Is Evaluative Conditioning Really Resistant to Extinction? Evidence for Changes in Evaluative Judgments without Changes in Evaluative Representations

Bertram Gawronski
University of Western Ontario, Canada

Anne Gast
University of Cologne, Germany

Jan De Houwer
Ghent University, Belgium

Evaluative conditioning (EC) is defined as the change in the evaluation of a conditioned stimulus (CS) due to its pairing with a positive or negative unconditioned stimulus (US). Although several individual studies suggest that EC is unaffected by unreinforced presentations of the CS without the US, a recent meta-analysis indicates that EC effects are less pronounced for post-extinction measurements than post-acquisition measurements. The disparity in research findings suggests that extinction of EC may depend on yet unidentified conditions. In an attempt to uncover these conditions, three experiments (N = 784) investigated the influence of unreinforced post-acquisition CS presentations on EC effects resulting from simultaneous versus sequential pairings and pairings with single versus multiple USs. For all four types of CS-US pairings, EC effects on self-reported evaluations were reduced by unreinforced CS presentations, but only when the CSs had been rated after the initial presentation of CS-US pairings. EC effects on an evaluative priming measure remained unaffected by unreinforced CS presentations regardless of whether the CSs had been rated after acquisition. The results suggest that reduced EC effects resulting from unreinforced CS presentations are due to judgment-related processes during the verbal expression of CS evaluations rather than genuine changes in the underlying evaluative representations.

Keywords: associative learning; evaluative conditioning; evaluative judgment; extinction

When an object repeatedly co-occurs with a positive or negative stimulus, the object tends to acquire the valence of the co-occurring stimulus. For example, many commercial advertisements rely on the idea that repeated pairings of a consumer product with a pleasant stimulus (e.g., depictions of a car with an attractive person) enhance consumers' liking of the product, thereby increasing the likelihood that they will actually buy it. Conversely, many health campaigns involve pairings of unhealthy products with unpleasant stimuli (e.g., graphic images on cigarette packages), which is assumed to reduce people's liking of these products, and thereby their consumption. In research on associative learning, such transfer effects are prominently captured by the concept of *evaluative conditioning* (EC), which is defined as the change in the evaluation of a conditioned stimulus (CS) due to its pairing with a valenced unconditioned stimulus (US) (De Houwer, 2007).

In addition to its value for various applied areas, EC represents a fascinating topic for basic research on human learning. What makes EC particularly interesting for learning theorists is that it has been claimed to have unique properties that distinguish it from other forms of conditioning (De Houwer, Thomas, & Baeyens, 2001; Walther, Nagengast, & Trasselli, 2005). One of these properties is resistance to extinction. Whereas most conditioned responses are attenuated by subsequent unreinforced presentations of the CS without the US, several studies have shown that EC effects are unaffected by unreinforced CS presentations (e.g., Baeyens, Crombez, Van den Bergh, & Eelen, 1988; Díaz, Ruiz, & Baeyens, 2005; Dwyer, Jarrat, & Dick, 2007; Vansteenwegen, Francken, Vervliet, De Clercq, & Eelen, 2006). These findings have fundamental implications for both basic and applied research. On the one hand, they impose major constraints on theories of the mental processes and representations underlying EC (e.g., Baeyens, Eelen, Crombez, & Van den Bergh, 1992; Field & Davey, 1999; Gawronski & Bodenhausen, 2006; Jones, Fazio, & Olson, 2009; Martin & Levey, 1978; Mitchell, De Houwer, & Lovibond, 2009; for a review, see Jones, Olson, & Fazio, 2010), in that these theories have to explain why EC is resistant to extinction. On the other hand, they are extremely interesting for applications of EC in real-world settings, in that EC effects seem to be unaffected by individual encounters of the relevant target objects without further reinforcement (e.g., repeated encounters of a consumer product in a store without the pleasant stimulus of the advertisement).

