in R. Huxley & E. Ingram (eds), *Language Acquisition: Models and Methods*, New York: Academic Press, pp. 3–28.
- Hyvärinen, J. (1982). *The Parietal Cortex of Monkey and Man*, Springer-Verlag, Berlin.
- Ingram, E. (1971). The requirements of model users, in R. Huxley & E. Ingram (eds), *Language Acquisition: Models and Methods*, New York: Academic Press, pp. 147–159.
- Ingvar, D. H. & Schwartz, M. S. (1974). Blood flow patterns induced in the dominant hemisphere by speech and reading, *Brain* **97**: 273–288.
- Inhelder, B. (1976). Observations on the operational and figurative aspects of thought in dysphasic children, in D. Morehead & A. Morehead (eds), *Normal and deficient child language*, Baltimore: University Park Press.
- Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory, *J. Exp. Psychol. Hum. Learn. Mem.* **6**: 1–12.
- Irwin, D. E. (1993a). Perceiving an integrated visual world, *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience—A silver jubilee*, The MIT Press: Cambridge, MA, pp. 121–142.
- Irwin, D. E. (1993b). Twenty-five years of learning and memory: Was the cognitive revolution a mistake?, *Attention and Performance XIV*, The MIT Press: Cambridge, MA, pp. 359–391.
- Irwin, D. E., Yantis, S. & Jonides, J. (1983). Evidence against visual integration across eye movements, *Percept. Psychophys.* **34**: 49–57.
- Jackendoff, R. (ed.) (1987). *Consciousness and the Computational Mind*, Cambridge, MA: MIT Press.
- Jacobson, S. & Trajanowski, J. Q. (1977). Prefrontal granular cortex in the rhesus monkey, *Brain Res.* **132**: 209–233.
- James, W. ([1884] 1994). The physical basis of emotion, *Psychological Review* **101**(2): 205–210. (Originally published: *Psychological Review*, 1, 516–529, 1884).
- James, W. ([1890] 1950). *The Principles of Psychology*, New York: Dover.
- Jensen, A. R. (1990). “g”: Outmoded theory or unconquered frontier?, *Creative Science and Technology* **11**: 16–29.
- Jiménez, L. & Cleeremans, A. (1994). Direct and indirect measures of implicit learning, *Proceedings of the 16th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Erlbaum, pp. 445–450.

- Johnson-Laird, P. N. (1983). *Mental Models*, Cambridge, MA: Harvard University Press.
- Johnson-Laird, P. N. (1988). *The Computer and the Mind*, Cambridge, MA: Harvard University Press.
- Johnson-Laird, P. N. (1989). Mental models, in M. I. Posner (ed.), *Foundations of Cognitive Science*, The MIT Press: A Bradford Book: Cambridge, MA, chapter 12, pp. 469–499.
- Johnson-Laird, P. N. (1995). Mental models, deductive reasoning, and the brain, in M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, A Bradford Book: The MIT Press: Cambridge, MA, pp. 999–1008.
- Jones, E. B. (1985). *The thalamus*, Plenum Press, New York.
- Jones, E. G. & Powell, T. P. S. (1969). Connections of the somatic sensory cortex in the rhesus monkey. I. Ipsilateral cortical connections, *Brain* **92**: 477–502.
- Jones, E. G. & Powell, T. P. S. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey, *Brain* **93**: 793–820.
- Jones, E. G. & Wise, S. P. (1977). Size, laminar and columnar distribution of efferent cells in the somatic-sensory cortex of monkeys, *J. Comp. Neurol.* **175**: 391–438.
- Jones-Gotman, M. & Milner, B. (1977). Design fluency: The invention of nonsense drawings after focal cortical lesions, *Neuropsychologia* **15**: 653–674.
- Jordan, M. I. (1986). Attractor dynamics and parallelism in a connectionist sequential machine, *Proceedings of the Eighth Annual Conference of the Cognitive Science Society, Amhurst, Massachusetts*, Lawrence Erlbaum, pp. 531–546.
- Jordan, M. I. (1989a). Indeterminate motor skill learning problems, in M. Jeannerod (ed.), *Attention and Performance, XIII*, Hillsdale, NJ: Erlbaum.
- Jordan, M. I. (1989b). Serial order: A parallel, distributed processing approach, in J. L. Elman & D. E. Rumelhart (eds), *Advances in Connectionist Theory: Speech*, Hillsdale, NJ: Erlbaum.
- Jordan, M. I. & Rosenbaum, D. A. (1989). Action, in M. I. Posner (ed.), *Foundations of Cognitive Science*, The MIT Press: A Bradford Book: Cambridge, MA, chapter 18, pp. 727–767.
- Jusczyk, P. (1997). *The Discovery of Spoken Language*, Cambridge, MA: The MIT Press.
- Kaas, J. H. (1983). What, if anything, is SI? Organization of first somatosensory area of cortex, *Physiol. Rev.* **63**: 206–231.
- Kaas, J. H. (1995). The reorganization of sensory and motor maps in adult mammals, in M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, A Bradford Book: The MIT Press: Cambridge, MA, pp. 51–71.

- Kaas, J. H., Merzenich, M. M. & Killackey, H. P. (1983). The reorganization of somatosensory cortex following peripheral nerve damage in adult and developing mammals, *Annual Review Neuroscience* **6**: 325–356.
- Kaas, J. H., Nelson, R. J., Sur, M., Lin, C. S. & Merzenich, M. M. (1979). Multiple representations of the body within the primary somatosensory cortex of primates, *Science* **204**: 521–523.
- Kahneman, D. (ed.) (1973). *Attention and Effort*, Englewood Cliff, N. J.: Prentice-Hall.
- Kahneman, D. & Treisman, A. M. (1984). Changing views of attention and automaticity, in R. Parasuraman & D. R. Davies (eds), *Varieties of Attention*, New York: Academic Press.
- Kaiser, P. K. & Boynton, R. M. (1996). *Human color vision: Second edition*, Optical Society of America.
- Kandel, E. R. & Schwartz, J. H. (1983). *Principles of Neural Science*, Elsevier/North-Holland.
- Kangas, J. (1989). Time-delayed self-organizing maps, *Proc. IJCNN'89 Int. Joint Conf. on Neural Networks II*: 331–336.
- Kangas, J. (1991). Time-Dependent Self-Organizing Maps for Speech Recognition, in T. Kohonen, K. Mäkisara, O. Simula & J. Kangas (eds), *Artificial Neural Networks - Volume 2 (Proceedings of the 1991 International Conference on Artificial Neural Networks (ICANN-91), Espoo, Finland, (June 1991)*, North-Holland, pp. 1591–1594.
- Kangas, J., Kohonen, T., Laaksonen, J., Simula, O. & Venta, O. (1989). Variants of self-organizing maps, *Proc. IJCNN'89 Int. Joint Conf. on Neural Networks II*: 517–522.
- Kapur, S. (1993). Uniform characterizations of various kinds of language learning, in K. P. Jantke, S. Kobayashi, E. Tomita & T. Yokomori (eds), *Algorithmic Learning Theory: Proceedings of the 4th International Workshop, ALT'93, Tokyo, Japan*, Springer-Verlag, pp. 197–208. Lecture Notes in Artificial Intelligence, No. 744.
- Kasamatsu, T. (1983). Neuronal plasticity maintained by the central norepinephrine system in the cat visual cortex, *Progress in Psychobiology and Physiological Psychology*, Vol. 10, Academic Press, pp. 1–112.
- Kelso, J. A. S., DeGuzman, G. C. & Holroyd, T. (1991). The self-organized phase attractive dynamics of coordination, pp. 41–62.
- Kemp, J. M. & Powell, T. P. S. (1970). The cortico-striate projection in the monkey, *Brain* **93**: 525–546.
- Khanna, T. (1990). *Foundations of Neural Networks*, Addison-Wesley.
- Kihlstrom, J. F. (1987). The cognitive unconscious, *Science* **237**: 1445–1452.
- Kimble, D. P. (1963). *Physiological Psychology*, Reading, MA.

- Klein, S. A. & Levi, D. M. (1985). Hyperacuity thresholds of 1.0 second: Theoretical predictions and empirical validation, *J. opt. Soc. Am.* **A2**: 1170–1190.
- Knudsen, E. I. (1984). Synthesis of a neural map of auditory space in the owl, in G. M. Edelman, W. E. Gall & W. M. Cowan (eds), *Dynamic Aspects of Neocortical Function*, John Wiley & Sons: A Neurosciences Institute Publication, chapter 11.
- Knudsen, E. I., DuLac, S. & Esterly, S. D. (1987). Computational maps in the brain, *Annual Review Neuroscience* **10**: 41–65.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps, *Biological Cybernetics* **43**: 59–69. Reproduced in *Readings in Machine Learning* (Shavlik & Dietterich 1990), pages 326–336.
- Kohonen, T. (1984). *Self-Organization and Associative Memory*, Springer-Verlag.
- Kohonen, T. (1989). Improved versions of learning vector quantization, *Proc. IJCNN'89 Int. Joint Conf. on Neural Networks I*: 545–550.
- Kohonen, T. (1990). The self-organizing map, *Proceedings of the IEEE: Special Issue on Neural Networks I* **78**(9): 1464–1479.
- Kohonen, T. (1993). Physiological Interpretation of the Self-Organizing Map Algorithm, *Neural Networks* **6**: 895–905.
- Kolb, B. & Whishaw, I. Q. (1990). *Fundamentals of Human Neuropsychology (Third Edition)*, New York: W. H. Freeman and Company.
- Kossut, M., Hand, P. J., Greenberg, J. & Hand, C. L. (1988). Single vibrissal cortical column in S-I cortex of rat and its afferations in neonatal and adult vibrissa-deafferented animals: A quantitative 2 DG study, *J. Neurophysiol.* **60**: 829–852.
- Kremer, S. C. (1995). On the computational power of Elman-style recurrent networks, *IEEE Transactions on Neural Networks* **6**(4).
- Krueger, L. E. (1975). Familiarity effects in visual information processing, *Psychological Bulletin* **82**: 949–974.
- Kubota, K., Iwamoto, T. & Suzuki, H. (1974). Visuokinetic activities of primate prefrontal neurons during delayed-response performance, *J. Neurophysiol.* **37**: 1197–1212.
- Kubota, K., Tonoike, M. & Mikami, A. (1980). Neuronal activity in the monkey dorsolateral prefrontal cortex during a discriminating task with delay, *Brain Res.* **183**: 29–42.
- Kuffler, S. W. (1952). Neurons in the retina: Organization, inhibition and excitation problems, *Cold Spring Harbor Symp. Quant. Biol.* **17**: 281–292.
- Kugler, P. N. & Turvey, M. T. (1987). *Information, natural law, and the self-assembly of rhythmic movement*, Hillsdale, NJ: Erlbaum.

