Measuring Unconditional Stimulus Expectancy during Evaluative Conditioning Strengthens Explicit Conditional Stimulus Valence

Camilla C. Luck & Ottmar V. Lipp

School of Psychology, Curtin University, Australia

ARC-SRI: Science of Learning Research Centre

Author Notes

Address for correspondence:

Camilla C. Luck, School of Psychology, Curtin University, GPO Box U1987 Perth WA 6845, Australia. E-mail: c.luck@curtin.edu.au

Data from all experiments can be found at https://osf.io/xzsca/

Acknowledgements:

This work was supported by grants number DP180100869 and SR120300015 from the Australian Research Council.

Word count: 9081

Abstract

During evaluative conditioning, a neutral conditional stimulus (CS) becomes pleasant or unpleasant after pairings with a positive/negative unconditional stimulus (US). Measures of US expectancy are commonly assessed during conditioning but it is unclear whether this affects evaluative learning. In Experiment 1, we examined whether the concurrent assessment of US expectancy alongside measures of CS valence would influence the acquisition, extinction, and reinstatement of explicit CS valence. Participants rated both valence and expectancy during conditioning (valence/expectancy group) or only CS valence (valence only group). Evaluative conditioning was acquired in both groups during acquisition, but its magnitude was enhanced in the valence/expectancy group. Measuring US expectancy did not influence the extinction or reinstatement of conditional valence. In Experiment 2, we confirmed the enhancement of evaluative conditioning due to concurrent measurement of US expectancy in an explicit measure, but did not find corresponding evidence in an implicit measure of conditional valence. In Experiment 3, we replicated the results using a different US expectancy scale and demonstrated that measuring CS valence multiple times throughout conditioning also strengthens conditional valence. Overall, the results suggest that the measurement of US expectancy and CS valence throughout conditioning draws attention to the contingencies and strengthens explicit evaluative learning.

Key words: evaluative conditioning; valence; US expectancy measures; reinstatement; CS valence measures

Co-occurrences in our environment have a large impact on what we like and dislike. We may like a new colleague more if they are associated with an old friend or we may buy a new brand of coffee if it is endorsed by our favorite celebrity. These phenomena are based upon a type of learning known as evaluative conditioning. During evaluative conditioning, a neutral stimulus (conditional stimulus, CS; e.g. a picture of a shape) is paired with a positive or negative outcome (unconditional stimulus, US; e.g. a pleasant/unpleasant picture). After repeated pairings, the CS acquires positive or negative valence (becomes pleasant or unpleasant; see De Houwer, Thomas, & Baeyens, 2001 and De Houwer, Perugini, Baeyens, & Crombes, 2010 for reviews). Evaluative conditioning occurs in a wide range of situations and has vast practical implications – including for prejudice reduction (see Olson & Fazio, 2006) and for the treatment of anxiety disorders (see Dour, Brown, & Craske, 2015; Luck & Lipp, 2017; Zbozinek, Hermans, Prenoveau, Lio, & Craske, 2015).

The mechanisms underlying evaluative conditioning are heavily debated with most arguments centering on how the CS and the US are encoded in memory – via associative or propositional links – and on whether the CS-US link is referential or signal based in nature. Associations merely encode that stimuli are related (A and B co-occur), while, propositions encode *how* these stimuli are related (Stimulus A causes Outcome B *or* Stimulus A is an effect of Outcome B; see De Houwer, 2009). Propositional theorists argue that learning occurs *solely* due to the formation of propositions between the CS and the US (see De Houwer, 2009; De Houwer, Baeyens, & Field, 2005), whereas dual process theorists argue that learning occurs due to the formation of *both* propositions and associations (e.g. Gawronski & Bodenhausen, 2006). The referential versus signal learning debate focuses on what happens when the CS activates the representation of the US. If the CS-US link is signal-based then the presentation of the CS will generate an expectancy that the US will occur. On the other hand, if the CS-US link is referential then the presentation of the CS will activate the US representation (for instance make the individual think of the US) without generating an expectancy that it will occur.

Debates concerning the nature of the CS-US link are driven by reports that evaluative conditioning behaves in a different manner to other types of Pavlovian conditioning. Evaluative conditioning is argued to resist extinction, be sensitive only to contiguity (rather than contingency), to be acquired in the absence of contingency awareness, and to resist some types of contingency instructions (see De Houwer et al., 2001). Whether or not the research supports these differences, however, is heavily debated. For instance, some report that evaluative conditioning extinguishes at the same rate as expectancy conditioning (Luck & Lipp, 2019), some report that evaluative conditioning extinguishes at a slower rate than expectancy based learning (see Hoffman, De Houwer, Perugini, Baeyens & Crombez, 2010), and others report that evaluative conditioning does not extinguish at all (Gawronski, Gast, & De Houwer, 2014). Similarly, some report that evaluative conditioning can be acquired without awareness, but others report no evaluative learning in participants who are truly unaware of the contingencies (for reviews see Sweldens, Corneille, & Yzerbyt, 2014 and Lovibond & Shanks, 2002).

Theoretical investigations rely on comparing results across paradigms (i.e. evaluative conditioning responds in a different manner to fear conditioning in response to manipulation X) or on the inclusion of measures of expectancy learning within the evaluative conditioning paradigm. Comparisons across paradigms are problematic as factors that can affect learning, such as overall emotional arousal, are not controlled for. Including measures of US expectancy is advantageous in many ways and has been recommended (De Houwer et al., 2001). For instance, for experiments on the role of contingency awareness, it is important that the contingency awareness assessment takes place as close as possible to assessments of evaluative learning (known as the immediacy criterion, Lovibond & Shanks, 2002). This ensures that it is not affected by factors such as forgetting or interference. Measuring US expectancy throughout learning allows the exact moment that an individual becomes ‘contingency aware’ to be determined. The inclusion of expectancy measures alongside valence measures in a single study also adds more weight to dissociations between expectancy and evaluative conditioning. For instance, in the case of instructed extinction, informing participants that the US will no longer be presented eliminates expectancy conditioning but leaves evaluative conditioning intact (Lipp, Mallan, Libera, and Tan; 2010; Luck and Lipp, 2015; for a review see Luck and Lipp; 2016). This null finding for CS valence could indicate that evaluative learning does not respond to instructed extinction but could also reflect a failure of the instructional manipulation (i.e. maybe participants did not believe or notice the instruction). If US expectancy is measured at the same time, however, and does reduce after the instruction then the null finding observed for evaluative conditioning is validated by the within-experiment dissociation.

Although the inclusion of expectancy measures in evaluative conditioning is clearly advantageous in some respects, it is not known whether this assessment at multiple times during each learning phase would inadvertently influence evaluative learning. Many have suggested that this inclusion may draw attention to the CS-US contingencies and strengthen learning. Indeed in evaluative conditioning, Baeyens, Eelen, and van den Bergh (1990) report that including an expectancy measure does increases the number of participants who are deemed ‘contingency aware’, but this did not translate into an increase in the magnitude of evaluative conditioning in the small sample used (*n* = 16). In fear conditioning, US expectancy measures have been argued to be valid measures of fear learning (see Boddez, Baeyens, Luyten, Vansteenwegen, Hermans,

& Beckers, 2013) and are commonly assessed continuously throughout conditioning. There is evidence, however, that this inclusion magnifies conditional physiological fear responses. Warren et al. (2014) report that the acquisition of fear potentiated startle is strengthened, extinction slowed, and relapse exaggerated when a concurrent measure of US expectancy is included. Due to the clear importance of US expectancy measures to research aiming to understand the mechanisms underlying evaluative conditioning we set out to examine whether assessing US expectancy influences the acquisition, extinction, and reinstatement of evaluative learning across three experiments. We included extinction and reinstatement phases to extend our investigation to determine whether US expectancy measures reduce or inflate extinction and reinstatement effects. Reinstatement occurs when extinguished conditional responding returns after the unsignalled presentation of the USs and has recently been demonstrated in evaluative conditioning (see Luck & Lipp, 2019). In Experiment 1, we examined whether the inclusion of a US expectancy measure throughout conditioning would influence explicit CS valence ratings. In Experiment 2, we examined whether measuring US expectancy would influence implicit measures of CS valence. In Experiment 3, to ensure the findings were generalizable to other evaluative conditioning paradigms, we replicated the findings using a different type of expectancy measure and varying the CS valence assessment (throughout the conditioning phases vs. after each phase only).

Experiment 1

To examine whether the concurrent measurement of US expectancy would influence evaluative learning, one group evaluated CS valence and US expectancy throughout conditioning, while, another only evaluated CS valence. During acquisition, one shape (CSp) was paired with pleasant images and another shape was paired with unpleasant images (CSu).