However, counter to the widespread assumption that EC is resistant to extinction, a recent meta-analysis by Hofmann, De Houwer, Perugini, Baeyens, and Crombez (2010) found that EC effects were less pronounced for post-extinction measurements than post-acquisition measurements. Hofmann et al. argued that previous failures to identify such reductions may be due to low statistical power of individual studies, which is overcome when the available data are aggregated across studies. Thus, although extinction may occur at slower rates for EC compared to other forms of conditioning, the claim that EC is resistant to extinction seems questionable on the basis of Hofmann et al.'s meta-analytic findings (see also Gawronski & Mitchell, 2014).

Another possibility is that specific features of the CS-US pairings determine whether extinction does or does not occur. In line with this contention, it has been argued that the functional properties of EC—such as its resistance to extinction—may depend on various procedural aspects of how a CS is paired with a US (e.g., De Houwer, 2007; Gast, Gawronski, & De Houwer, 2012). For example, whereas EC effects resulting from pairings with a single US have been shown to be reversed as a result of subsequent changes in the valence of the US, EC effects resulting from pairings with multiple USs of the same valence seem to be unaffected by US revaluation (Sweldens, Van Osselaer, & Janiszewski, 2010; see also Walther, Gawronski, Blank, & Langer, 2009). Moreover, research by Hütter and Sweldens (2012) suggests that, whereas EC effects resulting from simultaneous CS-US pairings can occur without recollective memory for these pairings, EC effects resulting from sequential pairings seem to require recollective memory. Because these findings may reflect differences in the underlying mental representations, it is possible that EC effects resulting from certain kinds of CS-US pairings are resistant to extinction, whereas EC effects resulting from other kinds of pairings are attenuated by unreinforced CS presentations. In line with this contention, it has been argued that EC effects can be due to either (1) the formation of a mental link between the CS and the US, a process that has been referred to as stimulus-stimulus (S-S) learning (Fulcher & Cocks, 1997), referential learning (Baeyens, Eelen, Van den Bergh, & Crombez, 1992), or indirect attitude transfer (Sweldens et al., 2010), or (2) the formation of a mental link between the CS and an evaluative response, a process that has been referred to as stimulus-response (S-R) learning (Fulcher & Cocks, 1997), intrinsic learning (Baeyens et al., 1992), or direct attitude change (Sweldens et al., 2010). Thus, to the extent that (1) procedural factors of the pairings influence the type of representation that is formed in response to CS-US pairings and (2) these representations differ in their resistance to extinction, procedural aspects of the CS-US pairings may be an important factor in the resistance of EC to extinction.

Demonstrating the relevance of such procedural moderators would impose further constraints on theories of the mental processes and representations underlying EC, because these theories should also explain why the functional properties of EC depend on the identified conditions. Thus, by investigating the conditions under which EC is resistant to extinction, the current research does not follow from a specific theoretical viewpoint, but rather from a more general meta-conditional approach to studying the functional properties of EC (De Houwer, 2007). The basic idea underlying this approach is that EC effects may be characterized by different functional properties depending on specific characteristics of the CS-US pairings. Toward this end, we conducted three experiments to investigate the influence of unreinforced CS presentations on EC effects that resulted from simultaneous versus sequential pairings and pairings with single versus multiple USs. Experiment 1 investigated effects of unreinforced CS presentations on self-reported CS evaluations using within-subjects comparisons to identify potential differences between post-acquisition and post-extinction measurements. Experiment 2 tested between-subjects differences between post-acquisition and post-extinction measurements, additionally including an evaluative priming task (Fazio, Jackson, Dunton, & Williams, 1995) to obtain an unobtrusive measure of evaluation. Finally, Experiment 3 investigated post-extinction EC effects on self-reported evaluations and an evaluative priming measure as a function of whether participants completed post-acquisition ratings of the CSs.[1]

Experiment 1

Participants in Experiment 1 were presented with CS-US pairings involving either simultaneous or sequential pairings. In addition, we manipulated whether a given CS was paired with a single US or multiple USs of the same valence. After the presentation of the CS-US pairings, the CSs were presented alone without further reinforcement. Participants were asked to rate their feelings toward the CSs after the initial presentation of the CS-US pairings and a second time after unreinforced CS presentations.