- Kuhl, P. K. (1978). Predispositions for the perception of speech-sound categories: A species-specific phenomenon?, in F. D. Minifie & L. L. Lloyd (eds), *Communicative and cognitive abilities: Early behavioral assessment*, Baltimore: University Park Press.
- Kuhn, L. S. (1970). *The Structure of Scientific Revolutions*, The University of Chicago Press, Chicago.
- Künzle, H. & Akert, K. (1977). Efferent connections of cortical area 8 (frontal eye field) in *Macaca fascicularis*: A reinvestigation using the autoradiographic technique, *J. Comp. Neurol.* **173**: 147–164.
- Künzle, H., Akert, K. & Wurtz, R. H. (1976). Projections of area 8 (frontal eye field) to superior colliculus in the monkey: An autoradiographic study, *Brain Res.* **117**: 487–492.
- Kuypers, H. G. J. M., Szwarcbart, M. K., Mishkin, M. & Rosvold, H. E. (1965). Occipitotemporal corticocortical connections in the rhesus monkey, *Exp. Neurol.* **11**: 245–262.
- Laboratory of Comparative Human Cognition (1983). Culture and cognitive development, in W. Kessen (ed.), *Handbook of Child Psychology, Vol. 1: History, theory, and methods*, New York: Wiley.
- Lakatos, I. (1978). in J. Worrall (ed.), *Philosophical Papers*, Cambridge: Cambridge University Press.
- Lakoff, G. (1986). *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*, The University of Chicago Press.
- Lal, R. & Friedlander, M. J. (1989). Gating of retinal transmission by afferent eye position and movement signals, *Science* **243**: 93–96.
- Lang, K. J., Waibel, A. H. & Hinton, G. E. (1990). A time-delay neural network architecture for isolated word recognition, *Neural Networks* **3**: 33–43. Reproduced in *Readings in Machine Learning* (Shavlik & Dietterich 1990), pages 150–170.
- Larsen, B., Skinhoj, E. & Lassen, N. A. (1978). Variation in regional cortical blood flow in the right and left hemispheres during automatic speech, *Brain* **101**: 193–209.
- Lashley, K. S. (1951). The problem of serial order in behavior, in L. A. Jeffress (ed.), *Cerebral Mechanisms in Behavior*, Wiley, New York, pp. 112–146.
- Lassen, N. A. & Larsen, B. (1980). Cortical activity in the left and right hemispheres during language-related brain functions, *Phonetica* **37**: 27–37.
- Lave, J. (1988). *Cognition in Practice: Mind, mathematics and culture in everyday life*, Cambridge: Cambridge University Press.
- Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*, Cambridge, UK: Cambridge University Press.

- Leibnitz, G. W. ([1702] 1969). Reply to the thoughts on the system of preestablished harmony . . . , *Gottfried Wilhelm Leibnitz: Philosophical Papers and Letters*, Reidel, Dordrecht.
- Leibniz, G. (1951). *Selections* (ed. P. Wiener), New York: Scribner.
- Leichnetz, G. R. & Astruc, J. (1977). The course of some prefrontal corticofugals to the pallidum, substantia innominata , and amygdaloid complex in monkeys, *Exp. Neurol.* **54**: 104–109.
- Leichnetz, G. R., Spencer, R. F., Hardy, S. G. P. & Astruc, J. (1981). The prefrontal corticotectal projection in the monkey: An anterograde and retrograde horseradish peroxidase study, *Neuroscience* **6**: 1023–1041.
- Leinonen, L., Hyvärinen, J. & Sovijarvi, A. R. A. (1980). Functional properties of neurons in the temporo-parietal association cortex of the awake monkey, *Exp. Brain Res.* **39**: 203–215.
- Lenneberg, E. H. (1967). *Biological foundations of language*, New York: Wiley.
- LeVay, S., Stryker, M. P. & Shatz, C. J. (1978). Ocular dominance columns and their development in layer IV of the cat's visual cortex, *Journal of Comparative Neurology* **179**: 223–244.
- LeVay, S., Wiesel, T. N. & Hubel, D. H. (1980). The development of ocular dominance columns in normal and visually deprived monkeys, *Journal of Comparative Neurology* **191**: 1–51.
- Levin, B. E. & Dunn-Meynell, A. A. (1991). Adult barrel cortex plasticity occurs at 1 week but not 1 day after vibrissotomy as demonstrated by the 2-deoxyglucose method, *Exp. Neurol.* **113**: 237–248.
- Lewicki, P., Czyzewska, M. & Hoffman, H. (1987). Unconscious acquisition of complex procedural knowledge, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **13**: 523–530.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action, *Behavioral and Brain Sciences* **8**: 529–566.
- Libet, B. (1993a). Epilogue I. Some implications of “time-on” theory, *Neurophysiology of Consciousness: Selected Papers and New Essays by Benjamin Libet*, Birkhäuser, Boston, pp. 385–392.
- Libet, B. (1993b). *Neurophysiology of Consciousness: Selected Papers and New Essays by Benjamin Libet*, Birkhäuser, Boston.
- Libet, B., Pearle, D. K., Morledge, D. E., Gleason, C. A., Hosobuchi, Y. & Barbaro, N. M. (1991). Control of the transition from sensory detection to sensory awareness in man by the duration of a thalamic stimulus: The cerebral ‘time-on’ factor, *Brain* **114**: 1731–1757.

- Libet, B., Wright Jnr., E. W., Feinstein, B. & Pearl, D. K. (1979). Subjective referral of the timing for a conscious experience: A functional role for the somatosensory specific projection system in man, *Brain* **102**: 193–224.
- Lieberman, P. (1984). *The Biology and Evolution of Language*, Cambridge, MA: Harvard University Press.
- Lightfoot, D. (1982). *The Language Lottery*, Cambridge, MA: The MIT Press.
- Lilly, J. C. (1958). Correlations between neurophysiological activity in the cortex and short-term behavior in the monkey, in H. F. Harlow & C. N. Woolsey (eds), *Biological and Biochemical Bases of Behavior*, University of Wisconsin Press, Madison, pp. 83–100.
- Linden, A., Dudbrak, T., Tietz, C. & Weber, F. (1992). An object-oriented framework for the simulation of neural nets, in C. L. Giles, S. J. Hanson & J. D. Cowan (eds), *Advances in Neural Information Processing Systems 5*, San Mateo, CA: Morgan Kaufmann.
- Linden, A. & Tienz, C. (1992). Combining multiple neural network paradigms and applications using SESAME, *Proceedings of the International Joint Conference on Neural Networks, Baltimore, IEEE*.
- Linden, A. & Tienz, C. (n.d.). SESAME—A software environment for combining multiple neural network paradigms and applications. Electronic site: <ftp://ftp.gmd.de/gmd/as/sesame>.
- Linsker, R. (1986). Perceptual neural organization: Some approaches based on network models and information theory, *Annu. Rev. Neurosci.* **13**: 257–281.
- Livingstone, M. & Hubel, D. (1988). Segregation of Form, Color, Movement, and Depth: Anatomy, Physiology, and Perception, *Science* **240**: 740–749. Reproduced in *Biology and Computation: A Physicist's Choice, Reprint Volume*, edited by H. Gutfreund and G. Toulouse. Advanced Series in Neuroscience – Vol. 3; World Scientific, 1994.
- Livingstone, M. S. & Hubel, D. H. (1984). Anatomy and physiology of a color system in the primate visual system, *J. Neurosci.* **4**: 309–356.
- Lo, Z.-P., Yu, Y. & Bavarian, B. (1993). Analysis of the convergence properties of topology preserving neural networks, *IEEE Transactions on Neural Networks* **4**(2): 207–220.
- Locher, P. J. & Nodine, C. F. (1974). The role of scanpaths in the recognition of random shapes, *Perception & Psychophysics* **15**(2): 308–314.
- Locke, J. (1690). *Essay Concerning Human Understanding*, London. Edition by P. H. Nidditch, Oxford, 1975.
- Locke, S. (1978). Motor programming and language behavior, in G. A. Miller & E. Lenneberg (eds), *Psychology and Biology of Language and Thought: Essays in Honor of Eric Lenneberg*, New York: Academic Press, pp. 111–118.

- Luria, A. R. (1961). *The role of speech in the regulation of normal and abnormal behavior*, New York: Liveright.
- Luria, A. R. (1966). *Higher Cortical Functions in Man*, Tavistock Publications, London.
- Luria, A. R. (1970). *Traumatic Aphasia*, Mouton, The Hague.
- Luria, A. R. (1982). *Language and Cognition*, John Wiley & Sons. Edited by James V. Wertsch.
- Lycan, W. (1987). *Consciousness*, Cambridge, MA: MIT Press.
- Lynch, J. C. (1980). The functional organization of the posterior parietal association cortex, *Behavioral and Brain Sciences* **3**: 485–499.
- MacKay, D. M. (1967). Ways of looking at perception, in W. Wathen-Dunn (ed.), *Models for the perception of speech and visual form*, Cambridge, MA: MIT Press, pp. 25–43.
- MacKay, D. M. (1973). Visual stability and voluntary eye movements, in R. Jung (ed.), *Handbook of sensory physiology*, Vol. VII/3A, Berlin: Springer, pp. 307–331.
- MacKay, D. M. (1991). *Behind the Eye*, Basil Blackwell.
- Mackintosh, N. J. (1974). *The Psychology of Animal Learning*, Academic Press, New York.
- Mackworth, N. H. & Morandi, A. J. (1967). The gaze selects informative details within pictures, *Percept. Psychophys.* **2**: 547–552.
- MacLennan, B. (1991). Gabor representations of spatiotemporal visual images, *Neuroprose electronic site: ftp://archive.cis.ohio-state.edu:/pub/neuroprose/macLennan.gabor.ps*. Technical Report CS-91-144, Computer Science Department, University of Tennessee, Knoxville, TN.
- MacNamara, J. (1972). Cognitive basis of language learning in infants, *Psychological Review* **79**: 1–13.
- MacWinney, B., Leinbach, J., Taraban, R. & McDonald, J. (1989). Language learning: Cues or rules?, *Journal of Memory and Language* **28**: 255–277.
- Maier, N. R. F. (1931). Reasoning in humans: II. The solution of a problem and its appearance in consciousness, *Journal of Comparative Psychology* **12**: 181–194.
- Malcolm, N. (1971). The myth of cognitive processes and structures, in T. Mischel (ed.), *Cognitive Development and Epistemology*, New York: Academic Press, pp. 385–392.
- Mammone, R. J. (ed.) (1993). *Artificial Neural Networks for Speech and Vision*, Chapman & Hill.
- Mandler, G. (1975). Consciousness: Respectable, useful and probably necessary, in R. Solso (ed.), *Information Processing and Cognition: The Loyola Symposium*, Hillsdale, NJ: Erlbaum, pp. 229–254.

- Marcel, A. (1983a). Conscious and unconscious perception: Experiments on visual masking and word recognition, *Cognitive Psychology* **15**: 197–237.
- Marcel, A. J. (1980). Surface dyslexia and beginning reading: A revised hypothesis of the pronunciation of print and its impairments, in M. Coltheart, K. Patterson & J. C. Marshall (eds), *Deep dyslexia*, London: Routledge & Kegan Paul, pp. 227–258.
- Marcel, A. J. (1983b). Conscious and Unconscious Perception: An Approach to the Relations between Phenomenal Experience and Perceptual Processes, *Cognitive Psychology* **15**: 238–300.
- Marr, D. (1982). *Vision*, San Francisco: Freeman.
- Marr, D. C. (1978). Representing visual information, in W. E. Reichardt & T. Poggio (eds), *Theoretical Approaches in Neurobiology*, The MIT Press, Cambridge, MA, pp. 151–166.
- Marr, D. & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes, *Proc. R. Soc. Lond.* **200**: 269–294.
- Marr, D. & Poggio, T. (1976). Cooperative computation of stereo disparity, *Science* **194**: 283–287. Reproduced in *Neurocomputing: Foundations of Research*, James A. Anderson and Edward Rosenfeld, eds., Cambridge, MA: The MIT Press, pp. 261–267.
- Marshall, J. A. (1990). Self-organizing neural networks for perception of visual motion, *Neural Networks* **3**: 45–74.
- Marshall, J. C. & Halligan, P. W. (1988). Blindsight and insight in visuo-spatial neglect, *Nature* **336**: 776–777.
- Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies, *Nature* **244**: 522–523.
- Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process, *Science* **189**: 226–228.
- Marslen-Wilson, W. D., Tyler, L. K. & Seidenberg, M. (1978). Sentence processing and the clause boundary, in W. J. M. Levelt & G. B. F. d'Arcais (eds), *Studies in the Perception of Language*, John Wiley & Sons, NY, chapter 7, pp. 219–246.
- Marslen-Wilson, W. D. & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech, *Cognitive Psychology* **10**: 29–63.
- Marslen-Wilson, W. & Tyler, L. K. (1980). The temporal structure of spoken language understanding, *Cognition* **8**: 1–71.
- Massaro, D. W. (1995). Unconscious knowledge of one thing after another, *American Journal of Psychology - Book Reviews* **108**(4): 599–603.