Acquisition training was broken into four blocks with valence and expectancy (valence/expectancy group) or valence (valence only group) assessed after each block. During extinction, the CSp and CSu were presented alone followed by reinstatement. Based on the results reported by Warren et al. (2014), we hypothesized that the magnitude of acquired conditional valence would be larger during acquisition, extinction, and reinstatement in the valence/expectancy group.

Method

Participants. 76 participants from the United States (30 female) aged between 19 and 60 (M = 35.84, SD = 9.53), were recruited via Amazon's Mechanical Turk (Buhrmester, Kwang, & Gosling, 2011), and volunteered participation in exchange for $4.70 US. Ethics approval was obtained from the Curtin University Human Research Ethics Committee. An a-priori power analysis conducted using G*Power 3.1.9.4 revealed that 68 participants would be required to be 80% confident of detecting a medium to large effect (f = .35) for the CS × Group interaction at an alpha of .05. This was based on an ANOVA F test (within-between interaction; 2 groups [Valence/Expectancy, Valence only]; 2 measurement points [CSp, CSu]; Effect size specification: as in Cohen (1988); nonsphericity correction: 1).Participants were randomly assigned to the valence/expectancy group (n = 37) or the valence only group (n = 39). Twelve participants failed the contingency assessment. The report of the results is based on the full sample of participants, however, in any instance where the exclusion of these participants results in a notable deviation in the pattern of results we report both sets of statistics.

Apparatus/Stimuli. The experiment was programed using Inquisit 4 software, hosted on the Inquisit millisecond server, and run on Amazon Mechanical Turk using the Turkprime platform (Litman, Robinson, & Abberbock, 2016). CSs were four 900 × 675 pixel pictures of geometric shapes (square, triangle, diamond, circle) with black outlines on a white background, presented centered at a size of 50% of the participants' screen. The two shapes used during conditioning and the shape used as CSp and CSu were counterbalanced across participants. USs were six pleasant (codes: 1460, 1710, 2154, 2340, 5825, 5833; mean valence rating = 7.92 [1 = *low pleasure*, 9 = *high pleasure*]; mean arousal rating = 5.22 [1 = *low arousal*, 9 = *high arousal]*; subject matter: happy family, animals, nature scenery) and six unpleasant (codes: 9560, 9340, 9295, 9220, 2800, 2703; mean valence rating = 2.32; mean arousal rating = 4.93; subject matter: rubbish, starving children, cemetery scenes, sick animals) images taken from the International Affective Picture System (IAPS; Lang, Bradley & Cuthbert, 2008). USs were presented centered at a size of 60% of the participants' screen. CSs and USs were presented on a black background and the inter-trial interval was a 2s black screen. The reinstatement manipulation involved the presentation of the pleasant and unpleasant USs for 2s each in a random stream, with a 2s blank screen between pictures. CS valence was measured on a 1-9 Likert scale with the anchors 1 = *unpleasant*, 5 = *neutral*, 9 = *pleasant*. US expectancy was measured by asking participants to predict what type of pictures would follow the CSs on a 1-7 Likert scale with the anchors (1 = *unpleasant pictures will always follow*, 4 = *no pictures will follow*, 7 = *pleasant pictures will always follow*)

Procedure. Participants provided informed consent, completed a pre-rating of CSp and CSu valence, and were instructed to watch the screen and pay attention to the pictures. Participants then received four acquisition training blocks, each consisting of six CSp–$US_{pleasant}$ pairings and six CSu–$US_{unpleasant}$ pairings. Both CSs and USs were presented for 2s each with the presentation of the US following the CS immediately (delay conditioning). Each US was paired with its respective CS once during each block and pairing order was randomised. At the end of each block, participants were prompted to rate CS valence and then US expectancy. CS valence was assessed first to avoid contamination of the valence evaluations from the immediately preceding expectancy ratings. CSp and CSu rating order was randomised in each assessment. After acquisition, participants were presented with six extinction blocks, each consisting of six randomised presentations of both the CSp and CSu, with valence and expectancy assessed after each block. To induce reinstatement, the 12 US pictures were presented one at a time in a randomised stream immediately after the last extinction rating. Participants were then prompted to rate CS valence and US expectancy (reinstatement assessment). Demographic information was assessed and participants were asked to identify which shapes had been paired with the pleasant and unpleasant US images during the initial part of the experiment (contingency assessment).

Data Preparation and Analyses. The data from the US expectancy ratings were not central to the research question and are presented in the supplementary material. All frequentist statistics were conducted with IBM SPSS Statistics 25 with an alpha cut-off of .05. Phillai's trace statistics of the multivariate solution are reported in all analyses[1]. The report of the results is based primarily on frequentist statistics, however, we added Bayesian comparisons to the follow-up analyses to allow us to quantify evidence for the null hypothesis and for the readers' interest. These Bayesian comparisons were performed by comparing the follow-up effects of interest with paired or independent samples Bayesian t-tests (two sided). These Bayesian t-tests were conducted in Jasp using default settings (Cauchy prior scale: 0.707) and BF_{10} values are reported for all tests. BF_{10} values less than 1 provide evidence for the null hypothesis, whereas, BF_{10} values greater than 1 provide evidence for the alternative hypothesis. As Bayes factors provide a graded measure of evidence we did not define strict cut-off values, however, the reader should

[1] The multivariate solution is used as it has been shown to be more robust against assumption violations (see Vasey & Thayer, 1987).

note that following Jeffreys' (1961) conventions $BF_{10} < 0.33$ and $BF_{10} > 3$ are generally considered to provide substantial support for the null and alternative hypothesis, respectively.

An independent samples t-test was performed to check for any baseline group differences in age and a Pearson's chi-square test was performed to check for any group difference in the male to female sex ratio. A Pearson's chi-square test was performed on the contingency pass to fail ratio to determine whether the manipulation induced a group difference in contingency awareness. The CS valence evaluations measured during baseline and acquisition were subjected to a 2 CS (CSp, CSu) × 2 Group (valence/expectancy, valence only) × 5 Block (Baseline, A1, A2, A3, A4) mixed-model factorial ANOVA. The CS valence evaluations measured during extinction were subjected to a CS (CSp, CSu) × 2 Group (valence/expectancy, valence only) × 7 Block (A4, E1, E2, E3, E4, E5, E6) mixed-model factorial ANOVA. The last block of acquisition is included in the analyses to assess the influence of the first extinction training block. The CS valence evaluations measured after extinction and after reinstatement were subjected to a 2 CS (CSp, CSu) × 2 Group (valence/expectancy, valence only) × 2 Phase (E6, reinstatement test) mixed-model factorial ANOVA. To provide an index of the magnitude of the conditional response, difference scores for the valence ratings were calculated by subtracting the CSu score from the CSp score at each time point. The difference scores were subjected to the above ANOVAs removing the CS factor to determine whether the magnitude of the conditional response changed across the experimental phases. The evaluations of the positive and negative US pictures were averaged into mean scores and subjected to a 2 Valence (pleasant, unpleasant) × 2 Group (valence/expectancy, valence only) factorial ANOVA.

Results

Preliminary Analyses. The groups did not differ in age (valence/expectancy: *M* = 35.11, *SD* = 9.12; valence only: *M* = 36.55, *SD* = 9.99), $t(73) = 0.65$, $p = .515$, $BF_{10} = 0.29$, male to female sex ratio (valence/expectancy: 23:14; valence only: 23:16), $\chi^2(1) = 0.08$, $p = .776$, or contingency pass to fail ratio (valence/expectancy: 32:5; valence only: 32:7;), $\chi^2(1) = 0.28$, $p = .596$. The US valence results are displayed in Table 1. A main effect of valence, $F(1, 74) = 1105.75$, $p < .001$, $\eta p^2 = .937$, was moderated by a Valence × Group interaction, $F(1, 74) = 4.00$, $p = .049$, $\eta p^2 = .051$. Both groups, evaluated pleasant US pictures as more pleasant than unpleasant US pictures, both $F(1, 74) > 501.55$, $p < .001$, $\eta p^2 > .871$, $BF_{10} > 2.56 \times 10^{17}$. The valence/expectancy group evaluated the pleasant US pictures as more pleasant than the valence only group, $F(1, 74) = 4.44$, $p = .039$, $\eta p^2 = .057$, $BF_{10} = 1.57$, but the evaluations of the unpleasant US pictures did not differ between the groups, $F(1, 74) = 2.25$, $p = .138$, $\eta p^2 = .029$, $BF_{10} = 0.62$, although the Bayesian evidence for these differences is indecisive. The main effect of group did not reach significance, $F(1, 74) = 0.25$, $p = .618$, $\eta p^2 = .003$.

