An Influence Network Model to Study Discrepancies in Expressed and Private Opinions [⋆]

Mengbin Ye [a,b], Yuzhen Qin [a], Alain Govaert [a], Brian D.O. Anderson [b,c,d], Ming Cao [a]

[a] *Faculty of Science and Engineering, University of Groningen, The Netherlands*

[b] *Research School of Electrical, Energy and Materials Engineering, the Australian National University, Canberra, A.C.T. 2601, Australia*

[c] *School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China*

[d] *Data61-CSIRO, Canberra, A.C.T. 2601, Australia*

Abstract

In many social situations, a discrepancy arises between an individual's private and expressed opinions on a given topic. Motivated by Solomon Asch's seminal experiments on social conformity and other related socio-psychological works, we propose a novel opinion dynamics model to study how such a discrepancy can arise in general social networks of interpersonal influence. Each individual in the network has *both a private and an expressed opinion*: an individual's private opinion evolves under social influence from the expressed opinions of the individual's neighbours, while the individual determines his or her expressed opinion under a pressure to conform to the average expressed opinion of his or her neighbours, termed the *local public opinion*. General conditions on the network that guarantee exponentially fast convergence of the opinions to a limit are obtained. Further analysis of the limit yields several semi-quantitative conclusions, which have insightful social interpretations, including the establishing of conditions that ensure every individual in the network has such a discrepancy. Last, we show the generality and validity of the model by using it to explain and predict the results of Solomon Asch's seminal experiments.

Key words: opinion dynamics; social network analysis; networks; agent-based model; social conformity

1 Introduction

The study of dynamic models of opinion evolution on social networks has recently become of interest to the systems and control community. Most models are agent-based, in which the opinion(s) of each individual (agent) evolve via interaction and communication with neighbouring individuals. This paper aims to develop a novel opinion dynamics model as a general theoretical framework to study how discrepancies arise in individuals' private and expressed opinions, and thus bridge the current gap between socio-psychological studies on conformity and dynamic models of interpersonal influence. Interested readers are referred to [1–3] for surveys on the many works on opinion dynamics models.

Discrepancies in private and expressed opinions of individuals can arise in many situations, with a variety of consequential phenomena. Over one third of jurors in criminal trials would have privately voted against the final decision of their jury [4]. Large differences between a population's private and expressed opinions can create discontent and tension, a factor associated with the Arab Spring movement [5] and the fall of the Soviet Union [6]. Access to the public action of individuals, without being able to observe their thoughts, can create informational cascades where all subsequent in-

dividuals select the wrong action [7]. Other phenomena linked to such discrepancies include pluralistic ignorance, where individuals privately reject a view but believe the majority of other individuals accept it [8], the "spiral of silence" [9,10], and enforcement of unpopular social norms [11,12]. Whether occurring in a jury panel, a company boardroom or in the general population for a sensitive political issue, the potential societal ramifications of large and persistent discrepancies in private and expressed opinions are clear, and serve as a key motivator for our investigations.

1.1 Existing Work

Conformity: Empirical Data and Static Models. One common reason such discrepancies arise is a pressure on an individual to conform in a group situation; formal study of such phenomena goes back over six decades. In 1951, Solomon E. Asch's seminal paper [13] showed an individual's public support for an indisputable fact could be distorted due to the pressure to conform to a unanimous group of others opposing this fact. Asch's work was among the many studies examining the effects of pressures to conform to the group standard or opinion, using both controlled laboratory experiments and data gathered from field studies. Many of the lab experiments focus on Asch-like studies, perhaps with various modifications. A meta-analysis of 125 such studies was presented in [14]. Pluralistic ignorance is often associated with pressures to conform to social norms [8,15,16]. With a focus on the seminal Asch experiments, a number of *static* models were proposed to describe a single individual conforming to a unanimous majority [17–19], with obvious common limitations in generalisation to dynamics on social networks.

Opinion Dynamics Models. Agent-based models (ABMs) have proved to be both versatile and powerful, with simple agent-level dynamics leading to interesting emergent network-level social phenomena. The seminal French–DeGroot model [20,21] showed that a network of individuals can reach a consensus of opinions via weighted averaging of their opinions, a mechanism modelling "social influence". Indeed, the term "influence network" arose to reflect the social influence exerted via the interpersonal network. Since then, the roles of homophily [22,23], bias assimilation [24], social distancing [25], and antagonistic interactions [26,27] in generating clustering, polarisation, and disagreement of opinions in the social network have also been studied. Individuals who remain somewhat attached to their initial opinions were introduced in the Friedkin–Johnsen model [28] to explain the persistent disagreements observed in real communities. However, a key assumption in most existing ABMs (including those above), is that *each individual has a single opinion* for a given topic. These models are unable to capture phenomena in which an individual holds, for the same topic, a private opinion different to

the opinion he or she expresses. A few complex ABMs do exist in which each agent has both an expressed opinion and a private opinion for the same topic. The work [11] studies norm enforcement and assumes that each agent has two binary variables representing private and public acceptance or rejection of a norm. We are motivated to consider opinions as continuous variables to better capture *discrepancies* in expressed and private opinions, since an individual's opinion may range in its intensity. The model in [29] does assume the expressed and private opinions take values in a continuous interval, but is extremely complex and nonlinear. The properties of the models in [11,29] have only been partially characterised by simulation-based analysis, which is computationally expensive if detailed analysis is desired.

We seek to expand from [11,29] to build an ABM of lower complexity that is still powerful enough to capture how discrepancies in expressed and private opinions might evolve in social networks, and to allow study by theoretical analysis, as opposed to only by simulation. Importantly also, a minimal number of parameters per agent makes data fitting and parameter estimation in experimental investigations a tractable process, as highlighted by the successful validations of the Friedkin–Johnsen model [30–32], whereas experiments for more complicated models are rare.

1.2 Contributions of This Paper

In this paper, we aim to bridge the gap between the literature on conformity and the opinion dynamics models, by proposing a model where each individual (agent) has both a private and an expressed opinion. Inspired by the Friedkin–Johnsen model, we propose that an individual's private opinion evolves under social influence exerted by the individual's network neighbours' expressed opinions, but each individual remains attached to his or her initial opinion with a level of stubbornness. Then, and motivated by existing works on the pressures to conform in a group situation, we propose that each individual has some resilience to this pressure, and each individual expresses an opinion *altered* from his or her private opinion to be closer to the average expressed opinion.

Rigorous analysis of the model is given, leading to a number of semi-quantitative conclusions with insightful social interpretations. We show that for strongly connected networks and almost all parameter values for stubbornness and resilience, individuals' opinions converge exponentially fast to a steady-state of persistent disagreement. We identify that the combination of (i) stubbornness, (ii) resilience, and (iii) connectivity of the network generically leads to every individual having a discrepancy between his or her limiting expressed and private opinions. We give a method for underbounding the disagreement among the limiting private opinions given limited knowledge of the network, and show that a

change in an individual's resilience to the pressure has a propagating effect on every other individual's expressed opinion. Last, we apply our model to the seminal experiments on conformity by Asch [13]. Asch recorded 3 different types of responses among test individuals who must choose between expressing support for an indisputable fact and siding with a unanimous majority claiming the fact to be false. We identify stubbornness and resilience parameter ranges for all 3 responses; this capturing of all 3 responses is a first among ABMs, and underlines our model's strength as a general framework for studying the evolution of expressed and private opinions.

Our work extends from (i) the static models of conformity, by generalising to opinion dynamics on arbitrary networks, and (ii) the dynamic agent-based models, by introducing mechanisms inspired by socio-psychological literature to model the expressed and private opinions of each individual separately. The result is a general modelling framework, which is shown to be consistent with empirical data, and may be used to further the study of phenomena involving discrepancies in private and expressed opinions in social networks.

The rest of the paper is structured as follows. The model is presented in Section 2, with theoretical results detailed in Section 3. Section 4 applies the model to Asch's experiments, with concluding remarks given in Section 5.

2 A Novel Model of Opinion Evolution Under Pressure to Conform

Before introducing the model, we define some notation, and introduce graphs, which are used to model the network of interpersonal influence.

Notations: The n-column vector of all ones and zeros is given by $\mathbf{1}_n$ and $\mathbf{0}_n$ respectively. The $n \times n$ identity matrix is given by $\boldsymbol{I}_n$. For a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times m}$ (respectively a vector $\boldsymbol{a} \in \mathbb{R}^n$), we denote the $(i,j)^{th}$ element as a_{ij} (respectively the ith element as a_i). A matrix $\boldsymbol{A}$ is said to be nonnegative, denoted by $\boldsymbol{A} \geq 0$ (respectively positive, denoted by $\boldsymbol{A} > 0$) if all of its entries a_{ij} are nonnegative (respectively positive). A nonnegative matrix $\boldsymbol{A}$ is said to be row-stochastic (respectively row-substochastic) if for all i, there holds $\sum_{j=1}^n a_{ij} = 1$ (respectively $\sum_{j=1}^n a_{ij} \leq 1$ and $\exists k : \sum_{j=1}^n a_{kj} < 1$).

Graphs: Given any nonnegative not necessarily symmetric $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, we can associate with it a graph $\mathcal{G}[\boldsymbol{A}] = (\mathcal{V}, \mathcal{E}[\boldsymbol{A}], \boldsymbol{A})$. Here, $\mathcal{V} = \{v_1, \ldots, v_n\}$ is the set of nodes, with index set $\mathcal{I} = \{1, \ldots, n\}$. An edge $e_{ij} = (v_i, v_j)$ is in the set of ordered edges $\mathcal{E}[\boldsymbol{A}] \subseteq \mathcal{V} \times \mathcal{V}$ if and only if $a_{ji} > 0$. The edge e_{ij} is said to be incoming with respect to j and outgoing with respect to i. We allow self-loops, i.e. e_{ii} is allowed to be in $\mathcal{E}$. The neighbour set of v_i is denoted by $\mathcal{N}_i = \{v_j \in \mathcal{V} : (v_j, v_i) \in \mathcal{E}\}$. A directed path

is a sequence of edges of the form $(v_{p_1}, v_{p_2}), (v_{p_2}, v_{p_3}), \ldots$, where $v_{p_i} \in \mathcal{V}, e_{p_j p_k} \in \mathcal{E}$. A graph $\mathcal{G}[\boldsymbol{A}]$ is strongly connected if and only if there is a path from every node to every other node [33], or equivalently, if and only if $\boldsymbol{A}$ is irreducible [33]. A cycle is a directed path that starts and ends at the same vertex, and contains no repeated vertex except the initial (also the final) vertex, and a directed graph is *aperiodic* if there exists no integer $k > 1$ that divides the length of every cycle of the graph [34].