Method

Participants and design. Two-hundred undergraduate students (138 women, 62 men) at the University of Western Ontario were recruited for a one-hour battery that included the current experiment and two unrelated studies. Participants completed the current study as the second one in this battery. Participants received research credit for an introductory psychology course. The study included a 2 (US Valence: positive vs. negative) × 2 (Time of Measurement: post-acquisition vs. post-extinction) × 2 (Number of USs: single vs. multiple) × 2 (Pairing Mode: simultaneous vs. sequential) mixed-model design with the first two variables as within-subjects factors and the last two as between-subjects factors.

[1] For all three experiments, we report all measures, all conditions, and all data exclusions. The predetermined sample size for Experiment 1 was 200, providing samples of 50 participants for each of the four types of CS-US pairings. The intended sample size in Experiments 2 and 3 was approximately 300 based on participant availability.

Materials. As CSs we adapted five computer-generated images of shapes with different color patterns from Gawronski, Balas, and Creighton (2014). Two of these images were paired with positive USs; two were paired with negative USs; one was not paired with a valenced picture to provide a neutral baseline. As USs, we used 16 positive and 16 negative images from various sources, including the International Affective Picture System (Lang, Bradley, & Cuthbert, 2008) and Google internet searches.[2]

CS-US pairings. The presentation of the CS-US pairings was introduced as a visual perception task (see Gawronski et al., 2014; Gawronski & Mitchell, 2014). Participants were told that they would be presented with various pictures on the screen, including computer-generated drawings and real-world photographs. Participants were further told that we would ask them a number of questions about the pictures later in the study and that they should pay close attention throughout the task. The procedural parameters of the CS-US pairings (e.g., presentation times, inter-trial intervals, etc.) were based on earlier research by Gawronski and Mitchell (2014) who found evidence for extinction of EC in a paradigm using sequential pairings with a single US. The trials in the current study included 8 presentations of each CS-US pair and the neutral baseline CS, summing up to a total of 40 trials. Each trial started with a fixation cross for 250 ms in the center of the screen. In the sequential pairing condition, the fixation cross was followed by the CS for 1000 ms, which was replaced by the US for 1000 ms. Both images were displayed in the center of the screen. For the neutral baseline CS, the screen turned blank for 1000 ms after the presentation of the CS. In the simultaneous pairing condition, the CS and the US were presented simultaneously on the screen for 1000 ms. On half of the trials, the CS was presented on the left side and the US on the right side. On the remaining half, the position of the CS and the US was reversed. The neutral baseline CS appeared individually on either the left or the right side of the screen. The inter-trial interval was 1500 ms in both presentation mode conditions. For half of the participants, each CS was paired with the same US on all trials of the task. For the remaining half, each CS was paired with 8 different USs of the same valence. Both the CSs and the USs were presented in a size of 300 × 255 pixels on a 1280 × 1024 monitor. The use of a given CS for pairings with positive USs, negative USs, or no US was counterbalanced by means of a Latin square.

Unreinforced CS presentations. The procedure of the unreinforced CS presentations was similar to the presentation of the CS-US pairings, the only difference being that the screen remained blank where the USs had been presented before (see Gawronski & Mitchell, 2014). The presentation times and number of trials were identical to the presentation of the CS-US pairings.

Measures. To measure participants' evaluations of the CSs, they were asked to rate how pleasant or unpleasant each image made them feel on 7-point scales ranging from 1 (*very unpleasant*) to 7 (*very pleasant*). The evaluation measure was administered twice, once after the presentation of the CS-US pairings and once after the unreinforced CS presentations.