- Maturana, H. R. (1978). Biology of language: The epistemology of reality, in G. A. Miller & E. Lenneberg (eds), *Psychology and Biology of Language and Thought: Essays in Honor of Eric Lenneberg*, New York: Academic Press, pp. 27–63.
- Maturana, H. R. (1980). Biology of cognition, in H. R. Maturana & F. Varela (eds), *Autopoiesis and Cognition: The Realization of the Living*, Dordrecht: Reidel, pp. 2–62.
- Maunsell, J. H., Sclar, G., Nealey, T. A. & DePriest, D. D. (1991). Extraretinal representations in area V4 in the macaque monkey, *Vis. Neurosci.* **7**: 561–573.
- McCarthy, J. & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence, in B. Meltzer & D. Michie (eds), *Machine Intelligence*, Edinburgh: Edinburgh University Press, pp. 463–502.
- McCarthy, R. A. & Warrington, E. K. (1988). Evidence for modality-specific meaning systems in the brain, *Nature* **334**: 428–430.
- McClelland, J. L. & Kawamoto, A. H. (1986). Mechanisms of sentence processing: Assigning roles to constituents of sentences, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models*, The MIT Press: A Bradford Book, chapter 19, pp. 272–325.
- McClelland, J. L. & Rumelhart, D. E. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 2: Psychological and Biological Models, MIT Press, Cambridge, MA: A Bradford Book.
- McClelland, J. L. & Rumelhart, D. E. (1988). *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises*, MIT Press, Cambridge, MA: A Bradford Book.
- McConkie, G. (1979). On the role and control of eye movements in reading, in P. A. Kollers, M. E. Wrolstad & H. Bouma (eds), *Processing of visual language*, New York: Plenum, pp. 37–48.
- McConkie, G. (1990). Where vision and cognition meet. Paper presented at the *H.F.S.P. Workshop on Object and Scene Perception*, Leuven, Belgium.
- McConkie, G. W., Kerr, P. W., Reddix, M. D. & Zola, D. (1988). Eye movement control during reading. I. The location of initial eye fixations on words, *Vision Res.* **28**: 1107–1118.
- McKenna, P. & Warrington, E. K. (1978). Category-specific naming preservation: A single case study, *J. Neurol. Neurosurg. Psychiatry* **41**: 571–574.
- McLeod, W. T. (ed.) (1988). *The New Collins Dictionary and Thesaurus*, William Collins Sons & Co Ltd: London & Glasgow.
- McNamara, J. (1971). The cognitive strategies of language learning, in T. Anderson (ed.), *Conference on child learning*, Quebec: Les Presses de l'Université.
- McNamara, J. (1972). Cognitive basis of language learning in infants, *Psychological Review* **79**: 1–13.

- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoincini, J. & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants, *Cognition* **29**: 143–178.
- Menzel, R. et al. (1984). Biology of invertebrate learning: Group report, in P. Marler & H. S. Terrace (eds), *The Biology of Learning*, Springer-Verlag, pp. 249–270.
- Merleau-Ponty, M. (1962). *Phenomenology of Perception*, Routledge & Kegan Paul. (Reprinted 1976).
- Merzenich, M. M. & Brugge, J. F. (1973). Representation of the cochlear partition on the superior temporal plane of the macaque monkey, *Brain Res.* **50**: 275–296.
- Merzenich, M. M., Jenkins, W. M. & Middlebrooks, J. C. (1984). Observations and hypotheses on special organizational features of the central auditory nervous system, in G. M. Edelman, W. E. Gall & W. M. Cowan (eds), *Dynamic Aspects of Neocortical Function*, John Wiley & Sons: A Neurosciences Institute Publication, chapter 12.
- Merzenich, M. M. & Kaas, J. H. (1982). Organization of mammalian somatosensory cortex following peripheral nerve injury, *Trends Neurosci.* **5**: 434–436.
- Mesulam, M.-M. (1983). The functional anatomy and hemispheric specialization for directed attention, *Trends Neurosci.* **6**: 384–387.
- Mesulam, M.-M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory, *Ann. Neurol.* **28**(5): 597–613.
- Mesulam, M.-M., Hoesen, G. W. V., Pandya, D. N. & Geschwind, N. (1977). Limbic and sensory connections of the inferior parietal lobule (area PG) in the rhesus monkey: A study with a new method for horseradish peroxidase histochemistry, *Brain Res.* **136**: 393–414.
- Metter, E. J., Riege, W. H., Hanson, W. R., Jackson, C. A., Kempler, D. & Van-Lancker, D. (1988). Subcortical structures in aphasia: An analysis based on (F-18)-fluorodeoxyglucose positron emission tomography, and computed tomography, *Arch. Neurol.* **45**: 1229–1234.
- Middleton, D. & Edwards, D. (1990). *Collective Remembering*, London: Sage.
- Miller, G. A. (1962). *Psychology: The Science of Mental Life*, New York: Harper & Row.
- Miller, G. A. (1994). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information, *Psychological Review* **101**(2): 343–352.
- Miller, G. A. & Chomsky, N. (1963). Finitary models of language users, in R. D. Luce, R. R. Bush & E. Galanter (eds), *Handbook of mathematical psychology*, Vol. II, New York: Wiley.
- Miller, K. D., Keller, J. B. & Stryker, M. P. (1989). Ocular dominance column development: Analysis and simulation, *Science* **245**: 605–615.

- Miller, K. D. & Stryker, M. P. (1990). Ocular dominance column formation: Mechanisms and models, *Connectionist Modeling and Brain Function: The Developing Interface*, Vol. 13, MIT Press, Cambridge, MA.
- Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man, *Br. Med. Bull.* **27**: 272–277.
- Milner, B. (1974). Hemispheric specialization: Its scope and limits, in F. O. Schmidt & F. G. Warden (eds), *The Neurosciences: Third Study Program*, Cambridge, MA: MIT Press, pp. 75–89.
- Milton, J., Mundel, T., an der Heiden, U., Spire, J.-P. & Cowan, J. (1995). Traveling activity waves, in M. A. Arbib (ed.), *The Handbook of Brain Theory and Neural Networks*, The MIT Press: A Bradford Book: Cambridge, MA, pp. 994–997.
- Minsky, M. (1967). *Computation: Finite and Infinite Machines*, New York: Prentice-Hall.
- Minsky, M. (1975). A framework for representing knowledge, in P. Winston (ed.), *The psychology of computer vision*, New York: McGraw Hill, chapter 3, pp. 95–128.
- Minsky, M. L. & Papert, S. (1969). *Perceptrons: An Introduction to Computational Geometry*, MIT Press: Cambridge, MA. (Expanded Version 1988), chapter 11 – Learning. Reproduced in *Readings in Machine Learning* (Shavlik & Dietterich 1990), pages 228–241.
- Minsky, M. L. & Papert, S. A. (1988). *Perceptrons: Expanded Edition*, MIT Press, Cambridge, MA. (Original printing 1969).
- Mishkin, M. (1964). Perseveration of central sets after frontal lesions in monkeys, in J. M. Warren & K. Akert (eds), *The Frontal Granular Cortex and Behavior*, McGraw-Hill, New York, pp. 219–241.
- Mishkin, M. (1972). Cortical visual areas and their interactions, in A. G. Karczman & J. C. Eccles (eds), *Brain and Human Behavior*, Springer-Verlag, Berlin, pp. 187–208.
- Mishkin, M., Ungerleider, L. G. & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways, *Trends Neurosci.* **6**: 414–417.
- Miura, T. (1987). Behavior oriented vision: Functional field of view and processing resources, in J. K. O'Regan & A. Lévy-Schoen (eds), *Eye Movements: From Physiology to Cognition*, North-Holland, pp. 563–572.
- Mixon, D. (1991). On not-doing and on trying and failing, in A. Still & A. Costall (eds), *Against Cognitivism: Alternate Foundations for Cognitive Psychology*, Harvester Press, pp. 27–37.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J. J. & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English, *Perception and Psychophysics* **18**: 331–340.

- Mize, R. R. & Murphy, E. H. (1970). Selective visual experience fails to modify receptive field properties of rabbit striate cortex neurons, *Science* **180**: 320–321.
- Mogilner, A., Grossman, J. A. I., Ribory, U., Joliet, M., Volkmann, J., Rapaport, D., Bensley, R. W. & Llinás, R. R. (1993). Somatosensory cortical plasticity in adult humans revealed by magnetoencephalography, *Proc. Natl. Acad. Sci. U.S.A.* **90**: 3593–3597.
- Mohler, C. W., Goldberg, M. E. & Wurtz, R. H. (1973). Visual receptive fields of frontal eye field neurons, *Brain Res.* **61**: 385–389.
- Montague, R. (1974). English as a formal language, in R. H. Thomason (ed.), *Formal Philosophy*, New Haven: Yale University Press.
- Morgan, J. L. & Demuth, K. (eds) (1996). *Signal to Syntax*, Hillsdale, NJ: Erlbaum.
- Morris, R. K., Rayner, K. & Pollatsek, A. (1990). Eye movement guidance in reading: The role of parafoveal letter and space information, *Journal of Experimental Psychology: Human Perception and Performance* **16**(2): 268–281.
- Mountcastle, V. B., Lynch, J. C., Georgopoulos, A., Sakata, H. & Acuna, C. (1975). Posterior parietal association cortex of the monkey: Command functions for operations within extrapersonal space, *J. Neurophysiol.* **38**: 871–909.
- Mozer, M. C. (1993). Neural net architectures for temporal sequence processing, in A. Weigend & N. Gershenfeld (eds), *Predicting the Future and Understanding the Past*, Redwood City, CA: Addison-Wesley.
- Mozer, M. C. & Bachrach, J. (1991). SLUG: A connectionist architecture for inferring the structure of finite-state environments, *Machine Learning* **7**(2/3): 139–160. Reproduced in: “Connectionist Approaches to Language Learning”, David Touretzky (ed.), Kluwer Academic Press, 1991.
- Mumford, D. (1991). On the computational architecture of the neocortex; I. The role of the Thalamo-cortical loop, *Biological Cybernetics* **65**: 135–145. Reproduced in *Biology and Computation: A Physicist's Choice, Reprint Volume*, edited by H. Gutfreund and G. Toulouse. Advanced Series in Neuroscience – Vol. 3; World Scientific, 1994.
- Murphy, J. T., Kwan, H. C., MacKay, W. A. & Wong, Y. C. (1977). Spatial organization of precentral cortex in awake primates, III, Input-output coupling, *J. Neurophysiol.* **41**: 1132–1139.
- Nauta, W. J. H. (1964). Some efferent connections of the prefrontal cortex in the monkey, in J. M. Warren & K. Akert (eds), *The Frontal Granular Cortex and Behavior*, McGraw-Hill, New York, pp. 397–407.
- Nauta, W. J. H. (1971). The problem of the frontal lobe: A reinterpretation, *J. Psychiatr. Res.* **8**: 167–187.
- Nauta, W. J. H. (1972). Neural associations of the frontal cortex, *Acta Neurobiol. Exp.* **32**: 125–140.