We are now ready to propose the agent-based model. For a population of n individuals, let $y_i(t) \in \mathbb{R}$ and $\hat{y}_i(t) \in \mathbb{R}$, $i = 1, \ldots, n$, represent, at time $t = 0, 1, \ldots$, individual i's private and expressed opinions on a given topic, respectively. In general, $y_i(t)$ and $\hat{y}_i(t)$ are not the same, and *we regard y_i as individual i's true opinion.* Individual i may refrain from expressing $y_i(t)$ for many reasons, e.g. political correctness when discussing a sensitive topic. For instance, *preference falsification* [35] occurs when an individual falsifies his or her view due to social pressure (be it imaginary or real), or deliberately, e.g. by a politician seeking to garner votes. In our model, an individual falsifies his or her opinion due to a pressure to conform to the group average opinion. The terms "opinion", "belief", and "attitude" all appear in the literature, with various related definitions. Our model is general enough to cover all these terms, since in all such instances, one can scale $y_i(t), \hat{y}_i(t)$ to be in some real interval $[a, b]$, where a and b represent the two extreme positions on the topic. For consistency, we will *only use "opinion"* unless explicitly stated otherwise.

The individuals discuss their expressed opinions $\hat{y}_i(t)$ over a network described by a graph $\mathcal{G}[\boldsymbol{W}]$, and as a result, their private and expressed opinions, $y_i(t)$ and $\hat{y}_i(t)$ evolve in a process qualitatively described in Fig. 1. Formally, individual i's private opinion evolves as

$$y_i(t+1) = \lambda_i w_{ii} y_i(t) + \lambda_i \sum_{j \neq i}^n w_{ij} \hat{y}_j(t) + (1-\lambda_i) y_i(0) \quad (1)$$

and expressed opinion $\hat{y}_i(t)$ is determined according to

$$\hat{y}_i(t) = \phi_i y_i(t) + (1 - \phi_i) \hat{y}_{i,\text{lavg}}(t-1). \quad (2)$$

In Eq. (1), the influence weight that individual i accords to individual j's expressed opinion $\hat{y}_j(t)$ is captured by $w_{ij} \geq 0$, satisfying $\sum_{j=1}^n w_{ij} = 1$ for all $i \in \mathcal{I}$. The term $w_{ii} \geq 0$ represents the self-confidence (if any) of individual i in i's own private opinion [1]. The constant $\lambda_i \in [0, 1]$ represents individual i's *susceptibility to interpersonal influence* changing i's private opinion

---

[1] In most situations, one can assume $w_{ii} > 0$, and models for studying the dynamics of w_{ii} exist [36,37]. Presence of $w_{ii} > 0$ can also ensure convergence of the opinions, e.g. in the DeGroot model [1].

$(1 - \lambda_i$ is thus i's stubbornness regarding initial opinion $y_i(0)$). Individual i is maximally or minimally susceptible if $\lambda_i = 1$ or $\lambda_i = 0$, respectively. In Eq. (2), the quantity $\hat{y}_{i,\text{lavg}}(t) = \sum_{j \in \mathcal{N}_i} m_{ij} \hat{y}_i(t)$ is specific to individual i, and includes only the expressed $\hat{y}_j(t)$ of i's neighbours. We assume that the weight $m_{ij} \geq 0$ satisfies $w_{ij} > 0 \Leftrightarrow m_{ij} > 0$ and $\sum_{j \in \mathcal{N}_i} m_{ij} = 1$; the matrix $\boldsymbol{M} = \{m_{ij}\}$ is therefore row-stochastic and $\mathcal{G}[\boldsymbol{M}]$ has the same connectivity properties as $\mathcal{G}[\boldsymbol{W}]$. A natural choice is $m_{ij} = |\mathcal{N}_i|^{-1}$ for all $j : e_{ji} \in \mathcal{E}[\boldsymbol{W}]$, while a reasonable alternative is $m_{ij} = w_{ij}, \forall i, j \in \mathcal{I}$. Thus, $\hat{y}_{i,\text{lavg}}(t)$ represents the group standard or norm as viewed by individual i at time t, and is termed the *local public opinion* as perceived by individual i. The constant $\phi_i \in [0, 1]$ encodes individual i's *resilience to pressures to conform* to the local public opinion (maximally 1, and minimally 0), or *resilience* for short. The initial expressed opinion is set to be $\hat{y}_i(0) = y_i(0)$, which means Eq. (1) comes into effect for $t = 1$. As it turns out, under mild assumptions on λ_i, the final opinion values are dependent on $y_i(0)$ but independent of $\hat{y}_i(0)$; one could also select other initialisations for $\hat{y}_i(0)$ with the final opinions unchanged (though the transient would change).

Sociology literature indicates that the pressure to conform causes an individual to express an opinion that is in the direction of the perceived group standard [13,38,10], which in our model is $\hat{y}_{i,\text{lavg}}(t)$. Some pressures of conformity may derive from unspoken traditions [39], or a fear or being different [13], and others arise because of a desire to be in the group, driven by e.g. monetary incentives, status or rewards [40]. Thus, Eq. (2) aims to capture individual i expressing an opinion equal to i's private opinion *modified or altered due to normative pressure* (proportional to $1 - \phi_i$) to be closer to the public opinion as perceived by individual i, which exerts a "force" $(1 - \phi_i)\hat{y}_{i,\text{lavg}}(t - 1)$. Heterogeneous ϕ_i captures the fact that some individuals are less inhibited/reserved than others when expressing their opinions. In addition, pressures are exerted (or perceived to be exerted), differentially for individuals, e.g. due to status [41,38].

Remark 1 *Use of a local public opinion $\hat{y}_{i,lavg}(t)$ ensures the model's scalability to large networks, but in small networks, e.g. a boardroom of 10 people, one could replace $\hat{y}_{i,lavg}(t)$ with the global public opinion $\hat{y}_{avg}(t) = \frac{1}{n} \sum_{j=1}^n \hat{y}_j(t)$ since it is likely to be discernible to every individual. It turns out that all but one of the high-level theoretical conclusions, including convergence, do not depend on the choice of weights of the local public opinion, nor on whether a local or global public opinion is used. However, preliminary observations in [42, Chapter 4] show that the distribution of the final opinion values can vary significantly depending on the aforementioned choices, and we leave characterisation of the difference to future investigations.*

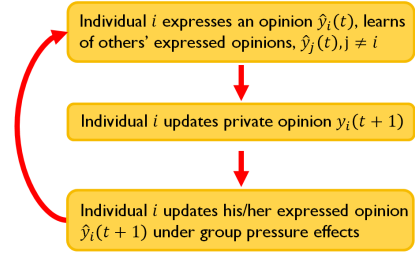**Remark 2** *A key feature in our model, departing from*

Fig. 1. The discussion process. Each individual i, at time step t, expresses opinion $\hat{y}_i(t)$ and learns of others' expressed opinions $\hat{y}_j(t), j \neq i$. Next, the privately held opinion $y_i(t+1)$ evolves according to Eq. (1). After this, individual i then determines the new $\hat{y}_i(t + 1)$ to be expressed in the next round of discussion, according to Eq. (2). Reprinted from [42] with permission by **SpringerNature**.

most existing models, is the associating of two states $y_i, \hat{y}_i$ for each individual and the restriction that only other $\hat{y}_j$ (and no y_j) may be available to individual i. Importantly, note that $\hat{y}_i(t)$ evolves dynamically via Eq. (2); $\hat{y}_i(t)$ is not simply an output variable. However, notice that setting $\phi_i = 1$ for all i recovers the Friedkin–Johnsen model, while $\phi_i = \lambda_i = 1$ for all i, recovers the DeGroot model [21]. One may also notice the time-shift in Eq. (2) of $\hat{y}_{i,lavg}(t-1)$, which ensures that Eq. (2) is consistent with the qualitative process described in Fig. 1. Thus, Eq. (2) aims to capture a natural manner, widely supported by sociology literature, in which an individual determines his or her expressed opinion under a pressure to conform.

2.1 The Networked System Dynamics

We now obtain a matrix form equation for the dynamics of all individuals' opinions on the network. Let $\boldsymbol{y} = [y_1, y_2, \ldots, y_n]^\top$ and $\hat{\boldsymbol{y}} = [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n]^\top$ be the stacked vectors of private and expressed opinions y_i and $\hat{y}_i$ of the n individuals in the influence network, respectively. The influence matrix $\boldsymbol{W}$ can be decomposed as $\boldsymbol{W} = \widetilde{\boldsymbol{W}} + \widehat{\boldsymbol{W}}$ where $\widetilde{\boldsymbol{W}}$ is a diagonal matrix with diagonal entries $\tilde{w}_{ii} = w_{ii}$. The matrix $\widehat{\boldsymbol{W}}$ has entries $\hat{w}_{ij} = w_{ij}$ for all $j \neq i$ and $\hat{w}_{ii} = 0$ for all i. Define $\boldsymbol{\Lambda} = \text{diag}(\lambda_i)$ and $\boldsymbol{\Phi} = \text{diag}(\phi_i)$. Substituting $\hat{y}_j(t)$ from Eq. (2) into Eq. (1), and recalling that $\hat{y}_{i,\text{lavg}} = \sum_{j \in \mathcal{N}_j} m_{ij} \hat{y}_j$, yields

$$y_i(t+1) = \lambda_i w_{ii} y_i(t) + \lambda_i \sum_{j \neq i}^n w_{ij} \phi_j y_j(t) + (1 - \lambda_i) y_i(0)$$

$$+ \lambda_i \sum_{j \neq i}^n w_{ij}(1 - \phi_j) \sum_{k \in \mathcal{N}_j} m_{jk} \hat{y}_k(t - 1).$$

$$(3)$$

From Eq. (3) and Eq. (2), one obtains

$$\begin{bmatrix} \boldsymbol{y}(t+1) \\ \hat{\boldsymbol{y}}(t) \end{bmatrix} = \boldsymbol{P} \begin{bmatrix} \boldsymbol{y}(t) \\ \hat{\boldsymbol{y}}(t-1) \end{bmatrix} + \begin{bmatrix} (\boldsymbol{I}_n - \boldsymbol{\Lambda})\,\boldsymbol{y}(0) \\ \boldsymbol{0}_n \end{bmatrix}, \quad (4)$$

where $\boldsymbol{P}$ consists of the following block matrices

$$\begin{bmatrix} \boldsymbol{\Lambda}(\widetilde{\boldsymbol{W}} + \widehat{\boldsymbol{W}}\boldsymbol{\Phi}) & \boldsymbol{\Lambda}\widehat{\boldsymbol{W}}(\boldsymbol{I}_n - \boldsymbol{\Phi})\boldsymbol{M} \\ \boldsymbol{\Phi} & (\boldsymbol{I}_n - \boldsymbol{\Phi})\boldsymbol{M} \end{bmatrix} = \begin{bmatrix} \boldsymbol{P}_{11} & \boldsymbol{P}_{12} \\ \boldsymbol{P}_{21} & \boldsymbol{P}_{22} \end{bmatrix} \quad (5)$$

As stated above, we set the initialisation as $\hat{\boldsymbol{y}}(0) = \boldsymbol{y}(0)$, yielding $\boldsymbol{y}(1) = (\boldsymbol{\Lambda}\boldsymbol{W} + \boldsymbol{I}_n - \boldsymbol{\Lambda})\boldsymbol{y}(0)$.