Results

To obtain baseline-corrected scores of self-reported CS evaluations, participants' ratings of the neutral baseline CS were subtracted from their ratings of each of the four CSs that had been paired with a positive or negative US (see Gawronski et al., 2014). Thus, higher values indicate more favorable evaluations of the CS compared to baseline. The resulting difference scores were then aggregated by averaging the baseline-corrected scores of the two CSs that had been paired with a US of the same valence before and after the presentation of the unreinforced trials (Cronbach's αs = .65 for post-acquisition positive, .70 for post-extinction positive, .68 for post-acquisition negative, and .76 for post-extinction negative).[3]

Submitted to a 2 (US Valence) × 2 (Time of Measurement) × 2 (Number of USs) × 2 (Pairing Mode) mixed-model ANOVA, CS evaluations revealed a significant main effect of US Valence, $F(1, 196) = 146.70$, $p < .001$, $\eta_p^2 = .428$, indicating that CSs that had been paired with positive USs were evaluated more favorably than CSs that had been paired with negative USs. This main effect was qualified by a significant two-way interaction of US Valence and Time of Measurement, $F(1, 196) = 65.16$, $p < .001$, $\eta_p^2 = .250$, indicating that EC effects were more pronounced for post-acquisition measurements, $F(1, 196) = 197.43$, $p < .001$, $\eta_p^2 = .502$, than post-extinction measurements, $F(1, 196) = 54.47$, $p < .001$, $\eta_p^2 = .217$. The two-way interaction of US Valence and Time of Measurement was statistically significant for all four kinds of CS-US pairings (all Fs > 8.40, all ps < .006), indicating that

[2] All materials are available from the authors upon request.

[3] Analyses using uncorrected raw scores revealed the same pattern of results. For the sake of consistency between studies and measures, we report baseline-corrected scores for all three experiments, because the evaluative priming measures in Experiments 2 and 3 require appropriate corrections to reduce measurement error resulting from baseline differences in responses to positive versus negative target words (see Wentura & Degner, 2010).

unreinforced CS presentations reduced EC effects regardless of whether they were due to simultaneous versus sequential pairings or pairings with single versus multiple USs (see Figure 1). Yet, despite the observed reductions as a result of unreinforced CS presentations, EC effects on post-extinction measurements remained statistically significant in all four conditions (all ts > 2.85, all ps < .007).

In addition to these effects, the ANOVA revealed a significant two-way interaction of US Valence and Number of USs, $F(1, 196) = 3.89$, $p = .05$, $\eta_p^2 = .019$, indicating that EC effects were more pronounced when the CSs had been paired with multiple USs of the same valence, $F(1, 98) = 98.16$, $p < .001$, $\eta_p^2 = .500$, than when they had been paired with a single US, $F(1, 98) = 51.95$, $p < .001$, $\eta_p^2 = .346$. More important for the current investigation, there was a significant three-way interaction of US Valence, Number of USs, and Time of Measurement, $F(1, 196) = 5.22$, $p = .02$, $\eta_p^2 = .026$. This interaction indicated that extinction effects were more pronounced for EC effects that resulted from pairings with multiple USs, $F(1, 98) = 46.72$, $p < .001$, $\eta_p^2 = .323$, compared with EC effects that resulted from pairings with a single US, $F(1, 98) = 19.66$, $p < .001$, $\eta_p^2 = .167$, although extinction effects were statistically significant in both conditions. Further inspection of the data revealed that the obtained difference in extinction effects was driven by significantly larger EC effects on post-acquisition measurements for pairings with multiple USs compared with pairings for single USs, $F(1, 196) = 7.36$, $p = .007$, $\eta_p^2 = .036$, whereas EC effects on post-extinction measurements did not differ for the two kinds of pairings, $F(1, 196) = 0.56$, $p = .45$, $\eta_p^2 = .003$ (see Figure 1). No other effects involving US Valence reached statistical significance (all Fs < 1, all ps > .40).