- Nazir, T. A. & O'Regan, J. K. (1990). Some results on translation invariance in the human visual system, *Spatial Vision* **5**: 81–100.
- Neisser, U. (1967). *Cognitive Psychology*, New York: Appleton-Century-Crofts.
- Neisser, U. & Becklen, R. (1975). Selective looking: Attending to visually specified events, *Cognitive Psychology* **7**: 480–494.
- Nelson, C. N. & Bignall, K. E. (1973). Interactions of sensory and non-specific thalamic inputs to cortical polysensory units in the squirrel monkey, *Exp. Neurol.* **40**: 189–206.
- Nelson, K. (1973). Structure and strategy in learning to talk, *Monographs of the Society for Research in Child Development* **38**((Serial No. 149)): 1–2.
- Nelson, K. (1974). Concept, word and sentence: Interrelations in acquisition and development, *Psychological Review* **81**: 267–285.
- Newell, A. (1980). Physical symbol systems, *Cognitive Science* **4**: 135–183.
- Newell, A. (1990). *Unified Theories of Cognition*, Harvard Univ. Press, Cambridge, MA.
- Newell, A. & Simon, H. (1981). Computer science as empirical inquiry: Symbols and search, in J. Haugeland (ed.), *Mind Design: philosophy, psychology, artificial intelligence*, Cambridge, MA: MIT Press, pp. 35–66.
- Newell, A. & Simon, H. (1988). GPS, A program that simulates human thought, *Reprinted in Readings in Cognitive Science*, A. Collins and E. E. Smith, eds., Morgan Kaufmann, San Mateo, CA, pp. 453–460.
- Nicolis, G. (1989). Physics of far-from-equilibrium systems and self-organisation, in P. Davies (ed.), *The New Physics*, Cambridge University Press, pp. 316–347.
- Niki, H. (1974). Differential activity of prefrontal units during right and left delayed response trials, *Brain Res.* **70**: 346–349.
- Nisbett, R. E. & Schachter, S. (1966). Cognitive manipulation of pain, *Journal of Experimental Social Psychology* **2**: 227–236.
- Nisbett, R. & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes, *Psychological Review* **84**(3): 231–259.
- Nissen, M. J. & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures, *Cognitive Psychology* **19**: 1–32.
- Norman, D. A. (1993). Cognition in the head and in the world: An introduction to the special issue on situated action, *Cognitive Science* **17**: 1–6.
- Noton, D. & Stark, L. (1971a). Eye movements and visual perception, *Scientific American* **224**(6): 34–43.
- Noton, D. & Stark, L. (1971b). Scanpaths in eye movements during pattern perception, *Science* **171**(3968): 308–311. Jan 22.

- Nye, R. D. (1992). *The Legacy of B. F. Skinner: Concepts and Perspectives, Controversies and Misunderstandings*, Brooks/Cole Publishing Company, CA.
- Olshausen, B. A. & Koch, C. (1995). Selective visual attention, in M. A. Arbib (ed.), *The Handbook of Brain Theory and Neural Networks*, The MIT Press: A Bradford Book: Cambridge, MA, pp. 837–840.
- Olson, C. R. & Freeman, R. D. (1975). Progressive changes in kitten striate cortex during monocular vision, *Journal of Neurophysiology* **38**: 26–33.
- O'Regan, J. K. (1990). Eye movements and reading, in E. Kowler (ed.), *Eye Movements and Their Role in Visual and Cognitive Processes*, Elsevier, pp. 395–453.
- O'Regan, J. K. (1992). Solving the “real” mysteries of visual perception: The world as an outside memory, *Canadian Journal of Psychology* **46**(3): 461–488.
- O'Regan, J. K. (1994). The world as an outside iconic memory—no strong internal metric means no problem of visual stability, *Behavioral and Brain Sciences* **17**: 258–2259. Open Peer Commentary on Target Article: *A theory of visual stability across saccadic eye movements*, by Bruce Bridgeman, A. H. C. Van der Heijden and Boris M. Velichkovsky.
- O'Regan, J. K. & Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: does trans-saccadic fusion exist?, *Vision Res.* **23**(8): 765–768.
- O'Regan, K. (1997). Personal correspondence.
- Ornstein, R. (1991). *The Evolution of Consciousness: The Origins of the Way We Think*, Simon & Schuster: A Touchstone Book.
- O'Rourke, J. (1993). Triple Book Review: Daniel C. Dennett, *Consciousness Explained*, Little Brown, Canada, 1991; Robert Ornstein, *The Evolution of Consciousness: Of Darwin, Freud and Cranial Fire: The Origins of the Way We Think*, Prentice Hall, NY, 1991; William Seager, *Metaphysics of Consciousness*, Routledge, London, 1991, *Artificial Intelligence* **60**: 303–312.
- O'Shea, R. P. (1991). Thumb's rule tested: visual angle of thumb's width is about 2 deg., *Perception* **20**(3): 415–418.
- Osherson, D. N. & Lasnik, H. (eds) (1990). *An Invitation to Cognitive Science*, The MIT Press: A Bradford Book.
- Osherson, D. N. & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts, *Cognition* **9**: 35–58.
- Osherson, D. N. & Smith, E. E. (1982). Gradedness and conceptual combination, *Cognition* **12**: 299–318.
- Osherton, D. N., Stob, M. & Weinstein, S. (1986). *Systems that Learn: An Introduction to Learning*, Cambridge, MA: MIT Press.
- Pandya, D. N., Dye, P. & Butters, N. (1971). Efferent cortico-cortical projections of the prefrontal cortex of the rhesus monkey, *Brain Res.* **31**: 35–46.

- Pandya, D. N. & Kuypers, H. G. J. M. (1969). Cortico-cortical connections in the rhesus monkey, *Brain Res.* **13**: 13–36.
- Pandya, D. N. & Seltzer, B. (1982a). Association areas of the cerebral cortex, *Trends Neurosci.* **5**: 386–390.
- Pandya, D. N. & Seltzer, B. (1982b). Intrinsic connections and architectonics of posterior parietal cortex in the rhesus monkey, *J. Comp. Neurol.* **204**: 196–210.
- Pandya, D. N. & Vignolo, L. A. (1971). Intra- and interhemispheric projections of the precentral, premotor and arcuate areas in the rhesus monkey, *Brain Res.* **26**: 217–233.
- Pandya, D. N. & Yeterian, E. H. (1985). Architecture and connections of cortical association areas, in A. Peters & E. G. Jones (eds), *Cerebral Cortex*, Vol. 4 Association and Auditory Cortices, Plenum Press: NY, chapter 1, pp. 3–61.
- Passingham, R. E. (1972). Visual discrimination learning after selective prefrontal ablations in monkeys (*Macaca mulatta*), *Neuropsychologia* **10**: 27–39.
- Passingham, R. E. & Ettliger, G. (1972). Tactile discrimination learning after selective prefrontal ablations in monkeys (*Macaca mulatta*), *Neuropsychologia* **10**: 17–26.
- Paterson, A. & Zangwill, O. L. (1944). Disorders of visual space perception associated with lesions of the right hemisphere, *Brain* **67**: 331–358.
- Paterson, A. & Zangwill, O. L. (1945). A case of topographical disorientation associated with unilateral cerebral lesion, *Brain* **68**: 188–212.
- Pearce, A. (1996). *Relational Evidence Theory and Spatial Interpretation Procedures*, PhD thesis, School of Computing, Curtin University of Technology, Perth, WA.
- Pearlmutter, B. A. (1989). Learning state space trajectories in recurrent neural networks, *Neural Computation* **1**: 263–269.
- Peterson, M. A. & Gibson, B. (1991). The initial identification of figure-ground relationship: Contributions from shape recognition processes, *Bell. Psychonomic Soc.* **29**: 199–202.
- Petrides, M. (1982). Motor conditional associative-learning after selective prefrontal lesions in the monkey, *Behav. Brain Res.* **5**: 407–413.
- Petrides, M. & Iversen, S. D. (1976). Cross-modal matching and the primate frontal cortex, *Science* **192**: 1023–1024.
- Petrides, M. & Milner, B. (1982). Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man, *Neuropsychologia* **20**: 249–262.
- Petrides, M. & Pandya, D. N. (1983). Projections to the frontal cortex from the posterior parietal region in the rhesus monkey, *Soc. Neurosci. Abstr.* **9**: 40.
- Phaf, R. H., Mul, N. M. & Wolters, G. (1994). A connectionist view of dissociations, *Attention and Performance*, Vol. 15, Amsterdam : North-Holland Pub. Co., chapter 29, pp. 725–751.

- Phillips, G. C. & Wilson, H. R. (1982). Orientation selectivity as revealed by psychophysical masking, *Invest. Ophthal. Visual Sci. Suppl.* **22**: 254.
- Piaget, J. ([1923] 1960). *The Language and Thought of the Child*, 3rd edn, London : Routledge & Kegan Paul.
- Piaget, J. ([1946] 1962). *Play, Dreams, and Imitation in Childhood*, London: Routledge & Kegan Paul.
- Piaget, J. & Garcia, R. (1991). *Towards A Logic Of Meanings*, Lawrence Erlbaum Associates.
- Pineda, F. J. (1988). Dynamics and architecture for neural computation, *J. Complexity* **4**: 216–245.
- Pineda, F. J. (1989). Recurrent backpropagation and the dynamical approach to adaptive neural computation, *Neural Computation* **1**: 161–172.
- Pinker, S. (1984). *Language Learnability and Language Development*, Cambridge, MA: Harvard University Press.
- Pinker, S. (1986). Visual cognition: An introduction, in S. Pinker (ed.), *Visual Cognition*, A Bradford Book: The MIT Press, pp. 1–63. Reprinted from “Cognition: International Journal of Cognitive Psychology”, volume 18.
- Pinker, S. (1989). *Learnability and Cognition*, Cambridge, MA: MIT Press.
- Pinker, S. (1994). *The Language Instinct*, New York, Morrow.
- Pinker, S. (1996). Letters, *Science* **276**.
- Pinto-Hamuy, T. & Linck, P. (1965). Effect of frontal lesions on performance of sequential tasks by monkeys, *Exp. Neurol.* **12**: 96–107.
- Polanyi, M. (1962). *Personal Knowledge : Towards a post-critical philosophy (Rev. ed.)*, London: Routledge & K. Paul. (Also published by: New York: Harper 1964).
- Pollack, J. B. (1991). The induction of dynamical recognizers, *Machine Learning* **7**(2/3): 227–252. Reproduced in: “Connectionist Approaches to Language Learning”, David Touretzky (ed.), Kluwer Academic Press, 1991.
- Pollatsek, A. & Rayner, K. (1989). Reading, in M. I. Posner (ed.), *Foundations of Cognitive Science*, The MIT Press: A Bradford Book: Cambridge, MA, chapter 10, pp. 401–436.
- Pollatsek, A., Rayner, K. & Collins, W. E. (1984). Integrating pictorial information across eye movements, *Journal of Experimental Psychology: General* **113**(3): 426–442.
- Popper, K. R. ([1956] 1983). *Realism and The Aim of Science*, Vol. 1 of the “Postscript to the Logic of Scientific Discovery”, London: Hutchinson. (Edited by W. W. Bartley, III).