3 Analysis of the Opinion Dynamical System

We now investigate the evolution of $y_i(t)$ and $\hat{y}_i(t)$, according to Eq. (1) and Eq. (2), for the n individuals interacting on the influence network $\mathcal{G}[\boldsymbol{W}]$. In order to place the focus on social interpretations, we first present the theoretical statements, and then discuss conclusions. Most of the proofs are deferred to the Appendix, while a few proofs are presented in the extended arXiv version of this paper [43]. Simulations are also provided in [43] for illustrative purposes. The key focus of this section is to secure conclusions via analysis of Eq. (4) regarding the *discrepancies between expressed and private opinions* that form over time. Throughout this section, we make the following assumption on the social network.

Assumption 1 *The network $\mathcal{G}[\boldsymbol{W}]$ is strongly connected and aperiodic, and $\boldsymbol{W}$ is row-stochastic. Furthermore, there holds $\lambda_i, \phi_i \in (0,1), \forall\, i \in \mathcal{I}$.*

It should be noted that for the purpose of convergence analysis, almost certainly one could relax the assumption to include graphs which are not strongly connected, and for $\phi_i, \lambda_i \in [0,1]$, which we leave for future work.

Notice that because $\sum_{j=1}^n w_{ij} = 1$ and $\lambda_i \in [0,1]$, Eq. (1) indicates that $y_i(t+1)$ is a convex combination of $y_i(0)$, $y_i(t)$, and $\hat{y}_j(t), j \in \mathcal{N}_i$. Similarly, $\hat{y}_i(t)$ is a convex combination of $y_i(t)$ and $\hat{y}_{i,\mathrm{lavg}}(t-1)$. It follows that

$$\mathcal{S} = \{y_i, \hat{y}_i : \min_{k \in \mathcal{I}} y_k(0) \le y_i, \hat{y}_i \le \max_{j \in \mathcal{I}} y_j(0), i \in \mathcal{I}\} \quad (6)$$

is a positive invariant set of the system Eq. (4), which is a desirable property. If $y_i(0) \in [a, b]$, where $a, b \in \mathbb{R}$ represent the two extremes of the opinion spectrum, and $\mathcal{S}$ is a positive invariant set of Eq. (4), then the opinions are always well defined.

3.1 Convergence

The main convergence theorem, and a subsequent corollary for consensus, are now presented.

Theorem 1 (Exponential Convergence) *Consider a network $\mathcal{G}[\boldsymbol{W}]$ where each individual i's opinions $y_i(t)$ and $\hat{y}_i(t)$ evolve according to Eq. (1) and Eq. (2), respectively. Suppose Assumption 1 holds. Then, the system Eq. (4) converges exponentially fast to the limit*

$$\lim_{t \to \infty} \boldsymbol{y}(t) \triangleq \boldsymbol{y}^* = \boldsymbol{R}\boldsymbol{y}(0) \quad (7)$$

$$\lim_{t \to \infty} \hat{\boldsymbol{y}}(t) \triangleq \hat{\boldsymbol{y}}^* = \boldsymbol{S}\boldsymbol{y}^*, \quad (8)$$

where $\boldsymbol{R} = (\boldsymbol{I}_n - (\boldsymbol{P}_{11} + \boldsymbol{P}_{12}\boldsymbol{S}))^{-1}(\boldsymbol{I}_n - \boldsymbol{\Lambda})$ and $\boldsymbol{S} = (\boldsymbol{I}_n - \boldsymbol{P}_{22})^{-1}\boldsymbol{P}_{21}$ are positive and row-stochastic, with $\boldsymbol{P}_{ij}$ defined in Eq. (5).

The above shows that the final private and expressed opinions depend on $\boldsymbol{y}(0)$, while $\hat{\boldsymbol{y}}(0)$ are forgotten exponentially fast; one could initialise $\hat{\boldsymbol{y}}(0)$ arbitrarily, though the transient will differ. The row-stochasticity of $\boldsymbol{R}$ and $\boldsymbol{S}$ implies that the final private and expressed opinions are a convex combination of the initial private opinions. Additionally, $\boldsymbol{R}, \boldsymbol{S} > 0$ means every individual i's initial $y_i(0)$ has an influence on every individual j's final opinions y_j^* and $\hat{y}_j^*$, a reflection of the strongly connected network. The following corollary establishes a condition for consensus of opinions, though one notes that part of the hypothesis for Theorem 1 is discarded.

Corollary 1 (Consensus of Opinions) *Suppose that $\phi_i \in (0,1)$, and $\lambda_i = 1$, for all $i \in \mathcal{I}$. Suppose further that $\mathcal{G}[\boldsymbol{W}]$ is strongly connected and aperiodic, and $\boldsymbol{W}$ is row-stochastic. Then, for the system Eq. (4), $\lim_{t \to \infty} \boldsymbol{y}(t) = \lim_{t \to \infty} \hat{\boldsymbol{y}}(t) = \alpha \boldsymbol{1}_n$ for some $\alpha \in \mathbb{R}$, exponentially fast.*

3.2 Discrepancies and Persistent Disagreement

This section establishes how disagreement among the opinions at steady state may arise. In the following theorem, let $z_{\max} \triangleq \max_{i=1,\ldots,n} z_i$ and $z_{\min} \triangleq \min_{i=1,\ldots,n} z_i$ denote the largest and smallest element of $\boldsymbol{z} \in \mathbb{R}^n$.

Theorem 2 *Suppose that the hypotheses in Theorem 1 hold. If $\boldsymbol{y}(0) \neq \alpha \boldsymbol{1}_n$ for some $\alpha \in \mathbb{R}$, then the final opinions obey the following inequalities*

$$y(0)_{\max} > y_{\max}^* > \hat{y}_{\max}^* \quad (9a)$$

$$y(0)_{\min} < y_{\min}^* < \hat{y}_{\min}^* \quad (9b)$$

and $\hat{y}_{\min}^* \neq \hat{y}_{\max}^*$. Moreover, given a network $\mathcal{G}[\boldsymbol{W}]$ and parameter vectors $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_n]^\top$ and $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_n]^\top$, the set of initial conditions $\boldsymbol{y}(0)$ for which precisely $m > 0$ individuals $i_j \in \{i_1, \ldots, i_m\} \subseteq \mathcal{I}$ have $y_{i_j}^* = \hat{y}_{i_j}^*$, i.e. $m \triangleq |\{i \in \mathcal{I} : y_i^* = \hat{y}_i^*\}|$, lies in a subspace of $\mathbb{R}^n$ with dimension $n - m$.

This result shows that for generic initial conditions there is a persistent disagreement of final opinions at the steady-state. This is a consequence of individuals not

being maximally susceptible to influence, $\lambda_i < 1 \, \forall \, i \in \mathcal{I}$. One of the key conclusions of this paper is that for *any individual i* in the network, $y_i^* \neq \hat{y}_i^*$ for generic initial conditions, which is a subtle but significant difference from Eq. (9). More precisely, *the presence of both stubbornness and pressure to conform, and the strong connectedness of the network creates a discrepancy between the private and expressed opinions of an individual.* Without stubbornness ($\lambda_i = 1, \forall \, i$), a consensus of opinions is reached, and without a pressure to conform ($\phi_i = 1$), an individual has the same private and expressed opinions. Without strong connectedness, some individuals will not be influenced to change opinions.

One further consequence of Eq. (9) is that $y_{\max}^* - y_{\min}^* > \hat{y}_{\max}^* - \hat{y}_{\min}^*$, which implies that the *level of agreement is greater* among the final expressed opinions when compared to the final private opinions. In other words, individuals are more willing to agree with others when they are expressing their opinions in a social network due to a pressure to conform. Moreover, the extreme final expressed opinions are upper and lower bounded by the final private opinions, which are in turn upper and lower bounded by the extreme initial private opinions, showing the effects of interpersonal influence and a pressure to conform.

Remark 3 *Theorem 2 states that generically, there will be no two individuals who have the same final private opinions, and no individual will have the same final private and expressed opinion. Let the parameters defining the system ($\boldsymbol{W}$, $\boldsymbol{\phi}$ and $\boldsymbol{\lambda}$) be given and suppose that one runs p experiments with $y_i(0)$ sampled independently from a distribution (uniform, normal, beta, etc.) over a non-degenerate interval[2] . If q is the number of those experiments which result in $y_i^* = \hat{y}_i^*$ for some $i \in \mathcal{I}$, then $\lim_{p\to\infty} q/p = 0$. From yet another perspective, the set of $\boldsymbol{y}(0)$ for which $y_i^* = \hat{y}_i^*$ for some $i \in \mathcal{I}$ belongs in a subspace of $\mathbb{R}^n$ that has a Lebesgue measure of zero. Similarly, $y_i^* = y_j^*$ for $i \neq j$ generically.*

3.3 Estimating Disagreement in the Private Opinions

We now give a quantitative method for underbounding the disagreement in the steady-state private opinions for a special case of the model, where we replace the local public opinion $\hat{y}_{i,\mathrm{lavg}}$ with the global public opinion $\hat{y}_{\mathrm{avg}} = n^{-1} \sum_{j=1}^n \hat{y}_i$ in Eq. (2) for all individuals.

Corollary 2 *Suppose that, for all $i \in \mathcal{I}$, $\hat{y}_{i,lavg}(t-1)$ in Eq. (2) is replaced with $\hat{y}_{avg} = n^{-1} \sum_{j=1}^n \hat{y}_i$. Let $\kappa(\boldsymbol{\phi}) = 1 - \frac{\phi_{\min}}{\phi_{\max}}(1 - \phi_{\max}) \in (0,1)$ and $\phi_{\max} = \max_{i \in \mathcal{I}} \phi_i$,*

---

[2] A statistical distribution is degenerate if for some k_0 the cumulative distribution function $F(x, k_0) = 0$ if $x < k_0$ and $F(x, k_0) = 1$ if $x \geq k_0$.

$\phi_{\min} = \min_{i \in \mathcal{I}} \phi_i$. *Suppose further that the hypotheses in Theorem 1 hold. Then,*

$$\frac{\hat{y}_{\max}^* - \hat{y}_{\min}^*}{\kappa(\boldsymbol{\phi})} \leq y_{\max}^* - y_{\min}^*. \tag{10}$$

For the purposes of monitoring the level of unvoiced discontent in a network (e.g. to prevent drastic and unforeseen actions or violence [5,6,29]), it is of interest to obtain more knowledge about the level of disagreement among the private opinions: $y_{\max}^* - y_{\min}^*$. A fundamental issue is that such information is by definition unlikely to be obtainable (except in certain situations like the post-experimental interviews conducted by Asch in his experiments, see Section 4). On the other hand, one expects that the level of expressed disagreement $\hat{y}_{\max}^* - \hat{y}_{\min}^*$ may be available. While one cannot expect to know every ϕ_i, we argue that $\phi_{\max}$ and $\phi_{\min}$ might be obtained, if not accurately then approximately. If the global public opinion $\hat{y}_{\mathrm{avg}}$ acts on all individuals, then Corollary 2 gives a method for computing a *lower bound* on the level of private disagreement given some limited knowledge.