Acquisition. The CS valence evaluations measured throughout the experiment are presented in Figure 1. A main effect of CS, $F(1, 74) = 93.17$, $p < .001$, $\eta p^2 = .557$, a main effect of block, $F(4, 71) = 4.57$, $p = .002$, $\eta p^2 = .205$, a CS × Group interaction, $F(1, 74) = 5.14$, $p = .026$, $\eta p^2 = .065$, and a CS × Block interaction, $F(4, 71) = 27.15$, $p < .001$, $\eta p^2 = .605$, were detected. The CS × Group interaction revealed that the valence/expectancy group, $F(1, 74) = 69.22$, $p < .001$, $\eta p^2 = .483$, $BF_{10} = 9.70 \times 10^7$, and the valence only group, $F(1, 74) = 28.01$, $p < .001$, $\eta p^2 = .275$, $BF_{10} = 1444.00$, evaluated the CSp as more pleasant than the CSu, but this difference was larger in the valence/expectancy group. The Bayesian evidence for this difference provided anecdotal support for the alternative hypothesis, $BF_{10} = 2.11$. The CS × Block interaction revealed that evaluations of CSp and CSu did not differ during baseline, $F(1, 74) =$

0.11, p = .739, ηp^2 = .002, BF_{10} = 0.13, but the CSp was evaluated as more pleasant than the CSu during all other blocks all $F(1, 74) > 37.15$, $p < .001$, $\eta p^2 > .334$, $BF_{10} > 2.52 \times 10^5$. The magnitude of conditional valence increased from baseline to the end of acquisition, $p < .001$, $BF_{10} = 2.72 \times 10^{12}$. A marginal CS × Block × Group interaction was also detected, $F(4, 71)$ = 2.33, p = .064, ηp^2 = .116. Follow-up of this marginal effect confirmed the pattern of results detected in the two way interactions reported above but also revealed that the enhancement in the magnitude of conditional valence in the valence/expectancy group was not present at baseline, $F(1, 74)$ = 0.41, p = .522, ηp^2 = .006, BF_{10} = 0.28, or during the first acquisition block, $F(1, 74)$ = 1.20, p = .278, ηp^2 = .016, BF_{10} = 0.40 (anecdotal support for the null hypothesis), but was during the remaining acquisition blocks, all $F(1, 74) > 4.87$, $p < .031$, $\eta p^2 > .061$, BF_{10}: blocks 2 = 6.28, 3 = 3.50, and 4 = 1.89 (anecdotal support for the alternative hypothesis). The remaining omnibus effects did not reach significance, $F < 1.87$, $p > .176$, $\eta p^2 < .025$. In the sample of participants who passed the contingency assessment, the CS × Block × Group interaction, F(4, 59) = 1.15, p = .342, ηp^2 = .072, did not reach significance, and the CS × Group interaction, F(1, 62) = 2.87, p = .095, ηp^2 = .044, only attained marginal significance.

Extinction. A main effect of CS, $F(1, 74)$ = 50.12, $p < .001$, ηp^2 = .404, was moderated by a CS × Block interaction, $F(6, 69)$ = 9.70, $p < .001$, ηp^2 = .457. Conditional valence was present during all blocks, all $F(1, 74) > 18.90$, $p < .001$, $\eta p^2 > .203$, $BF_{10} > 405.50$, but decreased from after acquisition to after extinction, $p < .001$, $BF_{10} = 1.44 \times 10^8$. The CS × Group interaction, $F(1, 74)$ = 2.15, p = .147, ηp^2 = .028, and CS × Block × Group interaction, $F(6, 69)$ = 0.71, p = .646, ηp^2 = .058, did not reach significance. Bayesian comparisons provided anecdotal evidence that conditional valence did not differ between the groups at the end of extinction, BF_{10}

= 0.34. The remaining omnibus effects did not reach significance, largest main effect of group, $F(1, 74) = 3.39$, $p = .069$, $\eta p^2 = .044$[2], all other, $F < 1.09$, $p > .379$, $\eta p^2 < .087$.

Reinstatement. A main effect of CS, $F(1, 74) = 33.59$, $p < .001$, $\eta p^2 = .312$, was moderated by a CS × Phase interaction, $F(1, 74) = 11.13$, $p = .001$, $\eta p^2 = .131$. Conditional valence was present after extinction, $F(1, 74) = 18.90$, $p < .001$, $\eta p^2 = .203$, $BF_{10} = 417.16$, and after reinstatement, $F(1, 74) = 35.64$, $p < .001$, $\eta p^2 = .325$, $BF_{10} = 1.45 \times 10^5$, but was larger after reinstatement, $BF_{10} = 18.93$. The CS × Phase × Group interaction did not reach significance, $F(1, 74) = 0.43$, $p = .513$, $\eta p^2 = 006$, and the Bayesian analyses provided anecdotal evidence that conditional valence did not differ between the groups during reinstatement, $BF_{10} = 0.45$. The main effects of phase, $F(1, 74) = 3.30$, $p = .073$, $\eta p^2 = .043$, and group, $F(1, 74) = 0.71$, $p = .402$, $\eta p^2 = .010$, did not reach significance.

Discussion

In Experiment 1, we examined whether assessing US expectancy throughout conditioning would influence the acquisition, extinction, and reinstatement of evaluative learning. Evaluative conditioning was successfully acquired in both groups but conditional valence was larger in the group rating US expectancy throughout conditioning. This conditioning enhancement only attained marginal significance in participants who could verbalize the experimental contingencies. The size of evaluative learning did not differ significantly between the groups during extinction but conditional valence was numerically larger in the valence/expectancy group. It is possible that our experiment was underpowered to detect a potentially smaller difference between the groups during extinction but both groups did show a clear reduction in the size of the conditional response from after acquisition to after extinction, suggesting that the

[2] This marginal main effect of group reveals that *overall* evaluations (collapsed across CSp and CSu) are more unpleasant in valence/expectancy group.

trajectory of learning was similar in the two groups. The Bayesian analyses also provided anecdotal evidence against a difference between the groups at the end of extinction. Reinstatement was observed in both groups, replicating the results of Luck and Lipp (2019), but did not differ between them. This suggests that measuring US expectancy throughout conditioning does not influence reinstatement.

It is not clear why conditional valence was larger in the valence/expectancy group during acquisition. Two mechanism seem most likely based on the current results – the US expectancy measure enhances conditional valence because it strengthens *learning* or the expectancy measure enhances conditional valence because it creates a *response bias*. Measuring US expectancy during evaluative conditioning is reported to increase contingency awareness and it is this increased awareness that has been proposed to strengthen learning. The results of Experiment 1 do not provide conclusive support for this proposition. Finding only a marginal enhancement in conditional valence in participants who passed the contingency assessment could suggest that the larger conditional response detected in the full sample was driven by a higher proportion of participants in the valence/expectancy group who were aware of the contingencies. The post-experimental contingency pass to fail ratio, however, did not differ between the groups. It is possible, that participants in both groups acquired contingency awareness to a similar extent by the *end* of acquisition but the US expectancy measure led participants to acquire this knowledge earlier during acquisition. This possibility is consistent with the Bayesian analysis which reveals the strongest evidence for a difference between the groups during block 2, followed by block 3, and then block 4.

Alternatively, the enhancement of conditional valence in the valence/expectancy group could be due to a response bias induced by the bi-polar nature of the expectancy scale. During acquisition, ‘correct’ reporting of US expectancies requires participants to use the entire range of the expectancy scale (i.e. a rating of 1 indicates that the CS reliably predicts unpleasant pictures and a rating of 7 indicates that the CS reliably predicts pleasant pictures). Using the entire range of the expectancy scale may lead participants to use a wider range on the valence scale as well and therefore provide more extreme valence ratings. If the enhancement of conditional valence could be replicated using an implicit measure of valence (which does not require participants to respond using a scale), then a response bias explanation would be unlikely.

Experiment 2

In Experiment 2, we examined whether conditional valence assessed with an implicit measure would be larger in participants providing US expectancy ratings in addition to CS valence ratings. After acquisition, participants completed an affective priming task (see Fazio & Olson, 2003). We hypothesized, that implicit and explicit differential conditional valence would be larger in participants providing both valence and expectancy ratings during acquisition.

Method

Participants. 156 participants from the United States (78 female; 1 other; 1 missing) aged between 19 and 75 (M = 32.85, SD = 11.39) were recruited via Amazon’s Mechanical Turk and volunteered participation in exchange for $3.65 US. Ethics approval was obtained from the Curtin University Human Research Ethics Committee. As implicit measures are typically less sensitive than explicit measures the power analysis was calculated based on an estimated medium effect (d = 0.5 [f = .25]). A total sample size of 128 participants was required to be 80% confident of detecting a significant effect at an alpha of .05 between the priming scores obtained in the expectancy groups. This was based on a t-test (means: difference between two independent means; two tailed; allocation ratio N2/N1 = 1). Participants were randomly assigned to the valence/expectancy group ($n = 78$) or the valence only group ($n = 78$). Thirty-eight participants failed the contingency assessment. The pattern of the results remains similar when these participants are excluded and results from the entire sample are reported.