- Porat, S. & Feldman, J. A. (1991). Learning automata from ordered examples, *Machine Learning* 7(2/3): 109–138. Reproduced in: “Connectionist Approaches to Language Learning”, David Touretzky (ed.), Kluwer Academic Press, 1991.
- Port, R. F. & van Gelder, T. (eds) (1995). *Mind as Motion: Explorations in the Dynamics of Cognition*, A Bradford Book; The MIT Press: Cambridge, MA.
- Posner, M. I., Cohen, Y. & Rafal, R. D. (1982). Neural systems control of spatial orienting, *Phil. Trans. R. Soc. Lond.* **B 298**: 187–198.
- Posner, M. I. & Synder, C. R. R. (1975). Attention and cognitive control, in R. L. Solso (ed.), *Information Processing and Cognition: The Loyola Symposium*, Hillsdale, N. J.: Erlbaum, pp. 55–85.
- Potter, M. C. (1983). Representational buffers: The eye-mind hypothesis in picture perception, reading, and visual search, in K. Rayner (ed.), *Eye Movement in Reading: Perceptual and Language Processes*, Academic Press, Inc., pp. 413–437.
- Prichard, R. M. (1961). Stabilized images on the retina, *Scientific American* **204**: 72–78.
- Prigatano, G. P. & Schacter, D. L. (1991). *Awareness of Deficit after Brain Injury*, New York: Oxford University Press.
- Prilleltensky, I. (1990). On the social and political implications of cognitive psychology, *Journal of Mind and Behavior* **11**: 127–136.
- Pylyshyn, Z. W. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*, Cambridge, MA: The MIT Press: A Bradford Book.
- Pylyshyn, Z. W. (1989). Computing in cognitive science, in M. I. Posner (ed.), *Foundations of Cognitive Science*, The MIT Press: A Bradford Book: Cambridge, MA, chapter 2, pp. 51–91.
- Raaijmakers, J. G. W. (1993). The story of the two-store model of memory: Past criticisms, current status and future directions, *Attention and Performance XIV*, The MIT Press: Cambridge, MA, pp. 467–488.
- Rager, J. E. (1992). Self-correcting connectionist parsing, in R. G. Reilly & N. E. Sharkey (eds), *Connectionist Approaches to Natural Language Processing*, Lawrence Erlbaum Associates, UK, pp. 143–167.
- Ramachandran, V. S. (1985). Apparent motion of subjective surfaces, *Perception* **14**: 127–134.
- Ramachandran, V. S. (1986). Capture of stereopsis and apparent motion by illusory contours, *Percept. Psychophys.* **39**: 361–373.
- Ramachandran, V. S. (1988). Perception of depth from shading, *Scientific American* **269**: 76–83.

- Ramachandran, V. S. (1992). Perception: A biological perspective, in G. A. Carpenter & S. Grossberg (eds), *Neural Networks for Vision and Image Processing*, A Bradford Book: The MIT Press, pp. 45–91.
- Ramachandran, V. S. & Anstis, S. M. (1983). Perceptual organization in moving patterns, *Nature* **304**: 529–531.
- Ramachandran, V. S. & Anstis, S. M. (1986). Perception of apparent motion, *Scientific American* **254**: 102–109.
- Randolph, M. & Semmes, J. (1974). Behavioral consequences of selective subtotal ablations in the postcentral gyrus of *Macaca mulatta*, *Brain Res.* **70**: 55–70.
- Rayner, K., McConkie, G. W. & Ehrlich, S. (1978). Eye movements and integrating information across fixations, *J. Exp. Psychol. Hum. Percept. Performance* **4**: 529–544.
- Rayner, K. & Pollatsek, A. (1983). Is visual information integrated across saccades?, *Percept. Psychophys.* **34**: 39–48.
- Rayner, K., Slowiaczek, M. L., Clifton, C. & Bertera, J. H. (1983). Latency of sequential eye movements: Implications for reading, *J. Exp. Psychol.: Hum. Percept. Perform.* **9**: 912–922.
- Reale, R. A. & Imig, T. J. (1980). Tonotopic organization in auditory cortex of the cat, *J. Comp. Neurol.* **192**: 265–291.
- Reber, A. S. (1967). Implicit learning of artificial grammars, *Journal of Verbal Learning and Verbal Behavior* **6**: 855–863.
- Reber, A. S. (1989). Implicit learning and tacit knowledge, *Journal of Experimental Psychology: General* **118**: 219–235.
- Reber, A. S. & Allen, R. (1978). Analogy and abstraction strategies in synthetic grammar learning: A functional interpretation, *Cognition* **6**: 189–221.
- Reeke Jr., G. N. & Edelman, G. M. (1984). Selective networks and recognition automata, *Annals of the New York Academy of Sciences* pp. 181–201.
- Reeke Jr., G. N. & Edelman, G. M. (1988). Real brains and artificial intelligence, *Daedalus: Journal of the American Academy of Arts and Sciences* **117**(1): 143–173.
- Reeke Jr., G. N. & Sporns, O. (1990). Selectionist models of perceptual and motor systems and implications for functionalist theories of brain function, *Physica D* **42**: 347–364.
- Reeke Jr., G. N. & Sporns, O. (1991). Neural darwinism and selective recognition automata: How selection shapes perceptual categories, in A. Babloyantz (ed.), *Self-Organization, Emergent Properties, and Learning*, Plenum Press, NY, pp. 199–218.
- Reeke Jr., G. N. & Sporns, O. (1993). Behaviorally based modelling and computational approaches to neuroscience, *Annual Review of Neuroscience* **16**: 597–623.

- Reeke Jr., G. N., Sporns, O. & Edelman, G. M. (1990). Synthetic neural modeling: The "Darwin" series of recognition automata, *Proceedings of the IEEE* **78**(9): 1498–1530.
- Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement, *Classical Conditioning II: Current Research and Theory*, Appleton-Century-Crofts, New York, pp. 64–69.
- Ritter, H. & Kohonen, T. (1989). Self-organizing semantic maps, *Biological Cybernetics* **61**: 241–254.
- Ritter, H., Martinetz, T. & Schulten, K. (1992). *Neural Computation and Self-Organizing Maps*, Addison-Wesley.
- Ritter, H. & Schulten, K. (1988). Convergence properties of Kohonen's topology conserving maps: Fluctuations, stability, and dimension selection, *Biological Cybernetics* **60**: 59–71.
- Robinson, D. A. & Fuchs, A. F. (1969). Eye movements evoked by stimulation of frontal eye fields, *J. Neurophysiol.* **32**: 637–648.
- Robinson, D. L. & Goldberg, M. E. (1978). Sensory and behavioral properties of neurons in posterior parietal cortex of the awake, trained monkey, *Fed. Proc.* **37**: 2258–2261.
- Robinson, W. P. & Rackstraw, S. J. (1967). Variations in mothers' answers to children's questions as a function of social class, verbal intelligence test scores, and sex, *Sociology* **1**: 259–276.
- Rockland, K. S. (1992a). Configuration, in serial reconstruction, of individual axons projecting from area V2 to V4 in the macaque monkey, *Cerebral Cortex* **2**: 353–374.
- Rockland, K. S. (1992b). Laminar distribution of neurons projecting from area V1 to V2 in macaque and squirrel monkeys, *Cerebral Cortex* **2**: 38–47.
- Rockland, K. S. & Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey, *Brain Res.* **179**: 3–20.
- Rockland, K. S. & Pandya, D. N. (1981). Cortical connections of the occipital lobe in the rhesus monkey: Interconnections between areas 17, 18, 19 and the superior temporal gyrus, *Brain Res.* **212**: 249–270.
- Rockland, K. S., Saleem, K. S. & Tanaka, K. (1992). Widespread feedback connections from areas V4 and TEO, *Soc. Neurosci. Abst.* **18**: 390.
- Rogoff, B. (1984). Introduction: Thinking and Learning in Social Context, in B. Rogoff & J. Lave (eds), *Everyday Cognition: Its Development in Social Context*, Harvard University Press, Cambridge, MA, chapter Introduction, pp. 1–8.
- Rogoff, B. & Lave, J. (eds) (1984). *Everyday Cognition: Its Development in Social Context*, Harvard University Press, Cambridge, MA.

- Rosch, E. (1988). Principles of categorization, in A. Collins & E. E. Smith (eds), *Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence*, Morgan Kaufmann, pp. 312–322.
- Rosch, E. & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories, *Cognitive Psychology* 7: 573–605.
- Rosene, D. L. & Pandya, D. N. (1983). Architectonics and connections of the posterior parahippocampal gyrus in the rhesus monkey, *Soc. Neurosci. Abstr.* 9: 222.
- Ross, S. M. & Ross, L. E. (1980). Saccade latency and warning signals: Effects of auditory and visual onset and offset, *Percept. Psychophys.* 29: 429–437.
- Rosvold, H. E. & Mishkin, M. (1961). Non-sensory effects of frontal lesions on discrimination learning and performance, in J. F. Delafresnaye (ed.), *Brain Mechanisms and Learning*, Blackwell, Oxford, pp. 555–567.
- Rumelhart, D. E. (1989). The architecture of mind: A connectionist approach, in M. I. Posner (ed.), *Foundations of Cognitive Science*, The MIT Press: A Bradford Book: Cambridge, MA, chapter 4, pp. 133–159.
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning internal representations by error propagation, in D. E. Rumelhart & J. L. McClelland (eds), *Parallel Distributed Processing*, Vol. 1, MIT Press: Cambridge, MA, chapter 8, pp. 318–362. Reproduced in *Readings in Machine Learning* (Shavlik & Dietterich 1990), pages 115–137.
- Rumelhart, D. E. & McClelland, J. L. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1: Foundations, MIT Press, Cambridge, MA: A Bradford Book.
- Rumelhart, D. E. & McClelland, J. L. (1987). Learning the past tense of English verbs: Implicit rules or parallel distributed processing, in B. MacWhinney (ed.), *Mechanisms of Language Acquisition*, Hillsdale, NJ: Erlbaum.
- Russell, J. (1984). *Explaining Mental Life: Some philosophical issues in psychology*, London: Macmillan.
- Ryle, G. (1949). *The Concept of Mind*, Hutchinson of London.
- Sacks, O. (1986). *The Man Who Mistook His Wife For A Hat*, Picador Books.
- Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996). Statistical learning by 8-month-old infants, *Science* 274: 1926–1928.
- Sakai, M. (1974). Prefrontal unit activity during visually guided lever pressing reaction in the monkey, *Brain Res.* 81: 297–309.
- Sakata, H. (1975). Somatic sensory responses of neurons in the parietal association area (area 5) in monkeys, in H. H. Kornhuber (ed.), *The Somatosensory System*, Thieme, Stuttgart, pp. 250–261.