It is obvious that if $\kappa(\boldsymbol{\phi})$ is small (if $\phi_{\max}$ is small and the ratio $\phi_{\min}/\phi_{\max}$ is close to 1), then even strong agreement among the expressed opinions (a small $\hat{y}_{\max}^* - \hat{y}_{\min}^*$) does not preclude significant disagreement in the final private opinions of the individuals. This might occur in e.g., an authoritarian government. The tightness of the bound Eq. (10) depends on the ratio $\phi_{\min}/\phi_{\max}$; the closer the ratio is to one (i.e. as the "force" of the pressure to conform felt by each individual becomes more uniform), the tighter the bound.

3.4 An Individual's Resilience Affects Everyone

An interesting result is now presented, that shows how individual i's resilience ϕ_i is propagated through the network.

Corollary 3 *Suppose that the hypotheses in Theorem 1 hold. Then, the matrix $\boldsymbol{S}$ in Eq. (8) has partial derivative $\frac{\partial(\boldsymbol{S})}{\partial \phi_i}$ with strictly positive entries in the ith column and with all other entries strictly negative.*

Recall below Theorem 1 that individual k's final expressed opinion $\hat{y}_k^*$ is a convex combination of all individuals' final private opinions y_j^*, with convex weights s_{kj}, $j = 1, \ldots, n$. Intuitively, increasing ϕ_k makes individual k more resilient to the pressure to conform, and this is confirmed by the above; $\frac{\partial s_{kk}}{\partial \phi_k} > 0$ and $\frac{\partial s_{kj}}{\partial \phi_k} < 0$ for any $j \neq k$ and thus $\hat{y}_k^* \to y_k^*$ as $\phi_k \to 1$.

More importantly, the above result yields a surprising and nontrivial fact; *every entry* of the kth column of $\frac{\partial(\boldsymbol{S})}{\partial \phi_k}$ is strictly positive, and all other entries of $\frac{\partial(\boldsymbol{S})}{\partial \phi_k}$ are

strictly negative. In context, any change in individual k's resilience directly impacts every other individual's final expressed opinion due to the network of interpersonal influences. In particular, as ϕ_k increases (decreases), an individual j's final expressed opinion $\hat{y}_j^*$ becomes closer to (further from) the final private opinion y_k^* of individual k, since $\frac{\partial s_{jk}}{\partial \phi_k} > 0$ (decreasing, since $\frac{\partial s_{jk}}{\partial \phi_k} < 0$).

4 Application to Asch's Experiments

We now use the model to revisit Solomon E. Asch's seminal experiments on conformity [13]. There are at least two objectives. For one, successfully capturing Asch's empirical data constitutes a form of soft validation for the model. Second, we aim to identify the values of the individual's susceptibility λ_i and resilience ϕ_i that determine the individual's reaction to a unanimous majority's pressure to conform, and thus give an agent-based model explanation of the recorded observations. In order for the reader to fully appreciate and understand the results, a brief overview of the experiments and its results are now given, and the reader is referred to [13] for full details on the results. In summary, the experiments studied an individual's response to "two contradictory and irreconcilable forces" [13] of (i) a clear and indisputable fact, and (ii) a unanimous majority of the others who take positions opposing this fact.

In the experiment, eight individuals are instructed to judge a series of line lengths. Of the eight individuals, one is in fact the test subject, and the other seven "confederates" [3] have been told a priori about what they should do. An example of the line length judging experiment is shown in Fig. 2. There are three lines of unequal length, and the group has open discussions concerning which one of the lines A, B, C is equal in length to the green line. Each individual is required to independently declare his choice, and the confederates (blue individuals) unanimously select the same wrong answer, e.g. B. The reactions of the test individual (red node) are then recorded, followed by a post-experiment interview to evaluate the test individual's private belief [4].

In order to apply our model, and with Fig. 2 as an illustrative example, we frame $y_i, \hat{y}_i \in [0,1]$ to be individual i's belief in the statement "the green line is of the same length as line A." Specifically, $y_i = 1$ (respectively $y_i = 0$) implies individual i is maximally certain the statement is true (respectively, maximally certain the statement is false). Asch found close to 100% of individuals in control groups had $y_i(0) = 1$. Without loss of

Fig. 2. Example of the Asch experiment. The individuals openly discuss their individual beliefs as to which one of A, B, C has the same length as the green line. Clearly A is equal in length to the green (left most) line. The test individual is the red (centre top) node. The confederates (seven blue nodes) unanimously express belief in the same wrong answer, e.g. B. Reprinted from [42] with permission by **SpringerNature**

Table 1
Types of test individuals and their susceptibility and resilience parameters

	λ_1	ϕ_1
Independent	low	high
Yielding, judgment distortion	high	any
Yielding, action distortion	low	low

generality, we therefore denote the test individual as individual 1 and set $y_1(0) = \hat{y}(0) = 1$. Confederates are set to have $y_i(0) = \hat{y}_i(0) = 0$, for $i = 2, \ldots, n$, with $\lambda_i = 0$ and $\phi_i = 1$. That is, they consistently express maximal certainty that "the green line of the same length as line A" is a **false** statement.

Asch reported that test individuals could be split into three types, based on their different reactions. Some test individuals were *independent* and expressed strong certainty even in the face of a unanimous majority; such individuals at the end of discussion should have final beliefs $\hat{y}_1^*, y_1^* \approx 1$. Others *yielded* to the confederates, and could be split into two further types. Some showed *distortion of judgment*, where eventually, individual 1 expressed and held a private belief in line with the confederate majority, represented by $y_1^*, \hat{y}_1^* \approx 0$. Others showed *distortion of action* whereby the test individual eventually expressed a belief similar to the confederate majority, but in the post-experiment interview maintained private belief in the correct answer; in our model this is represented by $y_1^* \approx 1$ and $\hat{y}_1^* \approx 0$. Based on Asch's qualitative descriptions of these 3 types of individuals [13], we assigned values for the parameters λ_1 and ϕ_1, summarised in Table 1. The arXiv version [43] contains additional explanatory details of the Asch experiments, including descriptions and quotes. It also provides theoretical calculations showing the functional dependence of the final opinions of test individual 1, $y_1^*, \hat{y}_1^*$, on the parameters $\boldsymbol{W}, \phi_1, \lambda_1$ and n.

---

[3] These other individuals have become referred to as "confederates" in later literature.

[4] In this section, we refer to $y_i, \hat{y}_i$ as beliefs, as the variables represent individual i's certainty on an issue that is provably true or false. As noted in Section 2, our model is general enough to cover both subjective and intellective topics.

The Asch experiments are simulated using the proposed model. An arbitrary $\boldsymbol{W}$ is generated with weights w_{ij} sampled randomly from a uniform distribution and normalised to ensure $\sum_{j=1}^{n} w_{ij} = 1$. The other parameters are described in the third paragraph of Section 4. In the following plots of Fig. 3a, 3b and 3c, the values of λ_1 and ϕ_1 are given. The red lines correspond to test individual 1, with the solid line showing private belief $y_1(t)$ and the dotted line showing expressed belief $\hat{y}_1(t)$. The blue line represents the confederates $k = 2, \ldots, 8$, who have $y_k(t) = \hat{y}_k(t) = 0$ for all t.

From Fig. 3a, it can be seen that both the private and expressed beliefs of an *independent* test individual are largely unaffected by the confederates' unanimous expressed belief and the pressure exerted. Note that $\hat{y}_1^* < y_1^*$; [13] reported that one independent test individual stated "You're *probably* right, but you may be wrong!", conceding slightly to the majority belief. Figure 3b shows a yielding test individual who exhibits *distortion of judgment*. Because of high susceptibility λ_1, y_1^* is heavily influenced by the unanimous majority; individual i is no longer privately certain that A is the correct answer. In contrast, a yielding test individual exhibiting *distortion of action* is shown in Fig. 3c. The observed belief evolution accurately reflect Asch's observations: individual 1 "yields because of an overmastering need to not appear different or inferior to others" [13], giving $\hat{y}_1^* \approx 0.1$. However, individual 1 is still able to "conclude that they [themselves] are not wrong" [13], i.e. $y_i^* \approx 0.93$.

Other simulations with values of λ_1, ϕ_1 in the neighbourhood of those used also display similar behaviour as in Fig. 3a to 3b, indicating a robust ability for our model to capture Asch's experiments is an intrinsic property of the model, and rather than resulting from careful reverse engineering. All three types of individual behaviours can be predicted by our model using pairs of parameters λ_i, ϕ_i, providing a measure of validation for our model. We have provided an agent-based model explanation of the empirical findings of Asch's experiments; it might now be possible to analyse the many subsequent works derived from Asch in a common framework, whereas existing static models of conformity are tied to specific empirical data (see Section 1.1). The Friedkin–Johnsen model has also been applied to the Asch experiments [30], but (unsurprisingly) was not able to capture all of the types of individuals reported because the Friedkin–Johnsen model does not assume that each individual has a separate private and expressed belief.

4.2 Threshold Variant and Asch's Second Experiments

The simulations above assumed that the individuals express a continuous real-valued opinion $\hat{y}_i(t)$, whereas it is perhaps more appropriate to set $\hat{y}_i(t)$ as a binary variable, with $\hat{y}_i(t) = 1$ and $\hat{y}_i(t) = 0$ denoting individual i picking A and not picking A as the correct answer. The proposed model can be modified to accommodate situations where the expressed variable denotes an action, or decision by replacing Eq. (2) with

$$\hat{y}_i(t) = \sigma_i \left(\phi_i y_i(t) + (1 - \phi_i)\hat{y}_{i,\text{lavg}}(t - 1) \right), \qquad (11)$$

where $\sigma_i(x) : [0, 1] \to \{0, 1\}$ is a threshold function satisfying $\sigma_i(x) = 0$ if $x \in [0, \tau_i]$ and $\sigma_i(x) = 1$ if $x \in (\tau_i, 1]$, for some threshold value $\tau_i \in (0, 1)$. Applying the threshold variant of the model with $\tau_i = 0.5$ yields no qualitative difference for the simulations in Section 4.1.

Asch conducted several variations to the original experiments, as reported in [13]. In one particular variation, one confederate also selected the correct answer; the frequency of individuals showing distortion of action or distortion of judgment decreased dramatically. Preliminary investigations to compare the original model Eq. (2) with the threshold model Eq. (11) were conducted using simulations to capture the aforementioned variation of the Asch experiments. Details of the simulations and figures can be found in the ArXiv version [43]; we provide only a summary here due to spatial constraints. It appears that the threshold model is better able to capture the outcomes of the experiments, especially the variation of Asch involving a truth-telling confederate, when compared to the original model. Moreover, when using the threshold model it seems that there is a range of λ_1 and ϕ_1 values that in Asch's First Experiment (Section 4.1) resulted in individual 1 supporting the confederates (distortion of action or judgment), but in Asch's Second Experiments setup resulted in individual 1 picking the correct answer.