Apparatus/Stimuli. The stimuli used during the conditioning task, the CS and US valence scales, and the US expectancy scale were the same as in Experiment 1. Ten unpleasant (unpleasant, bad, horrible, miserable, hideous, dreadful, painful, repulsive, awful, and ugly) and 10 pleasant target words (pleasant, good, outstanding, beautiful, magnificent, marvellous, excellent, appealing, delightful, and nice) were used during the priming task. The CSp and CSu, presented at a size of 40% of the participants' screen, were used as primes. Primes and target words were presented on a black screen with a 500ms inter-trial interval (black screen). A red 'X' appeared on the screen as feedback after incorrect trials.

Procedure. Participants provided informed consent, completed the CS pre-rating, and acquisition as in Experiment 1. After acquisition, the affective priming task was started. A trial started with a 500ms fixation cross followed immediately by the prime for 200ms and the presentation of a pleasant or unpleasant target word until a response was made. Participants were asked to press the 'E' key for negative words and the 'I' key for positive words and this information remained at the bottom of the screen during the task. Participants completed 20 practice trials before starting the main task. In the main task, each CS was paired with the pleasant and unpleasant target words twice, totalling 80 randomized trials made up of 4 different types (congruent: CSp–pleasant target, CSu–unpleasant target; incongruent: CSp–unpleasant target, CSu–pleasant target). After affective priming, participants rated US valence and completed the demographic questions and contingency assessment as in Experiment 1.

Data Preparation and Analyses. The explicit CS and US valence evaluations were analysed as in Experiment 1. Reaction times from trials where participants incorrectly classified the target word and reaction times that were faster than 200 ms or 3 standard deviations outside of the participant's individual means were removed and scored as errors. Participants who made more than 30% of errors during priming or had a mean reaction time > 1000ms were excluded (22 total; valence/expectancy: 13; valence only: 9). The remainder of the data preparation and analyses were conducted as in Experiment 1.

Results

Preliminary Analyses. The male to female to other sex ratio (valence/expectancy: 40:36:1; valence only: 36:42:0), $\chi^2(2) = 1.67$, $p = .435$, and the contingency pass to fail ratio (valence/expectancy: 61:16; valence only: 56:22), $\chi^2(1) = 1.16$, $p = .283$, did not differ between the groups. Unexpectedly, the valence/expectancy group ($M = 30.16$, $SD = 8.84$) was younger than the valence only group ($M = 35.47$, $SD = 12.95$), $t(134.48) = 2.96$, $p = .004$, $BF_{10} = 8.76$. The valence evaluations of US pictures are displayed in Table 1. Pleasant US pictures were evaluated as more pleasant than unpleasant US pictures, $F(1, 153) = 1258.09$, $p < .001$, $\eta p^2 = .892$, $BF_{10} = 3.15 \times 10^{72}$. The remaining omnibus effects did not reach significance, $F < 0.02$, $p > .889$, $\eta p^2 < .001$.

Acquisition. Main effects of CS, $F(1, 154) = 135.87$, $p < .001$, $\eta p^2 = .469$, and block, $F(4, 151) = 5.13$, $p = .001$, $\eta p^2 = .120$, a CS × Group, $F(1, 154) = 4.91$, $p = .028$, $\eta p^2 = .031$, and a CS × Block interaction, $F(4, 151) = 39.31$, $p < .001$, $\eta p^2 = .510$, were moderated by a CS × Group × Block interaction, $F(4, 151) = 4.25$, $p = .003$, $\eta p^2 = .101$. Conditional valence was not present during baseline in either group, both $F(1, 154) < 0.63$, $p > .429$, $\eta p^2 < .005$, $BF_{10} < 0.18$, but was present during all blocks of acquisition in both groups, $F(1, 154) > 19.22$, $p < .001$, $\eta p^2 > .110$, $BF_{10} > 616.16$. The magnitude of this difference increased in size from baseline to the end of acquisition in both groups, both $p < .001$, $BF_{10} > 1.56 \times 10^7$. Conditional valence did not differ between the groups during baseline or block 1, both $F(1, 154) < 0.05$, $p > .825$, $\eta p^2 < .001$, $BF_{10} < 0.18$, but during all other blocks, differential valence was larger in the valence/expectancy group, all $F(1, 154) > 5.46$, $p < .021$, $\eta p^2 > .034$, BF_{10}: blocks 2 = 2.09 (anecdotal evidence for the alternative hypothesis), 3 = 11.04, 4 = 4.19. The remaining omnibus effects did not reach significance, $F < 0.09$, $p > .774$, $\eta p^2 < .001$.

Affective Priming. The mean reaction times were subjected to a 2 CS (CSp, CSu) × 2 Group (valence/expectancy, valence only) × 2 Target Valence (positive, negative) mixed-model factorial ANOVA and are displayed in Table 2. A main effect of target valence, $F(1, 132) = 10.50$, $p = .002$, $\eta p^2 = .074$, was moderated by a CS × Target Valence interaction, $F(1, 132) = 7.01$, $p = .009$, $\eta p^2 = .050$. This interaction was followed up by calculating a priming score [(incongruent: CSp–unpleasant target + CSu–pleasant target) – (congruent: CSp–pleasant target + CSu–unpleasant target)], which was positive ($M = 16.10$, $SD = 69.52$) and different from 0, reflecting an implicit evaluative conditioning effect, $t(133) = 2.68$, $p = .008$, $BF_{10} = 2.96$, although the BF_{10} only provides anecdotal evidence for this effect. The CS × Group × Target Word Valence interaction did not reach significance (priming scores: valence/expectancy: $M = 7.78$, $SD = 61.61$; valence only: $M = 23.93$, $SD = 75.84$), $F(1, 132) = 1.82$, $p = .180$, $\eta p^2 = .014$, and a Bayesian independent samples *t*-test provided anecdotal evidence against a difference between the groups, $BF_{10} = 0.42$. The remaining omnibus effects did not reach significance, $F < 2.79$, $p > .097$, $\eta p^2 < .021$.

Discussion

In Experiment 2, we examined whether implicit conditional valence would be enhanced in participants rating both valence and expectancy during acquisition. Replicating the results of

Experiment 1, explicit evaluative conditioning was present in both groups but was stronger in the valence/expectancy group. Conditional valence was detected in the implicit measure but this was not affected by the US expectancy ratings. Failing to find the conditioning enhancement in affective priming is not likely to be due to a lack of experimental power as the Bayesian analyses provided evidence against a difference between the groups and the priming score was numerically *smaller* in the valence expectancy group. Not finding a group difference, however, also does not conclusively exclude a learning account. To provide strong support for a learning account of the enhancement in evaluative learning the response bias explanation must be eliminated.

Experiment 3

In Experiment 3, we aimed to exclude a response bias account by replicating the findings of Experiments 1 and 2 using an expectancy measure that used a different response format and did not involve predicting the contingencies on a Likert scale requiring a keystroke. Participants instead received a two part expectancy question that required them to use the mouse to select options. Participants were first asked to predict whether pictures would follow the CS in the next part of the experiment (yes or no). If they answered “yes” they were then asked to select the type of pictures they expected from a set of options displayed in a list format. As this type of expectancy measure does not require participants to use a keyboard scale it should not bias participants to use the entire range of the valence scale and therefore, any enhancement in conditional valence detected in participants rating US expectancy during conditioning with this scale cannot be due to a response bias.

In Experiments 1 and 2, all participants were asked to evaluate CS valence multiple times during each learning phase. These repeated valence assessment may draw attention to the experimental contingencies and exert a similar strengthening effect on evaluative conditioning as the US expectancy measure. Therefore, to examine this possibility and to ensure that our findings are generalizable to experiments were CS valence is only assessed following each phase we added this as a factor in Experiment 3. Thus, a 2 Expectancy Group (online vs. post-phase; between-participants) × 2 Valence Group (online vs. post-phase; between-participants) × 2 CS (CSp vs. CSu; within-participant) × 4 Phase (baseline, post-acquisition, post-extinction, post-reinstatement; within-participant) experimental design was used to examine the influence of continuous valence and expectancy assessments on the acquisition, extinction, and reinstatement of evaluative learning.

We hypothesized, that if the enhancement in conditional valence detected in Experiments 1 and 2 was due to an enhancement in *learning* processes than conditional valence would be larger in participants rating US expectancy throughout conditioning. We also expected that online conditional valence assessments may exert a similar enhancement on conditional valence. Finally, if conditional valence assessments do exert a learning enhancement via similar mechanisms, we expected that the influence of US expectancy ratings would be *larger* in those participants who were not evaluating CS valence throughout conditioning as removing the CS valence assessments would likely exaggerate the enhancement present due to the expectancy measure.