- Sampson, E. E. (1981). Cognitive psychology as ideology, *American Psychologist* **36**: 730–743.
- Sanides, F. (1969). Comparative architectonics of the neocortex of mammals and their evolutionary interpretation, *Ann. N.Y. Acad. Sci.* **167**: 404–423.
- Sanides, F. (1972). Representation in the cerebral cortex and its areal lamination patterns, in G. F. Bourne (ed.), *Structure and Function of Nervous Tissue*, Vol. 5, Academic Press, New York, pp. 329–453.
- Sapir, E. ([1929] 1949). Male and female forms of speech in Yara, in D. G. Mandelbaum (ed.), *Selected Writings of Edward Sapir*, Berkeley, CA: University of California Press.
- Scaife, M. & Bruner, J. S. (1975). The capacity for joint visual attention in the infant, *Nature* **253**: 265.
- Schacter, D. L. (1987a). Implicit memory: History and current status, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **13**(3): 501–518.
- Schacter, D. L. (1987b). Memory, amnesia, and frontal lobe dysfunction, *Psychobiology* **15**: 21–36.
- Schacter, D. L. (1989a). Memory, in M. I. Posner (ed.), *Foundations of Cognitive Science*, The MIT Press: A Bradford Book: Cambridge, MA, chapter 17, pp. 683–725.
- Schacter, D. L. (1989b). On the relation between memory and consciousness: Dissociable interactions and conscious experience, in H. L. Roediger, III & F. I. M. Craik (eds), *Varieties of Memory and Consciousness: Essays in Honor of Endel Tulving*, Hillsdale, NJ: Erlbaum.
- Schacter, D. L. (1992). Understanding implicit memory: A cognitive neuroscience approach, *Am. Psychol.* **47**: 559–569.
- Schacter, D. L. (1995). Implicit memory: A new frontier for cognitive neuroscience, in M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, A Bradford Book: The MIT Press: Cambridge, MA, pp. 815–824.
- Schacter, D. L., McAndrews, M. P. & Moscovitch, M. (1988). Access to consciousness: Dissociations between implicit and explicit knowledge in neuropsychological syndromes, in L. Weiskrantz (ed.), *Thought Without Language*, Oxford Science Publications, Clarendon Press, chapter 10, pp. 242–278.
- Schacter, S. (1971). *Emotion, Obesity, and Crime*, New York.
- Schaffer, H. R. (1984). *The child's entry into a social world*, New York: Academic Press.
- Schechter, P. B. & Murphy, E. H. (1975). Response characteristics of single cells in squirrel monkey frontal cortex, *Brain Res.* **96**: 66–70.

- Schlesinger, I. M. (1977). The role of cognitive development and linguistic input in language acquisition, *Child Language* 4: 153–169.
- Schneider, G. E. (1969). Two visual systems, *Science* 163: 895–902.
- Schneider, W. E. & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention, *Psychological Review* 84(1): 1–66.
- Schneider, W. E. & Shiffrin, R. M. (1985). Categorization (restructuring) and automatization: Two separable factors, *Psychological Review* 92(3): 424–428.
- Scholtes, J. C. (1991). Recurrent Kohonen Self-Organization in Natural Language Processing, in T. Kohonen, K. Mäkisara, O. Simula & J. Kangas (eds), *Artificial Neural Networks - Volume 2 (Proceedings of the 1991 International Conference on Artificial Neural Networks (ICANN-91), Espoo, Finland, June 1991)*, North-Holland, pp. 1751–1754.
- Schwartz, A. B. (1993). Motor cortical activity during drawing movements, *J. Neurophysiol.* 70: 28–36.
- Searle, J. (1989). *Mind, Brains and Science: The 1984 Reith lectures*, Harmondsworth: Penguin.
- Searle, J. R. (1980). Minds, brains, and programs, *The Behavioral and Brain Sciences* 3. Reproduced in “The Mind’s Eye” by Douglas R. Hofstadter and Daniel C. Dennett, pages 353–382.
- Segall, M. H., Campbell, D. T. & Herskovits, M. J. (1966). *The Influence of Culture on Visual Perception*, The Bobbs-Merrill Company, Inc.
- Segall, M. H., Dasen, P. R., Berry, J. W. & Poortinga, Y. H. (1990). *Human Behavior in Global Perspective: An Introduction to Cross-Cultural Psychology*, Pergamon General Psychology Series; Pergamon Press.
- Seidenberg, M. S. (1996). Language acquisition and use: Learning and applying probabilistic constraints, *Science* 275: 1599–1603.
- Seidl, D. R. & Lorenz, R. D. (1991). A structure by which a recurrent neural network can approximate a nonlinear dynamic system, *Proceedings of the 1991 International Joint Conference on Neural Networks (IJCNN), Seattle, WA, Vol. 2*, pp. 709–714.
- Sekuler, R. & Blake, R. (1990). *Perception (Second Edition)*, McGraw-Hill Publishing Company.
- Selfridge, O. G. (1958). Pandemonium: A paradigm for learning, *Mechanisation of Thought Processes: Proceedings of a Symposium Held at the National Physical Laboratory, London:HMSO*. Reproduced in *Neurocomputing: Foundations of Research*, James A. Anderson and Edward Rosenfeld, eds., Cambridge, MA: The MIT Press, pp. 117–122.
- Seltzer, B. & Pandya, D. N. (1976). Some cortical projections to the parahippocampal area in the rhesus monkey, *Exp. Neurol.* 50: 146–160.

- Seltzer, B. & Pandya, D. N. (1978). Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey, *Brain Res.* **149**: 1–24.
- Seltzer, B. & Pandya, D. N. (1980). Converging visual and somatic sensory cortical input to the intraparietal sulcus of the rhesus monkey, *Brain Res.* **192**: 339–351.
- Servan-Schreiber, D., Cleeremans, A. & McClelland, J. L. (1991). Graded state machines: The representation of temporal contingencies in simple recurrent networks, *Machine Learning* **7**(2/3): 161–193. Reproduced in: "Connectionist Approaches to Language Learning", David Touretzky (ed.), Kluwer Academic Press, 1991.
- Sestito, S. & Dillon, T. (1991). The use of sub-symbolic methods for the automation of knowledge acquisition for expert systems, *Proceedings of AVIGNON 91 Conference, France*.
- Sestito, S. & Dillon, T. S. (1994). *Automated knowledge acquisition*, New York : Prentice Hall.
- Shanon, B. (1988). Semantic representation of meaning: A critique, *Psychological Bulletin* **104**: 70–83.
- Shapiro, S. (ed.) (1987). *Encyclopedia of Artificial Intelligence*, New York : Wiley.
- Sharkey, N. E. & Sharkey, A. J. C. (1996). Separating learning and representation, in S. Wermter, E. Riloff & G. Scheler (eds), *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*, Springer, pp. 17–32.
- Shavlik, J. W. & Dietterich, T. G. (eds) (1990). *Readings in Machine Learning*, Morgan Kaufmann, San Mateo, CA.
- Shaw, R., Turvey, M. & Mace, W. (1982). Ecological psychology: The consequences of a commitment to realism, in W. Weimer & D. Palermo (eds), *Cognition and the symbolic processes*, Vol. 2, Hillsdale, NJ: Erlbaum.
- Sherman, S. M. & Koch, C. (1986). The control of retinogeniculate transmission in the mammalian lateral geniculate nucleus, *Exp. Brain Res.* **63**: 1–20.
- Shiffrin, R. M. (1975). Short-term store: The basis for a memory system, in F. Restle, R. M. Shiffrin, N. J. Castellon, H. Lindman & D. B. Pisoni (eds), *Cognitive Theory*, Vol. 1, Hillsdale, NJ: Erlbaum, pp. 193–218.
- Shiffrin, R. M. (1976). Capacity limitations in information processing, attention and memory, in W. K. Estes (ed.), *Handbook of learning and cognitive processes*, Vol. 4: Memory processes, Hillsdale, NJ: Erlbaum, pp. 177–236.
- Shiffrin, R. M. & Schneider, W. E. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory, *Psychological Review* **84**(2): 127–190.
- Shouval, H. Z. & Perrone, M. P. (1995). Post-Hebbian learning rules, in M. A. Arbib (ed.), *The Handbook of Brain Theory and Neural Networks*, The MIT Press: A Bradford Book: Cambridge, MA, pp. 745–748.

- Simpson, W. E. & Crandall, S. J. (1972). The perception of smiles, *Psychonomic Sci.* **29**: 197–200.
- Sinclair-deZwart, H. (1973). Language acquisition and cognitive development, in T. E. Moore (ed.), *Cognitive Development and the Acquisition of Language*, New York and London: Academic Press.
- Skarda, C. A. & Freeman, W. J. (1987). How brains make chaos in order to make sense of the world, *Behavioral and Brain Sciences* **10**: 161–195.
- Skinner, B. F. (1957). *Verbal behaviour*, New York: Appleton.
- Skinner, B. F. (1977). Why I am not a cognitive psychologist, *Behaviorism* **5**: 1–10.
- Skinner, B. F. (1985). Cognitive science and behaviourism, *British Journal of Psychology* **76**: 291–301.
- Slezak, P. (1993). Situated cognition: Minds in machines or friendly photocopiers. Personal Correspondence.
- Slobin, D. (1971). *Psycholinguistics*, Glenview, IL: Scott, Foresman & Company.
- Slobin, D. I. (1972). Seven questions about language development, in P. C. Dodwell (ed.), *New Horizons in Psychology, II*, Harmondsworth.
- Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar, in C. A. Ferguson & D. I. Slobin (eds), *Studies in Child Language Development*, New York: Holt, Rinehart and Winston, pp. 175–208.
- Smith, A. (1971). Objective indices of severity of chronic aphasia in stroke patients, *Journal of Speech and Hearing Disorders* **36**: 167–207.
- Smith, A. (1978). Lenneberg, Locke, Zangwill, and the neuropsychology of language and language disorders, in G. A. Miller & E. Lenneberg (eds), *Psychology and Biology of Language and Thought: Essays in Honor of Eric Lenneberg*, New York: Academic Press, pp. 133–150.
- Smith, B. C. (1985). Prologue to “Reflections and semantics in a procedural language” (1982), in R. J. Brachman & H. J. Levesque (eds), *Readings in Knowledge Representation*, Morgan Kaufmann, Los Altos, CA.
- Smith, E. E., Shoben, E. J. & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions, *Psychological Review* **81**: 214–241.
- Smith, F. & Miller, G. A. (eds) (1966). *The genesis of language*, Cambridge, MA: The MIT Press.
- Smoliar, S. W. (1989). Book Reviews: Gerald M. Edelman, *Neural Darwinism: The Theory of Neuronal Group Selection*, Basic Books, NY, 1987, *Artificial Intelligence* **39**: 121–139.