The threshold model variant appears to be better suited for certain more complicated scenarios, but carries with it the following cautionary remark. The threshold variant produces behaviour that is *not inconsistent* with the observations of Asch's second experiments, but further detailed study of the threshold model must be completed before concrete conclusions can be drawn. The model's behaviour becomes more difficult to characterise due to the nonlinear threshold function $\sigma(x)$, but promises richer and more varied phenomena.

5 Conclusions

We have proposed a novel agent-based model of opinion evolution on interpersonal influence networks, where each individual has separate expressed and private opinions that evolve in a coupled manner. Conditions on the network and the values of susceptibility and resilience for the individuals were established for ensuring that the opinions converged exponentially fast to a steady-state
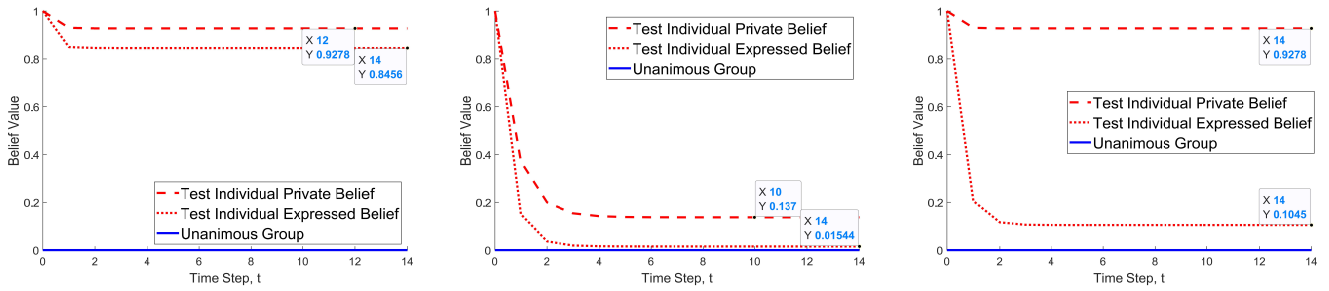
(a) An *independent* individual, with $\lambda_1 = 0.1, \phi_1 = 0.9$.

(b) A *yielding* individual with *distortion of judgment*, with $\lambda_1 = 0.9, \phi_1 = 0.1$.

(c) A *yielding* individual with *distortion of action*, with $\lambda_1 = 0.1, \phi_1 = 0.1$.

Fig. 3. Fig. 3a, 3c, and 3b show the evolution of beliefs for all three types of reactions recorded by Asch, as they appear in our model. The dashed and dotted red lines denote the private and expressed belief, respectively, of the test individual 1 (i.e. $y_1(t)$ and $\hat{y}_1(t)$). The solid blue line is the belief of the unanimous confederate group, who express a belief of $\hat{y}_i(t) = 0$. All three figures reprinted from [42] with permission by **SpringerNature**.

of persistent disagreement. Further analysis of the final opinion values yielded semi-quantitative conclusions that led to insightful social interpretations, including the conditions that lead to a discrepancy between the expressed and private opinions of an individual. We then used the model to study Asch's experiments [13], showing that all 3 types of reactions from the test individual could be captured within our framework. A number of interesting future directions can be considered. Preliminary simulations show that our model can also capture *pluralistic ignorance*, with network structure and placement of extremist nodes having a significant effect on the observed phenomena. Clearly the threshold model in Section 4.2 requires further study, and one could also consider the model in a continuous-time setting, or with asynchronous updating, or both.

Acknowledgements

References

[1] A. V. Proskurnikov and R. Tempo, "A tutorial on modeling and analysis of dynamic social networks. Part I," *Annual Reviews in Control*, vol. 43, pp. 65–79, 2017.

[2] ——, "A tutorial on modeling and analysis of dynamic social networks. Part II," *Annual Reviews in Control*, vol. 45, pp. 166–190, 2018.

[3] A. Flache, M. Mäs, T. Feliciani, E. Chattoe-Brown, G. Deffuant, S. Huet, and J. Lorenz, "Models of Social Influence: Towards the Next Frontiers," *Journal of Artificial Societies & Social Simulation*, vol. 20, no. 4, 2017.

[4] N. L. Waters and V. P. Hans, "A Jury of One: Opinion Formation, Conformity, and Dissent on Juries," *Journal of Empirical Legal Studies*, vol. 6, no. 3, pp. 513–540, 2009.

[5] J. Goodwin, "Why We Were Surprised (Again) by the Arab Spring," *Swiss Political Science Review*, vol. 17, no. 4, pp. 452–456, 2011.

[6] T. Kuran, "Sparks and prairie fires: A theory of unanticipated political revolution," *Public Choice*, vol. 61, no. 1, pp. 41–74, 1989.

[7] S. Bikhchandani, D. Hirshleifer, and I. Welch, "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, vol. 100, no. 5, pp. 992–1026, 1992.

[8] F. H. Allport, *Social Psychology*. Boston: Houghton Mifflin Company, 1924.

[9] E. Noelle-Neumann, *The Spiral of Silence: Public Opinion, Our Social Skin*. University of Chicago Press, 1993.

[10] D. G. Taylor, "Pluralistic Ignorance and the Spiral of Silence: A Formal Analysis," *Public Opinion Quarterly*, vol. 46, no. 3, pp. 311–335, 1982.

[11] D. Centola, R. Willer, and M. Macy, "The Emperor's Dilemma: A Computational Model of Self-Enforcing Norms," *American Journal of Sociology*, vol. 110, no. 4, pp. 1009–1040, 2005.

[12] R. Willer, K. Kuwabara, and M. W. Macy, "The False Enforcement of Unpopular Norms," *American Journal of Sociology*, vol. 115, no. 2, pp. 451–490, 2009.

[13] S. E. Asch, *Groups, Leadership, and Men*. Carnegie Press: Pittsburgh, 1951, ch. Effects of Group Pressure Upon the Modification and Distortion of Judgments, pp. 222–236.

[14] R. Bond, "Group Size and Conformity," *Group Processes & Intergroup Relations*, vol. 8, no. 4, pp. 331–354, 2005.

[15] D. A. Prentice and D. T. Miller, "Pluralistic Ignorance and Alcohol Use on Campus: Some Consequences of Misperceiving the Social Norm," *Journal of personality and social psychology*, vol. 64, no. 2, pp. 243–256, 1993.

[16] H. J. O'Gorman, "Pluralistic Ignorance and White Estimates of White Support for Racial Segregation," *Public Opinion Quarterly*, vol. 39, no. 3, pp. 313–330, 1975.

[17] S. Tanford and S. Penrod, "Social Influence Model: A Formal Integration of Research on Majority and Minority Influence Processes," *Psychological Bulletin*, vol. 95, no. 2, p. 189, 1984.

[18] B. Mullen, "Operationalizing the Effect of the Group on the Individual: A Self-Attention Perspective," *Journal of Experimental Social Psychology*, vol. 19, no. 4, pp. 295–322, 1983.

[19] G. Stasser and J. H. Davis, "Group decision making and social influence: A social interaction sequence model." *Psychological Review*, vol. 88, no. 6, p. 523, 1981.

[20] J. R. P. French Jr, "A Formal Theory of Social Power," *Psychological Review*, vol. 63, no. 3, pp. 181–194, 1956.

[21] M. H. DeGroot, "Reaching a Consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[22] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence models, analysis, and simulation," *Journal of Artificial Societies and Social Simulation*, vol. 5, no. 3, 2002.

[23] W. Su, G. Chen, and Y. Hong, "Noise leads to quasi-consensus of hegselmann–krause opinion dynamics," *Automatica*, vol. 85, pp. 448 – 454, 2017.

[24] P. Dandekar, A. Goel, and D. T. Lee, "Biased assimilation, homophily, and the dynamics of polarization," *Proceedings of the National Academy of Sciences*, vol. 110, no. 15, pp. 5791–5796, 2013.

[25] M. Mäs, A. Flache, and J. A. Kitts, "Cultural Integration and Differentiation in Groups and Organizations," in *Perspectives on Culture and Agent-based Simulations*. Springer, 2014, pp. 71–90.

[26] C. Altafini, "Consensus Problems on Networks with Antagonistic Interactions," *IEEE Transactions on Automatic Control*, vol. 58, no. 4, pp. 935–946, 2013.

[27] A. Proskurnikov, A. Matveev, and M. Cao, "Opinion dynamics in social networks with hostile camps: Consensus vs. polarization," *IEEE Transaction on Automatic Control*, vol. 61, no. 6, pp. 1524–1536, 2016.

[28] N. E. Friedkin and E. C. Johnsen, "Social Influence and Opinions," *Journal of Mathematical Sociology*, vol. 15, no. 3-4, pp. 193–206, 1990.

[29] P. Duggins, "A Psychologically-Motivated Model of Opinion Change with Applications to American Politics," *Journal of Artificial Societies and Social Simulation*, vol. 20, no. 1, pp. 1–13, 2017.

[30] N. E. Friedkin and E. C. Johnsen, *Social Influence Network Theory: A Sociological Examination of Small Group Dynamics*. Cambridge University Press, 2011, vol. 33.

[31] N. E. Friedkin and F. Bullo, "How truth wins in opinion dynamics along issue sequences," *Proceedings of the National Academy of Sciences*, vol. 114, no. 43, pp. 11 380–11 385, 2017.

[32] N. E. Friedkin, P. Jia, and F. Bullo, "A Theory of the Evolution of Social Power: Natural Trajectories of Interpersonal Influence Systems along Issue Sequences," *Sociological Science*, vol. 3, pp. 444–472, 2016.

[33] C. D. Godsil and G. Royle, *Algebraic Graph Theory*. Springer: New York, 2001, vol. 207.

[34] F. Bullo, J. Cortes, and S. Martinez, *Distributed Control of Robotic Networks*. Princeton University Press, 2009.

[35] T. Kuran, *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Harvard University Press, 1997.

[36] P. Jia, A. MirTabatabaei, N. E. Friedkin, and F. Bullo, "Opinion Dynamics and the Evolution of Social Power in Influence Networks," *SIAM Review*, vol. 57, no. 3, pp. 367–397, 2015.

[37] M. Ye, J. Liu, B. D. O. Anderson, C. Yu, and T. Başar, "Evolution of Social Power in Social Networks with Dynamic Topology," *IEEE Transaction on Automatic Control*, vol. 63, no. 11, pp. 3793–3808, Nov. 2019.

[38] R. L. Gorden, "Interaction Between Attitude and the Definition of the Situation in the Expression of Opinion," *American Sociological Review*, vol. 17, no. 1, pp. 50–58, 1952.

[39] F. Merei, "Group Leadership and Institutionalization," *Human Relations*, vol. 2, no. 1, pp. 23–39, 1949.

[40] L. Festinger, "Informal Social Communication," *Psychological Review*, vol. 57, no. 5, p. 271, 1950.