Method

Participants. 273 participants from the United States (105 female; 1 missing) aged between 20 and 72 (M = 37.00, SD = 11.06) were recruited via Amazon's Mechanical Turk and volunteered participation in exchange for $4.30 US. Ethics approval was obtained from the Curtin University Human Research Ethics Committee. An a-priori power analysis conducted using G-Power 3.1 revealed that 256 participants would be required to be 80% confident of detecting a medium effect (f = .25) for the highest order interaction (CS × Expectancy Group × Valence Group × Phase). This was conducted based on an ANOVA F test (within-between interaction; 4 groups [online-expectancy/online-valence; post-expectancy/online-valence; post-expectancy/online-valence; post-expectancy/post-valence]; 4 measurement points [baseline; post-acquisition; post-extinction; reinstatement Test]; Effect size specification: as in Cohen (1988); nonsphericity correction: 1) using the CS difference scores. Participants were randomly assigned to one of the four subgroups (online-expectancy/online-valence = 68; online-expectancy/post-valence = 68; post-expectancy/online-valence = 68; post-expectancy/post-valence = 69). Sixty-six participants failed the contingency assessment and one participant did not answer the contingency question. The pattern of the results does not change when these participants are excluded and results from the entire sample are reported.

Apparatus/Stimuli. US expectancy was measured using a two-step question. The CSs were displayed on the screen individually and participants were first asked “Do you expect pictures to follow this shape in the next part of the experiment?” Participants were required to click “No” or “Yes”. If participants, answered “No” the experiment progressed without the second question. If participants answered “Yes” they were then asked “What type of pictures do you expect to follow this shape?” while the CS continued to be displayed on the screen. Participants were required to click one of 4 options displayed horizontally in the following order on the bottom of the screen. “Pleasant pictures”; “Unpleasant Pictures”’ “Both Pleasant and Unpleasant Pictures” or “Unsure”. The rest of the apparatus/stimuli were the same as in Experiment 1.

Procedure. The baseline valence assessment, the acquisition and extinction training, the reinstatement manipulation, the US valence assessment, and the post-experimental questionnaire was the same as in Experiment 1. Participants in the post-valence group rated CS valence at only 4 time-points during the experiment – baseline (B), after the last block of acquisition (A4), after the last block of extinction (E6), and after reinstatement (R). Participants in the online-valence group rated valence after every block as in Experiments 1 and 2. Participants in the post-expectancy group rated US expectancy at only 3 time-points during the experiment – after the last block of acquisition (A4), after the last block of extinction (E6), and after reinstatement (R). Participants in the online-expectancy group rated valence after every block as in Experiments 1 and 2. Thus, participants in the online-expectancy/online-valence group provided expectancy ratings and CS valence evaluations online throughout conditioning; participants in the online-expectancy/post-valence group provided expectancy ratings throughout conditioning but CS valence evaluations only after each phase; participants in the post-expectancy/online-valence group provided expectancy ratings after each phase and valence evaluations throughout conditioning; and participants in the post-expectancy/post-valence group provided expectancy ratings and CS valence evaluations only at the end of each phase.

Data Preparation and Analyses. The CS valence evaluations provided at baseline, post-acquisition, post-extinction, and post-reinstatement were analyzed with a 2 Expectancy Group (online vs. post-phase; between-participants) × 2 Valence Group (online vs. post-phase; between-participants) × 2 CS (CSp vs. CSu; within-participant) × 4 Phase (baseline, post-acquisition, post-extinction, post-reinstatement; within-participant) mixed-model ANOVA and are presented in Figure 3. The categorical US expectancy ratings measured as part of the experimental manipulation were not of interest in the current investigation and are presented in the supplementary material. Pearson's chi-square tests and a 2 Expectancy Group (online vs. post-phase) × 2 Valence Group (online vs. post-phase) between-participants ANOVA were performed to check for baseline group differences in sex and age, respectively. The mean evaluations of the US pictures were analyzed with a 2 Expectancy Group (online vs. post-phase) × 2 Valence Group (online vs. post-phase) × 2 US Valence (pleasant, unpleasant) mixed-model ANOVA. Pearson's chi square tests were performed to determine whether the pass to fail contingency ratio differed between the valence and expectancy groups. The remainder of the data preparation and analyses were performed as in Experiment 1.

Results

Preliminary Analyses. The male to female ratio did not differ between the expectancy groups (online-expectancy: 87:49; post-expectancy: 80:56), $\chi^2(1) = 0.76$, $p = .383$, or the valence groups (online-valence: 85:50; post-valence: 82:55), $\chi^2(1) = 0.28$, $p = .598$. No significant differences in age were detected between the groups, all $F < 1.91$, $p > .168$, $\eta p^2 < .008$. The mean US valence evaluations are displayed in Table 1. Pleasant US pictures were evaluated as more pleasant than unpleasant US pictures, $F(1, 269) = 2107.52$, $p < .001$, $\eta p^2 = .887$, $BF_{10} = 2.66 \times 10^{126}$. The remaining omnibus effects did not reach significance, $F < 1.04$, $p > .309$, $\eta p^2 < .004$.

Contingency Awareness. The post-experimental contingency pass to fail ratio was higher in the online expectancy group (110:26) than in the post expectancy group (96:40), $\chi^2(1) = 3.92$, $p = .048$. Similarly, the contingency pass to fail ratio was higher in the online valence group (111:24) than in the post valence group (95:42), $\chi^2(1) = 6.14$, $p = .013$. Removing participants who did not pass the contingency assessment does not change the pattern of results and the report is based on the full sample.

Conditioning. The CS valence evaluations are displayed in Table 3 and the conditional valence scores are displayed in Figure 3. Main effects of CS, $F(1, 269) = 144.32$, $p < .001$, $\eta p^2 = .349$, and phase, $F(3, 267) = 11.28$, $p < .001$, $\eta p^2 = .113$, as well as CS × Phase, $F(3, 267) = 92.17$, $p < .001$, $\eta p^2 = .509$, CS × Valence Group, $F(1, 269) = 4.51$, $p = .035$, $\eta p^2 = .017$, CS × Valence Group × Phase, $F(3, 267) = 7.01$, $p < .001$, $\eta p^2 = .073$, and CS × Expectancy Group × Phase, $F(3, 267) = 3.72$, $p = .012$, $\eta p^2 = .040$, interactions were detected. The CS × Expectancy Group × Valence Group interaction, $F(1, 269) = 0.22$, $p = .643$, $\eta p^2 = .001$, and the CS × Expectancy Group × Valence Group × Phase interaction, $F(3, 267) = 0.33$, $p = .803$, $\eta p^2 = .004$, did not reach significance, $F(3, 267) = 0.33$, $p = .803$, $\eta p^2 = .004$.

The CS × Expectancy Group × Phase interaction revealed that conditional valence was not present in the post-expectancy group during baseline, $F(1, 269) = 0.01$, $p = .913$, $\eta p^2 < .001$, $BF_{10} = 0.10$. Unexpectedly, the online-expectancy group evaluated the CSu ($M = 5.93$, $SD = 1.70$) as more pleasant than the CSp ($M = 5.64$, $SD = 1.51$) during baseline, $F(1, 269) = 5.16$, $p = .024$, $\eta p^2 = .019$, $BF_{10} = 0.93$, however, the Bayesian evidence was indecisive. After acquisition, after extinction and during reinstatement, conditional valence was present in both groups, all $F(1, 269) > 14.37$, $p < .001$, $\eta p^2 > .050$, $BF_{10} > 329.42$. Conditional valence increased from baseline to post-acquisition, both $p < .001$, $BF_{10} > 2.05 \times 10^{14}$, and decreased from post-acquisition to post-extinction in both groups, both $p < .001$, $BF_{10} > 7.78 \times 10^7$. After reinstatement, the magnitude of conditional valence increased significantly in the post-expectancy group, $p =. 033$, $BF_{10} = 0.92$, but the Bayesian evidence was indecisive. Conditional valence did not change in the online-expectancy group, $p = .619$, $BF_{10} = 0.11$. The magnitude of conditional valence did not differ between the groups during baseline, $F(1, 269) = 2.35$, $p = .127$, $\eta p^2 = .009$, $BF_{10} = 0.40$ (anecdotal evidence for the null hypothesis), after extinction, $F(1, 269) = 2.01$, $p = .158$, $\eta p^2 =$

.007, BF_{10} = 0.35 (anecdotal evidence for the null hypothesis), or during reinstatement, $F(1, 269) = 0.32$, $p = .570$, $\eta p^2 = .001$, BF_{10} = 0.16, but was larger in the online-expectancy group after acquisition, $F(1, 269) = 6.29$, $p = .013$, $\eta p^2 = .023$, BF_{10} = 2.43 (anecdotal evidence for the alternative hypothesis).