- Softky, W. & Koch, C. (1995). Single-cell models, in M. A. Arbib (ed.), *The Handbook of Brain Theory and Neural Networks*, The MIT Press: A Bradford Book: Cambridge, MA, pp. 879–884.
- Sperling, G. (1960). The information available in brief visual presentations, *Psychological Monographs* **74**. (11, Whole No. 498).
- Sperry, R. W. (1966). Brain bisection and mechanisms of consciousness, in J. J. Eccles (ed.), *Brain and Conscious Experience*, New York: Springer, pp. 298–313.
- Sperry, R. W. & Gazzaniga, M. (1967). Language following surgical disconnection of the hemispheres, in C. Millikan & F. Darley (eds), *Brain Mechanisms Underlying Speech and Language*, New York: Grune and Stratton, pp. 108–115.
- Sporns, O., Gally, J. A., Reeke Jr., G. N. & Edelman, G. M. (1989). Reentrant signaling among simulated neuronal groups leads to coherency in their oscillatory activity, *Proc. Natl. Acad. Sci. USA* **86**: 7265–7269.
- Sporns, O., Tononi, G. & Edelman, G. M. (1991). Modeling perceptual grouping and figure-ground segregation by means of active reentrant connections, *Proc. Natl. Acad. Sci. USA* **88**: 129–133.
- St. John, M. F. & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension, *Artificial Intelligence* **46**: 217–257.
- Stanovich, K. E. & West, R. F. (1981). The effect of sentence context on ongoing word recognition: Tests of a two-process theory, *Journal of Experimental Psychology: Human Perception and Performance* **7**: 658–672.
- Stanovich, K. E. & West, R. F. (1983). On priming with a sentence context, *Journal of Experimental Psychology: General* **112**: 1–36.
- Stark, L. W. & Choi, Y. S. (1996). Experimental metaphysics: The scanpath as an epistemological mechanism, *Visual Attention and Cognition*, Elsevier, pp. 3–69.
- Steriade, M., Domich, L. & Oakson, G. (1986). Reticularis thalami neurons revisited: Activity changes during shifts in states of vigilance, *Neurosci.* **6**: 68–81.
- Steriade, M. & Llinas, R. R. (1988). The functional states of the thalamus and the associated neuronal interplay, *Physiol. Rev.* **68**: 649–742.
- Stoerig, P. & Cowey, A. (1989). Wavelength sensitivity in blindsight, *Nature* **342**: 916–918.
- Stone, G. O. (1986). An analysis of the delta rule and the learning of statistical associations, in D. E. Rumelhart & J. L. McClelland (eds), *Parallel Distributed Processing*, Vol. 1, MIT Press: Cambridge, MA, chapter 8, pp. 444–459.
- Stork, F. C. & Widdowson, J. D. A. (1974). *Learning about Linguistics*, Hutchinson Educational.
- Stratton, G. M. (1897a). Upright vision and the retinal image, *Psychol. Rev.* **4**: 182–187.

- Stratton, G. M. (1897b). Vision without inversion of the retinal image, *Psychol. Rev.* **4**: 341–360, 463–481.
- Stryker, M. P. & Harris, W. A. (1986). Binocular impulse blockage prevents the formation of ocular dominance columns in cat visual cortex, *Journal of Neuroscience* **6**: 2117–2133.
- Suchman, L. (1993). Response to Vera and Simon's situated action: A symbolic interpretation, *Cognitive Science* **17**: 71–75.
- Suchman, L. A. (1987). *Plans and Situated Actions: The Problem of Human Machine Communication*, Cambridge University Press.
- Sudnow, D. (19781). *Ways of the Hand*, New York: Bantam.
- Suga, N. (1990). Cortical computational maps for auditory imaging, *Neural Networks* **3**: 3–21.
- Suga, N. & O'Neill, W. E. (1979). Neural axis representing target range in the auditory cortex of the mustache bat, *Science* **206**: 351–353.
- Suzuki, H. & Azuma, M. (1977). Prefrontal neuronal activity during gazing at a light spot in the monkey, *Brain Res.* **126**: 497–508.
- Swindale, N. V. (1980). A model for the formation of ocular dominance stripes, *Proc. R. Soc. Lond.* **B208**: 243–264.
- Swindale, N. V. (1990). Is the cerebral cortex modular?, *Trends Neurosci.* **13**: 487–492.
- Swindale, N. V. (1992). A model for the coordinated development of columnar systems in primate striate cortex, *Biological Cybernetics* **66**: 217–230.
- Takeuchi, A. & Amari, S. (1979). Formation of Topographic Maps and Columnar Microstructures in Nerve Fields, *Biological Cybernetics* **35**: 63–72.
- Talmy, L. (1983). How language structures space, in H. L. Pick, Jr & L. P. Acredolo (eds), *Conference on Spatial Orientation—Theory, Research, and Application*, Plenum Press, NY, pp. 225–282.
- Tanabe, T., Yarita, H., Iino, M., Ooshima, Y. & Takagi, S. F. (1975). An olfactory projection area in orbitofrontal cortex of the monkey, *J. Neurophysiol.* **38**: 1269–1283.
- Tanaka, K. & Saito, H.-A. (1989). Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey, *J. Neurophysiol.* **62**: 626–641.
- Tanaka, K., Saito, H.-A., Fukada, Y. & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey, *J. Neurophysiol.* **66**: 170–189.
- Tanaka, S. (1990). Theory of self-organization of cortical maps: Mathematical framework, *Neural Networks* **3**: 625–640.

- Tavan, P., Grubmuller, H. & Kuhnel, H. (1990). Self-organization of associative memory and pattern classification: Recurrent signal processing on topological feature maps, *Biological Cybernetics* **64**: 95–105.
- Taylor, J. L. & McCloskey, D. I. (1990). Triggering of preprogrammed movements as reactions to masked stimuli, *Journal of Neurophysiology* **63**: 439–446.
- Terrace, H. S. (1985). In the beginning was the “name”, *American Psychologist* **40**(9): 1011–1028.
- Teuber, H. L. (1966). The frontal lobes and their functions: Further observations on rodents, carnivores, subhuman primates, and man, *Int. J. Neurol.* **5**: 282–300.
- Thompson, P. (1980). Margaret Thatcher: A new illusion, *Perception* **9**: 483–484.
- Tigges, J., Spatz, W. B. & Tigges, M. (1973). Reciprocal point-to-point connections between parastriate and striate cortex in the squirrel monkey (*Saimiri*), *J. Comp. Neurol.* **148**: 481–490.
- Tigges, J., Tigges, M., Ansel, S., Cross, N. A., Letbetter, W. D. & McBride, R. L. (1981). Areal and laminar distribution of neurons interconnecting the central visual cortical areas 17, 18, 19 and MT in squirrel monkey (*Saimiri*), *J. Comp. Neurol.* **202**: 539–560.
- Tigges, J., Tigges, M. & Perachio, A. A. (1977). Complementary laminar terminations of afferents to area 17 originating in area 18 and in the lateral geniculate nucleus in the squirrel monkey, *J. Comp. Neurol.* **176**: 87–100.
- Tranel, D. & Damasio, A. R. (1985). Knowledge without awareness: An autonomic index of facial recognition by prosopagnosics, *Science* **228**: 1453–1454.
- Treisman, A. M. & Gelade, G. (1980). A feature integration theory of attention, *Cognitive Psychology* **12**: 97–136.
- Tremblay, J.-P. & Sorenson, P. G. (1985). *The Theory and Practice of Compiler Writing*, McGraw-Hill Book Company.
- Trotter, Y., Celebrini, S., Stricanne, B., Thorpe, S. & Imbert, M. (1992). Modulation of neural stereoscopic processing in primate area V1 by viewing distance, *Science* **257**: 1279–1281.
- Tulving, E. (1985). How many memory systems are there?, *Am. Psychol.* **40**: 385–398.
- Tulving, E. & Schacter, D. L. (1990). Priming and human memory systems, *Science* **247**: 301–306.
- Tunturi, A. R. (1950). Physiological determination of the arrangement of the afferent connections to the middle ectosylvian auditory area in the dog, *Am. J. Physiol.* **162**: 489–502.
- Tunturi, A. R. (1952). The auditory cortex of the dog, *Am. J. Physiol.* **168**: 712–717.

- Turnbull, B. G. & Rasmusson, D. D. (1991). Chronic effects of total or partial digit denervation of racoon somatosensory cortex, *Somatosens. Mot. Res.* **8**: 201–213.
- Turvey, M. T. (1977). Contrasting orientations to the theory of visual information processing, *Psychological Review* **84**: 67–88.
- Ungerleider, L. G. & Mishkin, M. (1982). Two cortical visual systems, in D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (eds), *Advances in the Analysis of Visual Behavior*, MIT Press, Cambridge, MA, pp. 549–586.
- Vaadia, E., Jr, M. H. G., Heinz, R. D. & Benson, D. A. (1983). Single unit activity in frontal cortex of monkeys localizing sound sources, *Soc. Neurosci. Abstr.* **9**: 1084.
- Valiant, L. G. (1984). A theory of the learnable, *Communications of the ACM* **27**: 1134–1142. Reproduced in *Readings in Machine Learning* (Shavlik & Dietterich 1990), pages 192–200.
- Van Essen, D. (1985). Functional organization of primate visual cortex, in A. Peters & E. G. Jones (eds), *Cerebral Cortex*, Vol. 3, New York: Plenum Press, pp. 259–329.
- Van Essen, D. C. & Anderson, C. H. (1989). Information processing strategies and pathways in the primate retina and visual cortex, in S. F. Zornetzer, J. L. Davis & C. Lau (eds), *Introduction to Neural and Electronic Networks*, Academic Press, Orlando, FL, pp. 43–72.
- Van Essen, D. C., Anderson, C. H. & Felleman, D. J. (1992). Information processing in the primate visual system: An integrated systems perspective, *Science* **255**: 419–423.
- Van Essen, D. C. & DeYoe, E. A. (1995). Concurrent processing in the primate visual cortex, in M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, A Bradford Book: The MIT Press: Cambridge, MA, pp. 383–400.
- Van Essen, D. C. & Maunsell, J. H. R. (1983). Hierarchical organization and functional streams in the visual cortex, *Trends Neurosci.* **6**: 370–375.
- Van Hoesen, G. W. (1982). The parahippocampal gyrus, *Trends Neurosci.* **5**: 345–350.
- Van Hoesen, G. W., Pandya, D. N. & Butters, N. (1972). Cortical afferents to the entorhinal cortex of the rhesus monkey, *Science* **175**: 1471–1473.
- Van Hoesen, G. W., Vogt, B. A., Pandya, D. N. & McKenna, T. M. (1980). Compound stimulus differentiation behavior in the rhesus monkey following periarculate ablation, *Brain Res.* **186**: 365–378.
- van Santen, J. P. H. & Sperling, G. (1985). Elaborated Reichardt detectors, *J. Opt. Soc. Am. A* **2**(2): 300–321.
- Velmans, M. (1993). Continuing Commentary: *Commentary on Max Velmans (1991) Is Human Information Processing Conscious?*, *BBS* **14**: 651–726, *Behavioral and Brain Sciences* **16**(2): 404–416.

- Venkatesh, S. S. & Pancha, G. (1990). Shaping attraction basins in neural networks, *Neural Networks* **3**: 613–623.
- Vera, A. H. & Simon, H. A. (1993a). Situated action, *Cognitive Science* **17**: 7–48.
- Vera, A. H. & Simon, H. A. (1993b). Situated action: Reply to reviewers, *Cognitive Science* **17**: 77–86.
- Vera, A. H. & Simon, H. A. (1993c). Situated action: Reply to William Clancey, *Cognitive Science* **17**: 117–133.
- Verschure, P. F. M. J. (1993). Formal minds and biological brains, *IEEE Expert* pp. 66–75.
- Vesalius, A. (1543 [1901]). *De humani corporis fabrica*, Cambridge: Cambridge University Press. (Translated by M. Foster, *Lectures on the history of physiology*).
- Viviani, P. (1990). Eye movements in visual search: Cognitive, perceptual and motor control aspects, in E. Kowler (ed.), *Eye Movements and Their Role in Visual and Cognitive Processes*, Elsevier, pp. 353–393.
- Vogt, B. A. & Pandya, D. N. (1978). Corticocortical connections of somatic sensory cortex (areas 3, 1 and 2) in the rhesus monkey, *J. Comp. Neurol.* **177**: 179–192.
- von der Heydt, R., Peterhans, E. & Baumgartner, G. (1984). Illusory contours and cortical neuron responses, *Science* **224**: 1260–1262.
- von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex, *Kybernetik* **14**: 85–100. Reproduced in *Neurocomputing: Foundations of Research*, James A. Anderson and Edward Rosenfeld, eds., Cambridge, MA: The MIT Press, pp. 212–227.
- von der Malsburg, C. (1981). Internal report, *Technical Report 81-2*, Max Planck Institute for Biophysical Chemistry, Goettingen, F.R.G. Cited in Crick (1984).
- Vygotsky, L. S. ([1962] 1986). *Thought and Language*, Cambridge, MA: MIT Press.
- Vygotsky, L. S. & Luria, A. R. (1993). *Studies on the History of Behavior: Ape, Primitive, and Child*, Hillsdale, NJ: Lawrence Erlbaum. (Trans. Victor I. Golod and Jane E. Knox.).
- Wagman, I. H. & Mehler, W. R. (1972). Physiology and anatomy of the cortico-oculomotor mechanism, *Prog. Brain Res.* **37**: 619–635.
- Waibel, A., Hanazawa, T., Hinton, G., Shikano, K. & Lang, K. J. (1989). Phoneme recognition using time-delay neural networks, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **37**(3): 328–339.
- Walker-Smith, G. J., Gale, A. G. & Findlay, J. M. (1977). Eye movement strategies involved in face perception, *Perception* **6**: 313–326.