[41] S. Schachter, "Deviation, Rejection, and Communication," *The Journal of Abnormal and Social Psychology*, vol. 46, no. 2, pp. 190–207, 1951.

[42] M. Ye, *Opinion Dynamics and the Evolution of Social Power in Social Networks*. Springer, Cham, 2019.

[43] M. Ye, Y. Qin, A. Govaert, B. D. O. Anderson, and M. Cao, "An Influence Network Model to Study Discrepancies in Expressed and Private Opinions," 2018. [Online]. Available: https://arxiv.org/abs/1806.11236

[44] R. S. Varga, *Matrix Iterative Analysis*. Springer Science & Business Media, 2009, vol. 27.

[45] E. Seneta, *Non-negative Matrices and Markov Chains*. Springer Science & Business Media, 2006.

[46] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton University Press, 2009.

[47] W. J. Rugh, *Linear System Theory*, 2nd ed. Prentice Hall, Upper Saddle River, New Jersey, 1996.

A Preliminaries

In this section, we record some definitions, and notations to be used in the proofs of the main results. A square matrix $\boldsymbol{A} \geq 0$ is primitive if there exists $k \in \mathbb{N}$ such that $\boldsymbol{A}^k > 0$ [34, Definition 1.12]. A graph $\mathcal{G}[\boldsymbol{A}]$ is strongly connected and aperiodic if and only if $\boldsymbol{A}$ is primitive, i.e. $\exists k \in \mathbb{N}$ such that $\boldsymbol{A}^k$ is a positive matrix [34, Proposition 1.35]. We denote the ith canonical base unit vector of $\mathbb{R}^n$ as $\mathbf{e}_i$. The spectral radius of a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is given by $\rho(\boldsymbol{A})$.

Lemma 1 *If $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is row-substochastic and irreducible, then $\rho(\boldsymbol{A}) < 1$.*

Proof: This lemma is an immediate consequence of [44, Lemma 2.8]. □

A.1 Performance Function and Ergodicity Coefficient

In order to analyse the disagreement among the opinions at steady state, we introduce a performance function and a coefficient of ergodicity. For a vector $\boldsymbol{x} \in \mathbb{R}^n$, define the performance function $V(\boldsymbol{x}) : \mathbb{R}^n \mapsto \mathbb{R}$ as

$$V(\boldsymbol{x}) = \max_{i \in \{1,\dots,n\}} x_i - \min_{j \in \{1,\dots,n\}} x_j, \qquad (\text{A.1})$$

In context, $V(\boldsymbol{y})$ measures the "level of disagreement" in the vector of opinions $\boldsymbol{y}(t)$, and consensus of opinions, i.e. $\boldsymbol{y}(t) = \alpha \mathbf{1}_n, \alpha \in \mathbb{R}$, is reached if and only if $V(\boldsymbol{y}(t)) = 0$. Next consider the following coefficient of ergodicity,

$\tau(\boldsymbol{A})$ for a row-stochastic matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, defined [45] as

$$\tau(\boldsymbol{A}) = 1 - \min_{i,j \in \{1,\dots,n\}} \sum_{s=1}^{n} \min\{a_{is}, a_{js}\}. \qquad \text{(A.2)}$$

This coefficient of ergodicity satisfies $0 \leq \tau(\boldsymbol{A}) \leq 1$, and $\tau(\boldsymbol{A}) = 0$ if and only if $\boldsymbol{A} = \mathbf{1}_n \boldsymbol{z}^\top$ for some $\boldsymbol{z} \geq 0$. Importantly, there holds $\tau(\boldsymbol{A}) < 1$ if $\boldsymbol{A} > 0$. Also, there holds $V(\boldsymbol{A}\boldsymbol{x}) \leq \tau(\boldsymbol{A})V(\boldsymbol{x})$ (see [45])

A.2 Supporting Lemmas

Two lemmas are introduced to establish several properties of $\boldsymbol{P}$ and $(\boldsymbol{I}_{2n} - \boldsymbol{P})^{-1}$, which will be used to help prove the main results.

Lemma 2 *Suppose that Assumption 1 holds. Then, $\boldsymbol{P}$ given in Eq. (5) is nonnegative, the graph $\mathcal{G}[\boldsymbol{P}]$ is strongly connected and aperiodic, and there holds $\rho(\boldsymbol{P}) < 1$.*

Lemma 3 *Suppose that Assumption 1 holds. With $\boldsymbol{P}$ given in Eq. (5), define $\boldsymbol{Q}$ as*

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_n - \boldsymbol{P}_{11} & -\boldsymbol{P}_{12} \\ -\boldsymbol{P}_{21} & \boldsymbol{I}_n - \boldsymbol{P}_{22} \end{bmatrix}.$$

Then, $\boldsymbol{Q}_{11}, \boldsymbol{Q}_{22}$ are nonsingular, and $\boldsymbol{Q}^{-1} > 0$ is

$$\boldsymbol{Q}^{-1} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}, \qquad \text{(A.3)}$$

where $\boldsymbol{A} = (\boldsymbol{Q}_{11} - \boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21})^{-1}$, $\boldsymbol{D} = (\boldsymbol{Q}_{22} - \boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12})^{-1}$, $\boldsymbol{B} = -\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{D}$, $\boldsymbol{C} = -\boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21}\boldsymbol{A}$. Moreover, $\boldsymbol{R} = \boldsymbol{A}(\boldsymbol{I}_n - \boldsymbol{\Lambda})$ and $\boldsymbol{S} = -\boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21}$ are invertible, positive row-stochastic matrices.

B Proofs

B.1 Proof of Lemma 2

First, we verify that $\boldsymbol{P} \geq 0$ by using the fact that $\boldsymbol{W}$, $\boldsymbol{\Lambda}$, $\boldsymbol{I}_n - \boldsymbol{\Phi}$, $\boldsymbol{M}$ are all nonnegative. Next, observe that

$$\begin{bmatrix} \boldsymbol{\Lambda}(\widetilde{\boldsymbol{W}} + \widehat{\boldsymbol{W}}\boldsymbol{\Phi}) & \boldsymbol{\Lambda}\widehat{\boldsymbol{W}}(\boldsymbol{I}_n - \boldsymbol{\Phi})\boldsymbol{M} \\ \boldsymbol{\Phi} & (\boldsymbol{I}_n - \boldsymbol{\Phi})\boldsymbol{M} \end{bmatrix} \begin{bmatrix} \mathbf{1}_n \\ \mathbf{1}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}\mathbf{1}_n \\ \mathbf{1}_n \end{bmatrix}$$

because $\boldsymbol{M}$ and $\boldsymbol{W} = \widetilde{\boldsymbol{W}} + \widehat{\boldsymbol{W}}$ are row-stochastic. In the Appendix of the extended paper [43], we prove that $\mathcal{G}[\boldsymbol{P}]$ is strongly connected and aperiodic, which implies that $\boldsymbol{P}$ is irreducible. Since $\lambda_i < 1 \, \forall i$, $\boldsymbol{P}$ is row-substochastic, and Lemma 1 establishes that $\rho(\boldsymbol{P}) < 1$. $\qquad \square$

B.2 Proof of Lemma 3

Lemma 2 showed that $\mathcal{G}[\boldsymbol{P}]$ is strongly connected and aperiodic, which implies that $\boldsymbol{P}$ is primitive. Since $\boldsymbol{Q}^{-1} = (\boldsymbol{I}_{2n} - \boldsymbol{P})^{-1}$ ans $\rho(\boldsymbol{P}) < 1$, the Neumann series yields $\boldsymbol{Q}^{-1} = \sum_{k=0}^{\infty} \boldsymbol{P}^k > 0$. Next, it will be shown $\boldsymbol{Q}_{11}$, $\boldsymbol{Q}_{22}$ and $\boldsymbol{D} = \boldsymbol{Q}_{11} - \boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21}$ are all invertible, which will allow $\boldsymbol{Q}^{-1}$ to be expressed in the form of Eq. (A.3) by use of [46, Proposition 2.8.7, pg. 108–109]. Under Assumption 1, $\mathcal{G}_1[\boldsymbol{P}_{11}]$ and $\mathcal{G}_2[\boldsymbol{P}_{22}]$ are both strongly connected and aperiodic; Lemma 1 states that $\rho(\boldsymbol{P}_{11}), \rho(\boldsymbol{P}_{22}) < 1$. Since $\boldsymbol{Q}_{11} = \boldsymbol{I}_n - \boldsymbol{P}_{11}$ and $\boldsymbol{Q}_{22} = \boldsymbol{I}_n - \boldsymbol{P}_{22}$, the same method as above can be used to prove that $\boldsymbol{Q}_{11}, \boldsymbol{Q}_{22}$ are invertible, and satisfy $\boldsymbol{Q}_{11}^{-1}, \boldsymbol{Q}_{22}^{-1} > 0$.

In order to prove that $\boldsymbol{D}$ is invertible, we first establish some properties of $\boldsymbol{S} = -\boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21}$. Since $\boldsymbol{Q}_{22}^{-1} > 0$, it follows from the fact that $\boldsymbol{\Phi} = \mathrm{diag}(\phi_i)$ is a positive diagonal matrix, that $\boldsymbol{S} = \boldsymbol{Q}_{22}^{-1}\boldsymbol{\Phi} > 0$. To prove that $\boldsymbol{S}$ is row-stochastic, first note that $\det(\boldsymbol{S}) = \det(\boldsymbol{Q}_{22}^{-1})\det(\boldsymbol{\Phi}) \neq 0$ (we have $\phi_i \in (0,1), \forall i \Rightarrow \det(\boldsymbol{\Phi}) \neq 0$). Since $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$, observe that

$$\boldsymbol{S} = \left(\boldsymbol{\Phi}^{-1} - \boldsymbol{\Phi}^{-1}(\boldsymbol{I}_n - \boldsymbol{\Phi})\boldsymbol{M}\right)^{-1}. \qquad \text{(B.1)}$$

From Eq. (B.1), verify that $\boldsymbol{S}^{-1}\mathbf{1}_n = \mathbf{1}_n$, which implies $\boldsymbol{S}\boldsymbol{S}^{-1}\mathbf{1}_n = \boldsymbol{S}\mathbf{1}_n \Leftrightarrow \boldsymbol{S}\mathbf{1}_n = \mathbf{1}_n$, i.e. $\boldsymbol{S}$ is row-stochastic.