The CS × Valence Group × Phase interaction revealed that conditional valence was not present during baseline in the online-valence group, $F(1, 269) = 0.22$, $p = .642$, $\eta p^2 = .001$, BF_{10} = 0.10. Unexpectedly, the post-valence group evaluated the CSu (*M* = 5.89, *SD* = 1.52) as marginally more pleasant than the CSp (*M* = 5.65, *SD* = 1.49), $F(1, 269) = 3.69$, $p = .056$, $\eta p^2 = .014$, BF_{10} = 0.96, but the Bayesian evidence was indecisive. After acquisition, after extinction, and during the reinstatement test, conditional valence was present in both valence groups, all $F(1, 269) > 19.05$, $p < .001$, $\eta p^2 > .066$, $BF_{10} > 921.12$. Conditional valence increased from baseline to post-acquisition in both groups, both $p < .001$, $BF_{10} > 3.11 \times 10^{17}$, and decreased from post-acquisition to post-extinction in both groups, both $p < .001$, $BF_{10} > 2.20 \times 10^{6}$. While conditional valence increased after reinstatement in the online valence assessment group, $p < .001$, BF_{10} = 28.42, it did not change in the post valence assessment group, $p = .300$, BF_{10} = 0.17. The magnitude of conditional valence did not differ between the groups during baseline, $F(1, 269) = 1.05$, $p = .306$, $\eta p^2 = .004$, BF_{10} = 0.22, or after extinction, $F(1, 269) = 0.29$, $p = .592$, $\eta p^2 = .001$, BF_{10} = 0.15, but conditional valence was larger in the online valence assessment group after acquisition, $F(1, 269) = 10.79$, $p = .001$, $\eta p^2 = .039$, BF_{10} = 19.90, and was marginally larger in the online valence assessment group during reinstatement, $F(1, 269) = 3.44$, $p = .065$, $\eta p^2 = .013$, BF_{10} = 0.69, but the Bayesian evidence for this difference was indecisive. The remaining omnibus effects did not reach significance, $F < 2.02$, $p > .112$, $\eta p^2 < .023$.

Discussion

In Experiment 3, we examined whether the conditioning enhancement detected in Experiments 1 and 2 would replicate using a different US expectancy measure. We changed the nature of the US expectancy measure to a categorical expectancy question to eliminate the possibility that the scale format used in the previous experiments biased participants to use a wider range of the CS valence scale. We also added the timing of CS valence assessments (online vs. post) as a factor in the experiment to ensure that the US expectancy enhancement would replicate when CS valence was not measured online throughout conditioning. This also allowed us to examine whether measuring CS valence online would enhance conditional valence. Replicating the results of Experiments 1 and 2, all participants acquired evaluative conditioning but conditional valence was larger at the end of acquisition in participants rating US expectancy online during conditioning. As in Experiment 1, the US expectancy measure did not influence extinction with a clear reduction in conditional valence occurring after extinction in both groups. Some evidence of reinstatement was observed in participants rating US expectancy after each phase, but this evidence was indecisive in the Bayesian analyses. Unexpectedly, reinstatement was not observed in participants rating expectancy online during conditioning. This is not consistent with the findings of Experiment 1 or with the results of Luck and Lipp (2019) who found reinstatement when participants evaluated US expectancy throughout learning.

Interestingly, the CS valence measure exerted a similar, but independent, strengthening effect on evaluative learning. At the end of acquisition, conditional valence was present in both groups but was larger in participants evaluating CS valence online. Extinction was not influenced by the timing of the CS valence assessment, with a clear extinction effect observed in both groups and no difference in the strength of evaluative conditioning detected between the groups after extinction. Unexpectedly, reinstatement was only observed when participants evaluated CS valence online throughout learning and not in participants who evaluated CS valence only post-experimentally. It should be noted that reinstatement did not occur in these participants in the expectancy measure either and therefore this finding will require further exploration and should be interpreted with care.

We observed a difference in the pass to fail contingency ratio between both the expectancy and the valence groups – with participants in the online groups showing a higher pass to fail ratio. This finding provides evidence that the mechanism underlying the enhancement in evaluative conditioning in both the expectancy and valence groups is likely to be an increase in contingency salience. The pattern of evaluative conditioning does not differ when participants who failed the contingency assessment are removed, however, which suggests that the findings are not driven merely by a larger proportion of participants in the post assessment groups failing to learn.

General Discussion

The current investigation examined whether assessing US expectancy during conditioning would influence the acquisition, extinction, and reinstatement of evaluative conditioning. In Experiment 1, we demonstrated that measuring US expectancy during acquisition enhanced the magnitude of conditional valence after acquisition but did not influence the extinction or reinstatement of evaluative learning. In Experiment 2, we replicated this effect on the explicit ratings but found no evidence for an enhancement of implicit conditional valence. In Experiment 3, we used a different US expectancy measure to eliminate the possibility that the conditioning enhancement was caused by a response bias. As in Experiment 1, conditional evaluative learning was strengthened after acquisition in the participants rating US expectancy during conditioning, but conditional valence did not differ between the expectancy groups after extinction. This enhancement for participants rating US expectancy online was not influenced by whether participants evaluated CS valence online or only at the end of each experimental phase. Evaluating CS valence online, however, resulted in a similar, but independent, strengthening of conditional valence after acquisition but not after extinction.

In Experiment 3, the magnitude of evaluative conditioning did not differ between the expectancy groups during reinstatement but only the post-expectancy assessment group showed an increase in conditional valence after reinstatement. This is surprising as it would suggest that reinstatement does not occur if US expectancy is rated *online* but this conclusion is not consistent with the results of Experiment 1 or with the experiments reported by Luck and Lipp (2019) and therefore should be interpreted with care. Interestingly, the opposite pattern emerged for the CS valence assessment groups, with only the online-valence group showing evidence of reinstatement. Reinstatement effects are typically small and dependent on the size of the original conditional response and therefore the observation that reinstatement did not occur in the post-valence group could be due to this group developing weaker conditional valence during acquisition.

The results of Experiment 3 provide strong evidence that the enhancement of conditional valence is not simply due to a response bias. Instead, the observation that participants in the online-expectancy group of Experiment 3 showed a higher pass to fail ratio than participants in the post-expectancy group suggests that US expectancy assessments likely strengthen conditional valence by drawing attention to the experimental contingencies. This observation is consistent with previous proposals (see Lovibond & Shanks, 2002), with the findings of Baeyens et al. (1990), and with the observation that the conditional valence enhancement in participants rating expectancy throughout conditioning was weakened in Experiment1 when participants who failed the contingency assessment were removed. The findings cannot merely be driven by a larger proportion of participants failing to learn when US expectancy is not measured, or measured only at the end of each phase, as in Experiments 2 and 3 the pattern of results does not differ when participants who failed the contingency assessment are removed. Rather, it is likely that the US expectancy measure draws attention to experimental contingencies and this creates a stronger conditional response *and* makes a higher percentage of participants aware of the contingencies. The parallel results observed for participants evaluating valence online suggests that online CS valence measurements likely also enhance evaluative conditioning because they draw attention to the experimental contingencies.

If the mechanism underlying the US expectancy effect is an enhancement in learning processes it is not clear why it did not emerge in the priming results. The priming measure itself suffers from low reliability and may not be sensitive enough to detect a small enhancement in evaluative learning (see Gawronski & De Houwer, 2014). It is also possible that the measure *itself* reduced evaluative learning. During affective priming each CS is presented with positive and negative target words in rapid succession – a procedure which could be considered similar to evaluative conditioning with a 50% reinforcement schedule. Failing to find evidence for a difference between the groups in the implicit measure, however, raises the possibility that the enhancement in conditional valence may occur because the addition of the US expectancy measure may increase demand pressure during acquisition. Although this possibility cannot be entirely discounted, it is not consistent with the observation that more participants in the online expectancy group were aware of the contingencies in Experiment 3. It is also not consistent with the results of Warren et al. (2014) who in fear conditioning found that US expectancy enhanced conditional startle potentiation – a measure that is not likely to be sensitive to the effects of experimental demand (Lipp, 2006).

The results have implications for evaluative conditioning research as they suggest that measuring US expectancy and CS valence continuously during learning may change the type of learning processes involved. We are *not* proposing, however, that these measures should not be used. A mild enhancement in evaluative learning may be a worthwhile cost to pay in the investigation of research questions that require the assessment of evaluative and expectancy learning in the same sample in close temporal proximity. In some evaluative conditioning research, enhancing evaluative learning may also be highly valuable as it could magnify small effects allowing researchers to examine them with more power. Instead, the results indicate that researchers need to be aware that online assessments of both valence and expectancy may influence evaluative conditioning and will need to way up the potential benefits and drawbacks of including the measure/s. The observation that assessing both expectancy and valence did not influence extinction does not necessarily imply that the measure will not influence the results of other procedures or other experimental manipulations. Some experiments may be more susceptible to this influence than others – in instruction research for instance, increasing contingency salience – and therefore possibly propositional learning processes – could have a large impact on how the instructional manipulation works. In designs such as these where the US expectancy measure is highly useful researchers may need to conduct pilot work to ensure the measure does not influence the process they are observing. The assessment and timing of US expectancy will also need to be taken into account when comparing evaluative conditioning results across studies as evaluative conditioning effects will likely be larger in studies assessing

US expectancy throughout learning and could be a ‘boundary condition’ explaining some discrepancies in the literature.