Discussion

The results of Experiment 1 suggest that EC effects are not entirely resistant to extinction. Although EC effects were not fully attenuated after unreinforced presentations of the CSs, EC effects were more pronounced for post-acquisition measurements than post-extinction measurements. This reduction generalized across various kinds of CS-US pairings, in that EC effects were reduced by unreinforced CS presentations regardless of whether they were due to sequential versus simultaneous pairings or pairings with single versus multiple USs.

Experiment 2

Although the findings of Experiment 1 are consistent with the results of Hofmann et al.'s (2010) meta-analysis, a potential concern is that our study used within-subjects comparisons of post-acquisition and post-extinction measurements. Lipp and Purkis (2006) showed that such within-subjects comparisons can influence judgments on self-report measures, in that participants flexibly adjust the use of available information as a function of prior judgments. Specifically, Lipp and Purkis argued that prior judgments influence whether participants integrate all available information or instead rely on the most recent information to make an evaluative judgment. Because the second strategy is more likely when participants rated the same CSs before, a within-subjects manipulation of time of measurement could lead to reduced EC effects on self-report measures. However, such "extinction" effects might be the product of judgment-related processes during the verbal expression of CS evaluations (i.e., which information is used to make an evaluative rating) rather than genuine changes in the underlying representation of the CS. Thus, to investigate whether reduced EC effects in Experiment 1 are due to judgment-related processes or genuine changes in the underlying evaluative representations, Experiment 2 utilized a between-subjects comparison of post-acquisition and post-extinction measurements (cf. Díaz et al., 2005). In addition, we included an evaluative priming task (Fazio et al., 1995) to obtain an unobtrusive measure of evaluation. Because evaluative priming effects are inferred from response times on a speeded categorization task (rather than direct evaluative ratings), such scores are unaffected by judgment-related shifts in the use of available information (see Gawronski & De Houwer, 2014).

Method

Participants and design. Two-hundred-and-eighty-nine undergraduate students (211 women, 78 men) at the University of Western Ontario were recruited for a one-hour battery that included the current experiment and two unrelated studies. Participants completed the current study as the second one in this battery. One-hundred participants received \$10. One-hundred-and-eighty-nine participants received research credit for an introductory psychology course. The study included a 2 (US Valence: positive vs. negative) × 2 (Time of Measurement: post-acquisition vs. post-extinction) × 2 (Number of USs: single vs. multiple) × 2 (Pairing Mode: simultaneous vs. sequential) mixed-model design with the first variable as within-subjects factor and the other three as between-subjects factors. Data from one participant who showed random responses on the dichotomous categorization task of the evaluative priming measure (i.e., error rate of 49%) were excluded from the analysis.

Procedure. The procedure and all materials were identical to Experiment 1, the only differences being that (1) we added an evaluative priming task as an unobtrusive measure of evaluation, and (2) time of measurement was manipulated between-subjects rather than within-subjects. Half of the participants completed the two evaluation measures after the presentation of the CS-US pairings; the remaining half completed the two evaluation measures after the presentation of unreinforced CS presentations. The order of the two evaluation measures was counterbalanced across conditions.

Measures. The measure of self-reported CS evaluations was identical to Experiment 1. In addition, participants were asked to complete an evaluative priming task that included the CSs as primes and positive and negative adjectives as targets. The procedural details of the evaluative priming task were adapted from an earlier EC study by Gawronski et al. (2014). Each trial started with a fixation cross that was displayed for 500 ms in the center of the screen. The fixation cross was followed by a prime stimulus, which was replaced by the target word after 200 ms. Participants' task was to press a right-hand key (*Numpad 5*) as quickly as possible when the target word was positive and a left-hand key (*A*) when the target word was negative. The target words remained on the screen until participants made their response. Incorrect responses were followed by the word *ERROR!* for 1500 msec. The inter-trial interval was 500 ms. The positive target words were: *pleasant*, *good*, *outstanding*, *beautiful*, *magnificent*, *marvelous*, *excellent*, *appealing*, *delightful*, *nice*; the negative target words were: *unpleasant*, *bad*, *horrible*, *miserable*, *hideous*, *dreadful*, *painful*, *repulsive*, *awful*, *ugly*. Each CS was presented once with each of the 10 positive target words and once with each of the 10 negative words, summing up to a total of 100 trials.