- Walley, A. C., Pisoni, D. B. & Aslin, R. N. (1981). The role of early experience in the development of speech perception, in G. J. Whitehurst (ed.), *Development of perception: Psychobiological perspectives*, Vol. 1. Audition, somatic perception, and chemical senses, New York: Academic Press.
- Walzl, E. M. & Woolsey, C. N. (1943). Cortical auditory areas of the monkey as determined by electrical excitation of nerve fibres in the osseous lamina and by click stimulation, *Fed. Proc.* **2**: 52.
- Warren, W. H., J. & Whang, S. (1987). Visual guidance of walking through apertures: Body-scaled information for affordances, *Journal of Experimental Psychology: Human Perception and Performance* **13**: 371–383.
- Warrington, E. K. (1975). The selective impairment of semantic memory, *Q. J. Exp. Psychol.* **27**: 635–657.
- Warrington, E. K. & McCarthy, R. A. (1983). Category specific access dysphasia, *Brain* **106**: 859–878.
- Warrington, E. K. & McCarthy, R. A. (1987). Categories of knowledge, *Brain* **110**: 1273–1296.
- Warrington, E. K. & Shallice, T. (1984). Category-specific impairments, *Brain* **107**: 829–854.
- Webster, K. E. (1965). The cortico-striatal projection in the cat, *J. Anat* **99**: 329–337.
- Weinstein, E. A., Cole, M., Mitchell, M. S. & Lyerly, O. G. (1964). Anosognosia and aphasia, *Archives of Neurology* **10**: 376–386.
- Weinstein, E. A. & Friedland, R. P. (1977). Behavioral disorders associated with hemi-inattention, in E. A. Weinstein & R. P. Friedland (eds), *Advances in Neurology*, Vol. 18, New York: Raven Press, pp. 51–62.
- Weiskrantz (1987). Residual Vision in a Scotoma, *Brain* **110**: 77–92.
- Weiskrantz, L. (1973). *Blindsight: A Case Study and Implications*, Oxford, England: Clarendon.
- Weiskrantz, L. (1980). Varieties of residual experience, *Quarterly Journal of Experimental Psychology* **32**: 365–386.
- Weiskrantz, L. & Mishkin, M. (1958). Effects of temporal and frontal cortical lesions on auditory discrimination in monkeys, *Brain* **81**: 406–414.
- Welch, K. & Stuteville, P. (1958). Experimental production of neglect in monkeys, *Brain* **81**: 341–347.
- Welker, W. I. & Seidenstein, S. (1959). Somatic sensory representation in the cerebral cortex of the racoon (*Procyon lotor*), *J. Comp. Neurol.* **111**: 469–501.
- Westheimer, G. & McKee, S. P. (1977). Spatial configurations for visual hyperacuity, *Vision Res.* **17**: 941–947.

- Wexler, K. & Culicover, P. (1980). *Formal Principles of Language Acquisition*, Cambridge, MA: MIT Press.
- Weyland, T. G. & Malpeli, J. G. (1989). Responses of neurons in primary visual cortex are influenced by eye position, *Soc. Neurosci. Abstr.* **15**: 1016.
- White, L. (1991). Second language competence versus second language performance: UG or processing strategies, in L. Eubank (ed.), *Point Counterpoint: Universal Grammar in the Second Language*, Language Acquisition & Language Disorders 3, John Benjamins Publishing Co., Amsterdam/Philadelphia, pp. 167–189.
- Whitehead, S. D. & Ballard, D. (1990). Active perception and reinforcement learning, *Neural Comp.* **2**: 409–419.
- Whitehead, S. D. & Ballard, D. (1991). Connectionist designs on planning, *Neural Information Processing Systems 3*, Morgan Kaufmann, San Mateo, CA, pp. 357–370.
- Whorf, B. L. (1956). *Language, Thought, and Reality*, New York: Wiley.
- Whorf, B. L. (1971). Languages and Logic, in J. B. Carroll (ed.), *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, The MIT Press, Cambridge, MA, pp. 233–245. Reprinted from *Technol. Rev.*, 43:250–252, 266, 268, 272, 1941.
- Wickens, C. (1980). The structure of attentional processes, in R. S. Nickerson (ed.), *Attention and Performance*, Hillsdale, N. J.: Erlbaum.
- Wickens, C. (1984). Processing resources in attention, in R. Parasuraman & D. R. Davies (eds), *Varieties of Attention*, New York: Academic Press.
- Wiesendanger, M. (1981). Organization of secondary motor areas of cerebral cortex, in V. B. Brooks (ed.), *Nervous System*, Vol. 2 (Part 2), American Physiological Society, Bethesda, pp. 1121–1147.
- Williams, M. (1985). Wittgenstein's rejection of scientific psychology, *Journal for the Theory of Social Behaviour* **15**: 203–223.
- Williams, R. J. & Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks, *Neural Computation* **1**(2): 270–280.
- Willingham, D. B., Nissen, M. J. & Bullemer, P. (1989). On the development of procedural knowledge, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **15**: 1047–1060.
- Wilson, H. R. (1986). Responses Of Spatial Mechanisms Can Explain Hyperacuity, *Vision Research* **26**(3): 453–469.
- Wilson, H. R. & Gelb, D. J. (1984). Modified line-element theory for spatial-frequency and width discrimination, *J. Opt. Soc. Am. A* **1**(1): 124–131.
- Wilson, H. R. & Regan, D. (1984). Spatial-frequency adaption and grating discrimination: predictions of a line-element model, *J. Opt. Soc. Am. A* **1**(11): 1091–1096.

- Winograd, T. (1975). Frame representations and the declarative-procedural controversy, in D. G. Bobrow & A. M. Collins (eds), *Representation and Understanding: Studies in Cognitive Science*, New York: Academic Press.
- Winograd, T. (1983). *Language as a Cognitive Process*, Vol. 1: Syntax, Addison-Wesley Publishing Company.
- Winograd, T. & Flores, F. (1986). *Understanding Computers and Cognition: A New Foundation for Design*, Addison-Wesley.
- Winston, P. H. (1975). Learning structural descriptions from examples, in P. H. Winston (ed.), *The psychology of computer vision*, New York: McGraw Hill.
- Wittgenstein, L. ([1948] 1982). *Last Writings on the Philosophy of Psychology*, Vol. 1 (trans. corrected), Chicago: University of Chicago Press.
- Wittgenstein, L. (1960a). *The Blue and Brown Books*, Oxford, Eng: Basil Blackwell.
- Wittgenstein, L. (1960b). *Tractatus Logico-Philosophicus*, London: Routledge & Kegan Paul.
- Wittgenstein, L. (1963). *Philosophical Investigations*, Basil Blackwell: Oxford.
- Woolgar, S. (1987). Reconstructing man and machine: A note on sociological critiques of cognitivism, in W. E. Bijke, T. P. Hughes & T. J. Pinch (eds), *The Social Construction of Technological Systems: New directions in the sociology and history of technology*, Cambridge, MA: MIT Press, pp. 311–328.
- Woolsey, C. N. (1958). Organization of somatic sensory and motor areas of the cerebral cortex, in H. F. Harlow & C. N. Woolsey (eds), *Biological and Biochemical Bases of Behavior*, University of Wisconsin Press, Madison, pp. 63–81.
- Wright, C. E. (1990). Controlling sequential motor activity, in D. N. Osherson, S. M. Kosslyn & J. M. Hollerbach (eds), *An Invitation to Cognitive Science*, Vol. 2—Visual Cognition and Action, The MIT Press: A Bradford Book, chapter 5, pp. 285–316.
- Yamadori, A. & Albert, M. L. (1973). Word category aphasia, *Cortex* 9: 112–125.
- Yarbus, A. L. (ed.) (1967). *Eye Movements and Vision*, New York: Plenum Press.
- Yin, R. K. (1969). Looking at upside-down faces, *Journal of Experimental Psychology* 81(1): 141–145.
- Young, M. P. (1992). Objective analysis of the topological organization of the primate cortical visual system, *Nature* 358: 152–155.
- Youniss, J. & Damon, W. (1992). Social construction in Piaget's theory, in H. Beilin & P. B. Pufall (eds), *Piaget's Theory: Prospects and Possibilities*, Lawrence Erlbaum Associates, Publ: NJ, chapter 12.
- Zangwill, O. L. (1978). Aphasia and the concept of brain centers, in G. A. Miller & E. Lenneberg (eds), *Psychology and Biology of Language and Thought: Essays in Honor of Eric Lenneberg*, New York: Academic Press, pp. 119–132.

- Zeki, S. (1980). The representation of colours in the cerebral cortex, *Nature* **284**: 412–418.
- Zeki, S. (1983). The distribution of wavelength and orientation selective cells in different areas of monkey visual cortex, *Proc. R. Soc. Lond. [Biol.]* **217**: 449–470.
- Zeki, S. (1993). *A Vision of the Brain*, Oxford: Blackwell Scientific Publications.
- Zeki, S. M. (1978). Functional specialization in the visual cortex of the rhesus monkey, *Nature* **274**: 423–428.
- Zeki, S. & Shipp, S. (1988). The functional logic of cortical connections, *Nature* **335**: 311–317.
- Zeng, Z., Goodman, R. M. & Smyth, P. (1994). Discrete recurrent neural networks for grammatical inference, *IEEE Transactions on Neural Networks* **5**(2): 320–330.
- Zetsche, C. & Caelli, T. (1989). Invariant Pattern Recognition Using Multiple Filter Image Representations, *Computer Vision, Graphics, and Image Processing* **45**: 251–262.
- Zhang, J. (1989). Dynamical self-organization and formation of cortical maps, *Proc. IJCNN'89 Int. Joint Conf. on Neural Networks III*: 487–492.
- Zihl, J. (1980). 'Blindsight': Improvement of visually guided eye movements by systematic practice in patients with cerebral blindness, *Neuropsychologia* **18**: 71–77.
- Zrenner, E., Abramov, I., Akita, M., Cowley, A., Livingstone, M. & Valberg, A. (1990). Color perception: Retina to cortex, in L. Spillman & J. S. Werner (eds), *Visual Perception: The Neurophysiological Foundations*, Academic Press, pp. 163–204.
- Zuber, B. L., Stark, L. & Cook, G. (1965). Microsaccades and the velocity amplitude relationship for saccadic eye movement, *Science* **15**: 1459–1460.