References

Baeyens, F., Eelen, P., & Bergh, O. v. d. (1990). Contingency awareness in evaluative conditioning: A case for unaware affective-evaluative learning. *Cognition and Emotion*, *4*(1), 3-18. doi:10.1080/02699939008406760

Boddez, Y., Baeyens, F., Luyten, L., Vansteenwegen, D., Hermans, D., & Beckers, T. (2013). Rating data are underrated: Validity of US expectancy in human fear conditioning. *Journal of Behavior Therapy and Experimental Psychiatry, 44*(2), 201-206. doi:10.1016/j.jbtep.2012.08.003

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science, 6*, 3-5. doi:10.1177/1745691610393980

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. New York, NY: Routledge Academic.

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior, 37*, 1-20. doi:10.3758/LB.37.1.1

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*, 853-869. doi:10.1037/0033-2909.127.6.853

De Houwer, J. D., Baeyens, F., & Field, A. P. (2005). Associative learning of likes and dislikes: Some current controversies and possible ways forward. *Cognition & emotion, 19*, 161-174. doi:10.1080/02699930441000265

Fazio, R. H., & Olson, M. A. (2003). Implicit Measures in Social Cognition Research: Their Meaning and Use. *Annual Review of Psychology, 54*, 297-327. doi:10.1146/annurev.psych.54.101601.145225

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and Propositional Processes in Evaluation: An Integrative Review of Implicit and Explicit Attitude Change. *Psychological Bulletin, 132*, 692-731. doi:10.1037/0033-2909.132.5.692

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H.T. Reis & C.M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (2nd ed., pp. 283-310). New York, NY: Cambridge University Press.

Gawronski, B., Gast, A., & De Houwer, J. (2015). Is evaluative conditioning really resistant to extinction? Evidence for changes in evaluative judgements without changes in evaluative representations. *Cognition and Emotion, 29*(5), 816-830. doi:10.1080/02699931.2014.947919

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin, 136*, 390-421. doi:10.1037/a0018916

Jeffreys, H. (1961). *Theory of Probability* (3 ed.). Oxford, UK: Oxford University Press.

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical Report A-8.* University of Florida, Gainesville, FL.

Lipp, O. V. (2006). Human fear learning: Contemporary procedures and measurement. In M. G. Craske, D. Hermans & D. Vansteenwegen (Eds.), (2006). Fear and learning: From basic processes to clinical implications (pp. 37-52). Washington: APA Books.

Lipp, O. V., Mallan, K. M., Libera, M., & Tan, M. (2010). The effects of verbal instruction on affective and expectancy learning. *Behaviour Research and Therapy, 48*, 203-209. doi:10.1016/j.brat.2009.11.002

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*, 433-442. doi:10.3758/s13428-016-0727-z

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. Journal of Experimental Psychology: Animal Behavior Processes, 28(1), 3-26. doi:10.1037/0097-7403.28.1.3

Luck, C. C., & Lipp, O. V. (2015). A potential pathway to the relapse of fear? Conditioned negative stimulus evaluation (but not physiological responding) resists instructed extinction. *Behaviour Research and Therapy, 66,* 18-31. doi:10.1016/j.brat.2015.01.001

Luck, C. C., & Lipp, O. V. (2016). Instructed extinction in human fear conditioning: History, recent developments, and future directions. *Australian Journal of Psychology, 68*, 209-227. doi:10.1111/ajpy.12135

Luck, C. C., & Lipp, O. V. (2019). Relapse of evaluative learning – Evidence for reinstatement, renewal, but not spontaneous recovery, of extinguished evaluative learning in a picture-picture evaluative conditioning paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* In press. doi: 10.1037/xlm0000785

O'Brien, F., & Cousineau, D. (2014). Representing Error bars in within-subject designs in typical software packages. Tutorials in Quantitative Methods for Psychology, 10, 56-67. doi:10.20982/tqmp.10.1.p056

Olson, M. A., & Fazio, R. H. (2006). Reducing Automatically Activated Racial Prejudice Through Implicit Evaluative Conditioning. *Personality and Social Psychology Bulletin, 32,* 421-433. doi:10.1177/0146167205284004

Sweldens, S., Corneille, O., & Yzerbyt, V. (2014). The Role of Awareness in Attitude Formation Through Evaluative Conditioning. Personality and social psychology review, 18(2), 187-209. doi:10.1177/1088868314527832

Vasey, M. W., & Thayer, J. F. (1987). The Continuing Problem of False Positives in Repeated Measures ANOVA in Psychophysiology: A Multivariate Solution. *Psychophysiology, 24*(4), 479-486. doi:10.1111/j.1469-8986.1987.tb00324.x

Vervliet, B., Craske, M. G., & Hermans, D. (2013). Fear Extinction and Relapse: State of the Art. *Annual Review of Clinical Psychology, 9*(1), 215-248. doi:doi:10.1146/annurev-clinpsy-050212-185542

Warren, V. T., Anderson, K. M., Kwon, C., Bosshardt, L., Jovanovic, T., Bradley, B., & Norrholm, S. D. (2014). Human fear extinction and return of fear using reconsolidation update mechanisms: The contribution of on-line expectancy ratings. *Neurobiology of Learning and Memory, 113,* 165-173. doi:10.1016/j.nlm.2013.10.014

Table 1

Means and Standard Deviations for the Unconditional Stimuli in Experiments 1, 2, 3.

Group	Positive Pictures	Negative Pictures
Experiment 1		
Valence/Expectancy	8.30 (0.67)	1.70 (0.66)
Valence Only	7.88 (1.03)	2.03 (1.16)
Experiment 2		
Valence/Expectancy	7.89 (0.94)	2.12 (1.44)
Valence Only	7.92 (0.99)	2.10 (1.33)
Experiment 3		
Online-Expectancy/Online-Valence	7.95 (1.26)	2.15 (1.59)
Online-Expectancy/Post-Valence	8.07 (0.89)	2.22 (1.49)
Post-Expectancy/Online-Valence	8.07 (0.83)	2.02 (1.44)
Post-Expectancy/Post-Valence	7.88 (0.90)	2.29 (1.50)

Table 2

Mean Reaction Times and Standard Deviations for the Prime-Target Categories assessed during Affective Priming.

Prime-Target Category	**Valence/Expectancy**	**Valence Only**
CSp—Positive Target	584.73 (127.23)	610.10 (130.32)
CSp—Negative Target	596.24 (118.00)	643.15 (150.56)
CSu—Positive Target	590.17 (113.86)	621.97 (128.79)
CSu—Negative Target	593.90 (110.11)	631.08 (137.34)

Table 3

Mean and Standard Deviations for Conditional Stimulus Valence Evaluations Measured in Experiment 3.

Variable	**Online-Expectancy/Online Valence**	**Online-Expectancy/Post-Valence**	**Post-Expectancy/Online Valence**	**Post-Expectancy/Post-Valence**
Baseline				
CSp	5.81 (1.70)	5.47 (1.29)	5.94 (1.37)	5.83 (1.65)
CSu	6.03 (1.83)	5.82 (1.57)	5.84 (1.47)	5.96 (1.47)
Post-Acquisition				
CSp	7.60 (1.76)	7.07 (1.60)	7.13 (1.70)	6.65 (1.89)
CSu	3.06 (2.43)	3.81 (2.21)	3.54 (2.44)	4.52 (2.38)
Post-Extinction				
CSp	6.18 (1.69)	6.21 (1.61)	5.63 (1.76)	6.10 (1.73)
CSu	4.99 (2.03)	4.81 (1.91)	4.85 (1.84)	5.19 (1.97)
Reinstatement				
CSp	6.37 (1.75)	5.85 (1.69)	6.07 (1.44)	5.94 (1.68)
CSu	4.82 (2.21)	4.65 (1.75)	4.46 (1.59)	5.17 (1.84)