We now turn to proving that $\boldsymbol{D}$ is invertible. Notice that $\boldsymbol{S}, -\boldsymbol{Q}_{12} = \boldsymbol{P}_{12}$, and $\boldsymbol{\Lambda}(\widehat{\boldsymbol{W}} + \widehat{\boldsymbol{W}}\boldsymbol{\Phi})$ are all nonnegative. We write $\boldsymbol{D} = \boldsymbol{I}_n - \boldsymbol{U}$ where $\boldsymbol{U} = \boldsymbol{P}_{11} + \boldsymbol{P}_{12}\boldsymbol{S} \geq 0$. Observe that $\boldsymbol{U}\mathbf{1}_n = \boldsymbol{P}_{11}\mathbf{1}_n + \left(\boldsymbol{\Lambda}\widehat{\boldsymbol{W}}(\boldsymbol{I}_n - \boldsymbol{\Phi})\right)\mathbf{1}_n = \boldsymbol{\Lambda}\mathbf{1}_n$ because $(\widehat{\boldsymbol{W}} + \widetilde{\boldsymbol{W}})\mathbf{1}_n = \mathbf{1}_n$. In other words, the ith row of $\boldsymbol{U}$ sums to $\lambda_i < 1$ (see Assumption 1), which implies that $\|\boldsymbol{U}\|_\infty < 1 \Rightarrow \rho(\boldsymbol{U}) < 1$. Because it was shown in the proof of Lemma 2 that $\mathcal{G}[\boldsymbol{P}_{11}]$ is strongly connected and aperiodic, it is straightforward to show that $\mathcal{G}[\boldsymbol{U}]$ is also strongly connected and aperiodic. It follows that $\boldsymbol{U}$ is primitive, which implies that $\boldsymbol{D}^{-1} > 0$ from the Neumann series $\boldsymbol{D}^{-1} = \sum_{k=0}^{\infty} \boldsymbol{U}^k$. Thus, $\boldsymbol{R} = \boldsymbol{D}^{-1}(\boldsymbol{I}_n - \boldsymbol{\Lambda}) > 0$, because $\boldsymbol{I}_n - \boldsymbol{\Lambda}$ is a positive diagonal matrix. Finally, one can verify that $\boldsymbol{R}$ is row-stochastic with the following computation: $\boldsymbol{D}\mathbf{1}_n = (\boldsymbol{I}_n - \boldsymbol{U})\mathbf{1}_n = (\boldsymbol{I}_n - \boldsymbol{\Lambda})\mathbf{1}_n \Rightarrow \boldsymbol{R}\mathbf{1}_n = \boldsymbol{D}^{-1}(\boldsymbol{I}_n - \boldsymbol{\Lambda})\mathbf{1}_n = \boldsymbol{D}^{-1}\boldsymbol{D}\mathbf{1}_n = \mathbf{1}_n$. This completes the proof. $\qquad \square$

B.3 Proof of Theorem 1 and Corollary 1

We provide a sketch here, and refer the reader to the arXiv paper for details [43].

Proof of Theorem 1: Lemma 2 established that the time-invariant matrix $\boldsymbol{P}$ satisfies $\rho(\boldsymbol{P}) < 1$. Standard linear systems theory [47] is used to conclude that the linear, time-invariant system Eq. (4), with constant input

$\left[((\boldsymbol{I}_n - \boldsymbol{\Lambda})\boldsymbol{y}(0))^\top, \ \boldsymbol{0}_n^\top\right]^\top$, converges exponentially fast to the steady state given in Eq. (7) and Eq. (8). $\qquad \square$

Proof of Corollary 1: The assumption that $\boldsymbol{\Lambda} = \boldsymbol{I}_n$ implies that $\boldsymbol{P}$ is nonnegative and row-stochastic. The proof of Lemma 2 established that $\mathcal{G}[\boldsymbol{P}]$ is strongly connected and aperiodic, and this remains unchanged when $\boldsymbol{\Lambda} = \boldsymbol{I}_n$. Standard results on the DeGroot model [1] then imply that consensus is achieved exponentially fast, i.e. $\lim_{t\to\infty} \boldsymbol{y}(t) = \hat{\boldsymbol{y}}(t) = \alpha \boldsymbol{1}_n$ for some $\alpha \in \mathbb{R}$. $\qquad \square$

B.4 Proof of Theorem 2

If $\boldsymbol{y}(0) = \alpha \boldsymbol{1}_n$, for some $\alpha \in \mathbb{R}$ (i.e. the initial private opinions are at a consensus), then $\boldsymbol{y}^* = \hat{\boldsymbol{y}}^* = \alpha \boldsymbol{1}_n$ because $\boldsymbol{R}$ and $\boldsymbol{S}$ are row-stochastic. In what follows, it will be proved that if the initial private opinions are not at a consensus, then there is disagreement at steady state.

First, we establish $y^*_{\min} \neq y^*_{\max}$. Note that $V(\boldsymbol{y}^*) = 0$ if and only if $\boldsymbol{y}^* = \beta \boldsymbol{1}_n$, for some $\beta \in \mathbb{R}$. Next, observe that $\boldsymbol{y}^* = \beta \boldsymbol{1}_n$ if and only if $\boldsymbol{R}\boldsymbol{y}(0) = \beta \boldsymbol{1}_n$, for some $\beta \in \mathbb{R}$. Note that $\boldsymbol{R}$ is invertible, because it is the product of two invertible matrices (see Lemma 3). Moreover, because $\boldsymbol{R}$ is row-stochastic, there holds $\boldsymbol{R}\boldsymbol{1}_n = \boldsymbol{1}_n \Leftrightarrow \boldsymbol{R}^{-1}\boldsymbol{R}\boldsymbol{1}_n = \boldsymbol{R}^{-1}\boldsymbol{1}_n \Leftrightarrow \boldsymbol{R}^{-1}\boldsymbol{1}_n = \boldsymbol{1}_n$. Thus, premultiplying by $\boldsymbol{R}^{-1}$ on both sides of $\boldsymbol{R}\boldsymbol{y}(0) = \beta \boldsymbol{1}_n$ yields $\boldsymbol{y}(0) = \beta \boldsymbol{R}^{-1}\boldsymbol{1}_n = \beta \boldsymbol{1}_n$. In other words, a consensus of the final private opinions, $\boldsymbol{y}^* = \beta \boldsymbol{1}_n$, occurs if and only if the initial private opinions are at a consensus. Recalling the theorem hypothesis that $\boldsymbol{y}(0) \neq \alpha \boldsymbol{1}_n$, for some $\alpha \in \mathbb{R}$, it follows that $\boldsymbol{y}^*$ is not at a consensus. Thus, $y^*_{\min} \neq y^*_{\max}$ as claimed.

Next, the inequalities Eq. (9a) and Eq. (9b) are proved. Since $\boldsymbol{R}, \boldsymbol{S} > 0$ are row-stochastic, $\tau(\boldsymbol{R}), \tau(\boldsymbol{S}) < 1$. Because $\boldsymbol{R}$ is invertible, $\boldsymbol{R} \neq \boldsymbol{1}_n \boldsymbol{z}^\top$ for some $\boldsymbol{z} \in \mathbb{R}^n$. This means that $\tau(\boldsymbol{R}) > 0$ (see below Eq. (A.2)). Similarly, one can prove that $\tau(\boldsymbol{S}) > 0$. In the above paragraph, it was shown that if there is no consensus of the initial private opinions, then $V(\boldsymbol{y}^* = \boldsymbol{R}\boldsymbol{y}(0)) > 0$. By recalling that $V(\boldsymbol{A}\boldsymbol{x}) \leq \tau(\boldsymbol{A})V(\boldsymbol{x})$ (see Appendix A.1) and the above facts, we conclude that $0 < V(\boldsymbol{y}^* = \boldsymbol{R}\boldsymbol{y}(0)) < V(\boldsymbol{y}(0))$, which establishes the left hand inequality of Eq. (9a) and Eq. (9b). Following steps similar to the above, but which are omitted, one can show that $0 < V(\hat{\boldsymbol{y}}^* = \boldsymbol{S}\boldsymbol{y}^*) < V(\boldsymbol{y}^*)$, which establishes the right hand inequality of Eq. (9a) and Eq. (9b), and also establishes that $\hat{y}^*_{\min} \neq \hat{y}^*_{\max}$.

Last, it remains to prove that for generic initial conditions, $y^*_i \neq \hat{y}^*_i$. Observe that $\hat{y}^*_i = y^*_i \Rightarrow \hat{y}_{avg} = \boldsymbol{1}_n^\top \hat{\boldsymbol{y}}^*/n$. Thus, $\hat{y}^*_i = y^*_i$ for m specific individuals if and only if there are m independent equations satisfying $(\mathbf{e}_i - \frac{1}{n}\boldsymbol{1}_n)^\top \boldsymbol{y}^* = 0$. This implies that $\hat{\boldsymbol{y}}^*$ must lie in an $n - m$-dimensional subspace of $\mathbb{R}^n$, denoted as $\mathcal{D}$. From Theorem 1, one has $\boldsymbol{y}^* = \boldsymbol{R}\boldsymbol{S}\boldsymbol{y}(0)$. It follows that $\hat{y}^*_i = y^*_i$ for m specific individuals only if $\boldsymbol{y}(0)$ belongs to the inverse image (by $\boldsymbol{R}\boldsymbol{S}$) of $\mathcal{D}$, and the inverse image has dimension $n - m$ because $\boldsymbol{R}, \boldsymbol{S}$ are invertible. This completes the proof. $\qquad \square$

B.5 Proof of Corollary 2

Recall the definition of V in Appendix A.1. From Theorem 1, one has that $V(\hat{\boldsymbol{y}}^*) = V(\boldsymbol{S}\boldsymbol{y}^*) \leq \tau(\boldsymbol{S})V(\boldsymbol{y}^*)$, which implies that there holds $V(\hat{\boldsymbol{y}}^*)/\tau(\boldsymbol{S}) \leq V(\boldsymbol{y}^*)$. Thus, Eq. (10) can be proved by showing that $\tau(\boldsymbol{S}) \leq \kappa(\boldsymbol{\phi})$. Note that since global public opinion $\hat{y}_{avg}$ is used, $\boldsymbol{M}$ in Eq. (5) becomes $\boldsymbol{M} = n^{-1}\boldsymbol{1}_n\boldsymbol{1}_n^\top$. Recall that $\boldsymbol{Q}_{22}^{-1}$ can be expressed as $\boldsymbol{Q}_{22}^{-1} = \sum_{k=0}^\infty \boldsymbol{P}_{22}$. Since $\boldsymbol{P}_{22} = n^{-1}(\boldsymbol{I}_n - \boldsymbol{\Phi})\boldsymbol{1}_n\boldsymbol{1}_n^\top$ and $\boldsymbol{Q}_{21} = -\boldsymbol{\Phi}$, we obtain $\boldsymbol{S} = \boldsymbol{\Phi} + \boldsymbol{H}$ where $\boldsymbol{H} \triangleq \sum_{k=1}^\infty \left[(\boldsymbol{I}_n - \boldsymbol{\Phi})\frac{\boldsymbol{1}_n\boldsymbol{1}_n^\top}{n}\right]^k \boldsymbol{\Phi} > 0$.

Let $\underline{a} = \min_{i,j} a_{ij}$ denote the smallest element of a matrix $\boldsymbol{A}$, and observe that $\underline{s} = \underline{h}$ because $\boldsymbol{S} = \boldsymbol{\Phi} + \boldsymbol{H}$ has the same offdiagonal entries as $\boldsymbol{H}$, and the ith diagonal entry of $\boldsymbol{S}$ is greater than that of $\boldsymbol{H}$ by $\phi_i > 0$. Since $\boldsymbol{S} > 0$, Eq. (A.2) yields $\tau(\boldsymbol{S}) \leq 1 - n\underline{s} \leq 1 - n\underline{h}$. We now analyse $\boldsymbol{H}$. For any $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, there holds

$$n^{-1}(\boldsymbol{I}_n - \boldsymbol{\Phi})\boldsymbol{1}_n\boldsymbol{1}_n^\top \boldsymbol{A}$$
$$= \frac{1}{n}\begin{bmatrix} (1-\phi_1)\sum_{j=1}^n a_{1j} & \cdots & (1-\phi_1)\sum_{j=1}^n a_{nj} \\ \vdots & \ddots & \vdots \\ (1-\phi_n)\sum_{j=1}^n a_{1j} & \cdots & (1-\phi_n)\sum_{j=1}^n a_{nj} \end{bmatrix}.$$

By recursion, we obtain that the $(i,j)^{th}$ entry of $[(\boldsymbol{I}_n - \boldsymbol{\Phi})\frac{\boldsymbol{1}_n\boldsymbol{1}_n^\top}{n}]^k$ is given by $\frac{(1-\phi_i)}{n^k}\gamma_k$, where

$$\gamma_k = \underbrace{\left[\sum_{p_1=1}^n \sum_{p_2=1}^n \cdots \sum_{p_{k-1}=1}^n (1-\phi_{p_1})(1-\phi_{p_2})\cdots(1-\phi_{p_{k-1}})\right]}_{\text{k-1 summation terms}}$$

This is obtained by recursively using $\sum_{i=1}^n \sum_{j=1}^n a_i b_j = \left(\sum_{i=1}^n a_i\right)\sum_{j=1}^n b_j = \sum_{i=1}^n a_i\left(\sum_{j=1}^n b_j\right)$. Next, define $\boldsymbol{Z}^k = [(\boldsymbol{I}_n - \boldsymbol{\Phi})\frac{\boldsymbol{1}_n\boldsymbol{1}_n^\top}{n}]^k \boldsymbol{\Phi}$. From the above, one can show that the $(i,j)^{th}$ element of $\boldsymbol{Z}^k$ is given by $z_{ij}(k) = \frac{1}{n^k}(1-\phi_i)\phi_j\gamma_k$. It follows that the smallest element of $\boldsymbol{Z}^k$, denoted by $\underline{z}(k)$, is bounded as follows

$$\underline{z}(k) \geq \frac{1}{n^k}(1-\phi_{\max})\phi_{\min}\gamma_k. \qquad (B.2)$$

Observe that $1 - \phi_i \geq 1 - \phi_{\max}, \forall i \Rightarrow \sum_{a=1}^n 1 - \phi_a \geq n(1 - \phi_{\max})$. It follows that

$$\underline{z}(k) \geq \frac{1}{n}\phi_{\min}(1-\phi_{\max})^k. \qquad (B.3)$$

Since $\boldsymbol{H} = \sum_{k=1}^{\infty} \boldsymbol{Z}^k$, there holds $\underline{h} \geq \sum_{k=1}^{\infty} \underline{z}(k) \geq \phi_{\min}(1 - \phi_{\max})(n\phi_{\max})^{-1}$. We can obtain this by noting that for any $r \in (-1, 1)$, the geometric series is $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r} \Leftrightarrow \sum_{k=1}^{\infty} r^k = \frac{1}{1-r} - 1$, and $0 < 1 - \phi_{\max} < 1$. From $\tau(\boldsymbol{S}), \tau(\boldsymbol{H}) \leq 1 - n\underline{h}$, and the above arguments, we obtain $\tau(\boldsymbol{S}) \leq 1 - n\underline{h} = 1 - \frac{\phi_{\min}}{\phi_{\max}}(1 - \phi_{\max}) = \kappa(\boldsymbol{\phi})$ as in the corollary statement. Since $0 < \phi_{\min}/\phi_{\max} < 1$ and $0 < 1 - \phi_{\max} < 1$, one has $0 < \kappa(\boldsymbol{\phi}) < 1$ and thus $\tau(\boldsymbol{S}) \leq \kappa(\boldsymbol{\phi})$ holds $\forall \phi_i \in (0, 1)$. $\square$

Key to the proof is that the coefficient of ergodicity for $\boldsymbol{S}$ is bounded from above as $\tau(\boldsymbol{S}) \leq \kappa(\boldsymbol{\phi})$. The tightness of $\tau(\boldsymbol{S}) \leq \kappa(\boldsymbol{\phi})$ depends on $\phi_{\min}/\phi_{\max}$: this can be concluded by examining the proof, and noting that the key inequalities in Eq. (B.2) and Eq. (B.3) involve $\phi_{\min}$ and $\phi_{\max}$. If $\phi_{\min}/\phi_{\max} = 1$, then $\tau(\boldsymbol{S}) = \kappa(\boldsymbol{\phi})$.

B.6 Proof of Corollary 3

First, verify that $\boldsymbol{S}$ is invertible, and continuously differentiable, for all $\phi_i \in (0, 1)$. From [46, Fact 10.11.20] we obtain

$$\frac{\partial \boldsymbol{S}(\boldsymbol{\phi})}{\partial \phi_i} = -\boldsymbol{S}(\boldsymbol{\phi}) \left(\frac{\partial \boldsymbol{S}^{-1}(\boldsymbol{\phi})}{\partial \phi_i} \right) \boldsymbol{S}(\boldsymbol{\phi}). \tag{B.4}$$

Below, the argument $\boldsymbol{\phi}$ will be dropped from $\boldsymbol{S}(\boldsymbol{\phi})$ and $\boldsymbol{S}^{-1}(\boldsymbol{\phi})$ when there is no confusion. Note that $\frac{\partial \boldsymbol{\Phi}^{-1}}{\partial \phi_i} = -\phi_i^{-2} \mathbf{e}_i \mathbf{e}_i^\top$. Using Eq. (B.1) and Eq. (B.4), one obtains $\frac{\partial \boldsymbol{S}(\boldsymbol{\phi})}{\partial \phi_i} = \phi_i^{-2} \boldsymbol{S} \mathbf{e}_i \left(\mathbf{e}_i^\top - \boldsymbol{m}_i^\top \right) \boldsymbol{S}$, where $\boldsymbol{m}_i^\top$ is the ith row of $\boldsymbol{M}$. It suffices to prove the corollary claim, if it can be shown that the row vector $\left(\mathbf{e}_i^\top - \boldsymbol{m}_i^\top \right) \boldsymbol{S}$ has a strictly positive i^{th} entry and all other entries are strictly negative. This is because $\boldsymbol{S} > 0 \Rightarrow \boldsymbol{S} \mathbf{e}_i > 0$. We achieve this

by showing that

$$(\mathbf{e}_i^\top - \boldsymbol{m}_i^\top)\boldsymbol{S}\mathbf{e}_i > 0 \tag{B.5}$$
$$(\mathbf{e}_i^\top - \boldsymbol{m}_i^\top)\boldsymbol{S}\mathbf{e}_j < 0 , \; \forall j \neq i. \tag{B.6}$$

Observe the following useful quantity:

$$\begin{aligned} \mathbf{e}_i^\top \boldsymbol{S}^{-1} &= \mathbf{e}_i^\top \left(\boldsymbol{\Phi}^{-1} - \boldsymbol{\Phi}^{-1}(\boldsymbol{I}_n - \boldsymbol{\Phi})\boldsymbol{M} \right) \\ &= \phi_i^{-1}\mathbf{e}_i^\top - (\phi_i^{-1} - 1)\boldsymbol{m}_i^\top. \end{aligned} \tag{B.7}$$

Postmultiplying by $\boldsymbol{S}$ on both sides of Eq. (B.7) yields $\mathbf{e}_i^\top = \phi_i^{-1}\mathbf{e}_i^\top \boldsymbol{S} - (\phi_i^{-1} - 1)\boldsymbol{m}_i^\top \boldsymbol{S}$. Rearranging this yields

$$\mathbf{e}_i^\top \boldsymbol{S} = \phi_i \mathbf{e}_i^\top + (1 - \phi_i)\boldsymbol{m}_i^\top \boldsymbol{S} \tag{B.8}$$
$$\boldsymbol{m}_i^\top \boldsymbol{S} = (1 - \phi_i)^{-1} \left(\mathbf{e}_i^\top \boldsymbol{S} - \phi_i \mathbf{e}_i^\top \right). \tag{B.9}$$

By using the equality of Eq. (B.8) for substitution, observe that the left hand side of Eq. (B.6) is

$$\begin{aligned} &(\mathbf{e}_i^\top \boldsymbol{S} - \boldsymbol{m}_i^\top \boldsymbol{S})\mathbf{e}_j \\ &= \left(\phi_i \mathbf{e}_i^\top + (1 - \phi_i)\boldsymbol{m}_i^\top \boldsymbol{S} - \boldsymbol{m}_i^\top \boldsymbol{S} \right)\mathbf{e}_j = -\phi_i \boldsymbol{m}_i^\top \boldsymbol{S}\mathbf{e}_j, \end{aligned}$$

because $\mathbf{e}_i^\top \mathbf{e}_j = 0$ for any $j \neq i$. Note that $\boldsymbol{m}_i^\top \boldsymbol{S}\mathbf{e}_j > 0$ because $\boldsymbol{M}$ being irreducible implies $\boldsymbol{m}_i^\top \neq \boldsymbol{0}_n^\top$. Thus, $-\phi_i \boldsymbol{m}_i^\top \boldsymbol{S}\mathbf{e}_j/n < 0$, which proves Eq. (B.6). Substituting the equality in Eq. (B.9), observe that the left hand side of Eq. (B.5) is

$$\begin{aligned} (\mathbf{e}_i^\top \boldsymbol{S} - \boldsymbol{m}_i^\top \boldsymbol{S})\mathbf{e}_i &= \mathbf{e}_i^\top \boldsymbol{S}\mathbf{e}_i - \frac{1}{1 - \phi_i} \left(\mathbf{e}_i^\top \boldsymbol{S}\mathbf{e}_i - \phi_i \mathbf{e}_i^\top \mathbf{e}_i \right) \\ &= \frac{\phi_i}{1 - \phi_i} \left(1 - \mathbf{e}_i^\top \boldsymbol{S}\mathbf{e}_i \right) > 0. \end{aligned} \tag{B.10}$$

The inequality is obtained by observing that 1) $\phi_i \in (0, 1) \Rightarrow \phi_i/(1 - \phi_i) > 0$, and 2) $1 - \mathbf{e}_i^\top \boldsymbol{S}\mathbf{e}_i > 0$ because $0 < \mathbf{e}_i^\top \boldsymbol{S}\mathbf{e}_i = s_{ii} < 1$. This proves Eq. (B.5). $\square$