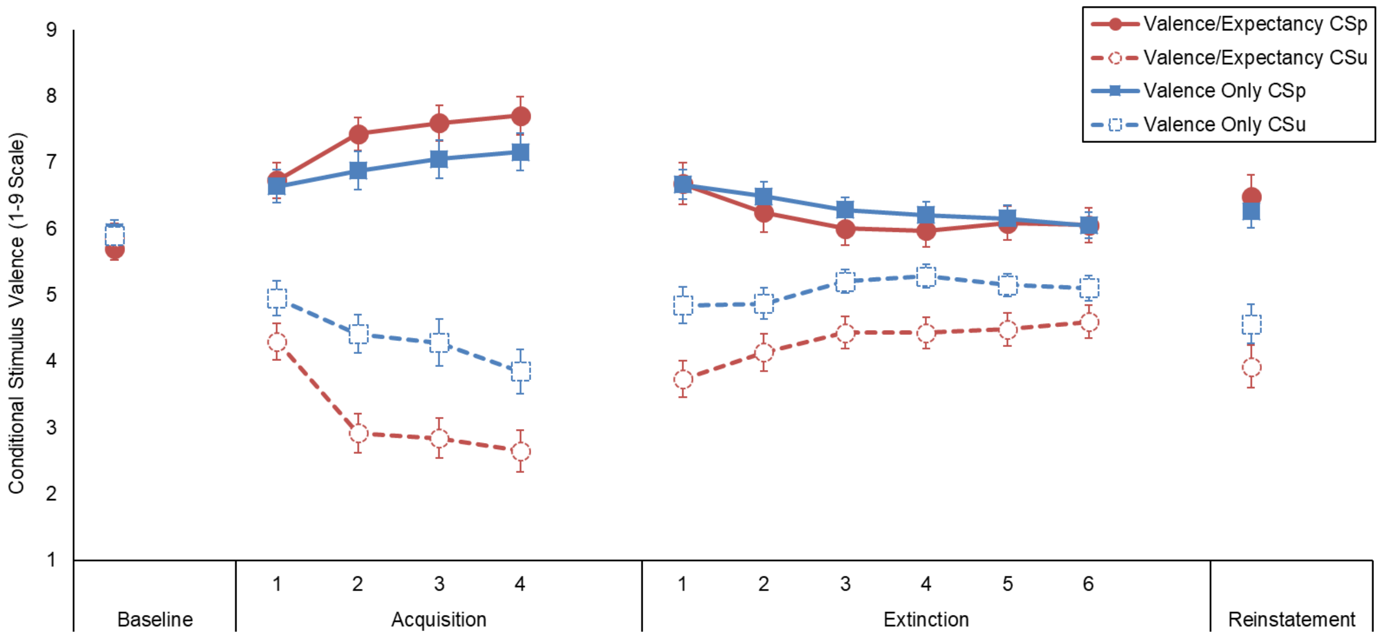

Figure 1. Conditional stimulus valence evaluations measured during baseline, acquisition, extinction, and the reinstatement test of Experiment 1. Higher values indicate higher stimulus pleasantness. Error bars represent SEMs for within-participant designs based on O'Brien and Cousineau (2014).

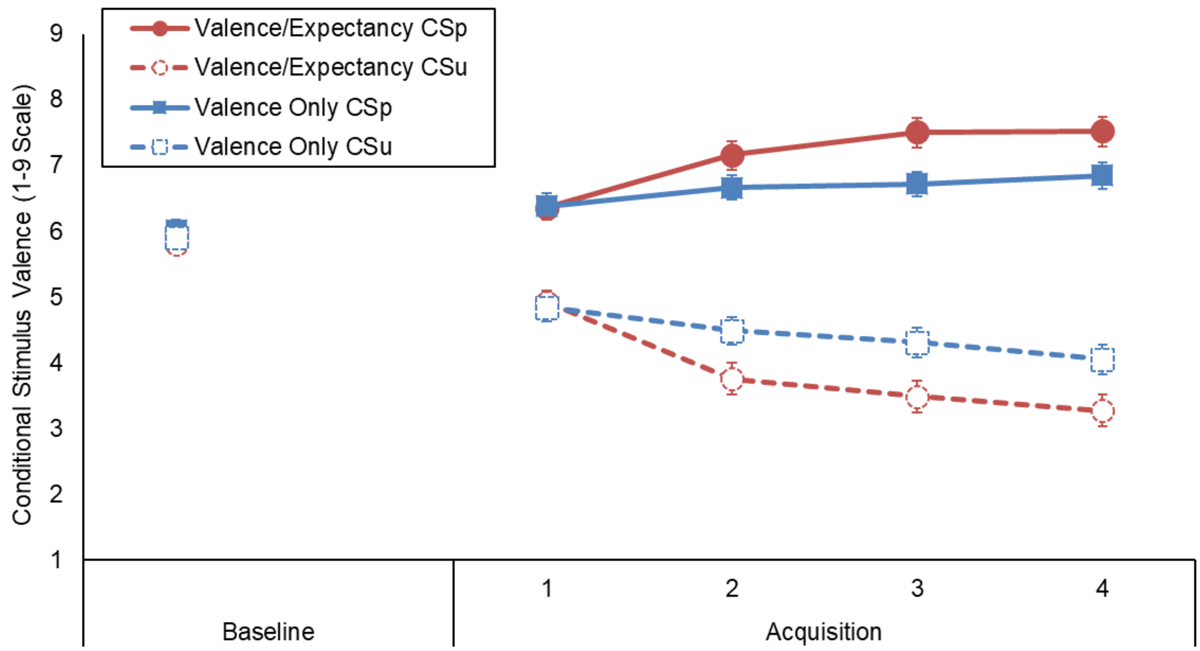

Figure 2. Conditional stimulus valence evaluations measured during baseline and acquisition of Experiment 2. Higher values indicate higher stimulus pleasantness. Error bars represent SEMs for within-participant designs based on O'Brien and Cousineau (2014).

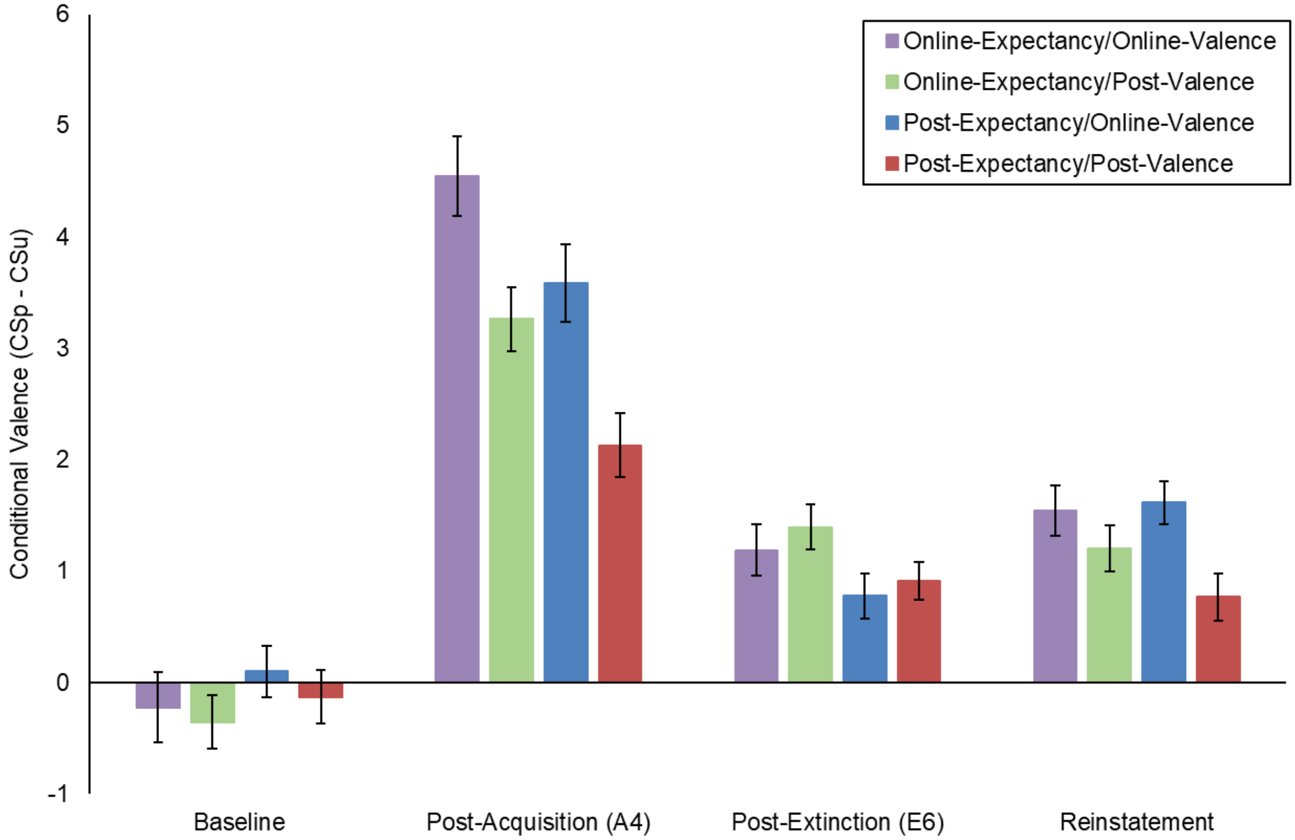

Figure 3. Conditional valence scores (CSp – CSu difference scores) measured during baseline, post-acquisition, post-extinction, and reinstatement of Experiment 3. Positive values indicate the expected pattern of conditional valence based on the stimulus contingencies. Error bars represent SEMs for within-participant designs based on O'Brien and Cousineau (2014).