Results

Evaluative ratings. Baseline-corrected scores of self-reported evaluations were calculated according to the procedures in Experiment 1. The resulting difference scores were aggregated by averaging the baseline-corrected scores of the two CSs that had been paired with a US of the same valence (Cronbach's α = .67 and .75, respectively). Means and standard deviations of self-reported evaluations are presented in Table 1. A 2 (US Valence) × 2 (Time of Measurement) × 2 (Number of USs) × 2 (Pairing Mode) mixed-model ANOVA on these scores revealed a significant main effect of US Valence, $F(1, 280) = 215.33$, $p < .001$, $\eta_p^2 = .435$, indicating that CSs that had been paired with positive USs were evaluated more favorably than CSs that had been paired with negative USs (*M*s = 1.07 vs. -0.78). This main effect was qualified by a significant two-way interaction of US Valence and Number of USs, $F(1, 280) = 13.55$, $p < .001$, $\eta_p^2 = .046$, indicating that EC effects were again more pronounced when the CSs had been paired with multiple USs of the same valence (*M*s = 1.34 vs. -0.94), $F(1, 139) = 153.01$, $p < .001$, $\eta_p^2 = .524$, than when they had been paired with a single US (*M*s = 0.79 vs. -0.59), $F(1, 141) = 67.01$, $p < .001$, $\eta_p^2 = .322$. Importantly, US Valence did not show any significant interactions with Time of Measurement (all *F*s < 1.23, all *p*s > .28), indicating that EC effects on self-reported evaluations remained unaffected by unreinforced CS presentations. The two-way interaction of US Valence and Time of Measurement failed to reach statistical significance for any of the four kinds of CS-US pairings (all *F*s < 1.15, all *p*s > .28).

Evaluative priming. Before we aggregated the response latency data of the evaluative priming task, we excluded latencies from trials with incorrect responses (5.8%) and truncated latencies higher than 1500 ms (2.4%) (see Gawronski, Bodenhausen, & Becker, 2007). Following the scoring procedure by Fazio et al. (1995), a priming index was calculated for each CS, reflecting the positivity versus negativity of the response elicited by the CS compared to baseline. First, a positivity index was calculated for each CS that had been paired with a US by subtracting the mean response latency to positive target words preceded by the CS from the mean response latency to positive target words preceded by the neutral baseline CS. This index reflects the extent to which a given CS facilitates responses to positive target words, which can be interpreted as an index of the positivity of the response elicited by the prime (Wentura & Degner, 2010). Second, a negativity index was calculated for each CS that had been paired with a US by subtracting the mean response latency to negative target words preceded by the CS from the mean response latency to negative target words preceded by the neutral baseline CS. This index reflects the extent to which a given CS facilitates responses to negative target words, which can be interpreted as an index of the negativity of the response elicited by the prime (Wentura & Degner, 2010). To obtain a single priming index of the evaluative response elicited by a given CS, the negativity scores of each CS were subtracted from the positivity scores of the same CS. Thus, higher values on this priming index indicate more favorable responses to the CS compared to baseline (Wentura & Degner, 2010).[4] Following the procedure by Gawronski et al. (2014), the resulting scores were aggregated by averaging the priming scores of the two CSs that had been paired with a US of the same valence (Cronbach's α = .77 and .72, respectively).

[4] Note that the overall size of this priming index is statistically equivalent the two-way interaction of prime valence and target valence using the four baseline-corrected difference scores (see Wentura & Degner, 2010).

Means and standard deviations of the priming scores are presented in Table 2. A 2 (US Valence) × 2 (Time of Measurement) × 2 (Number of USs) × 2 (Pairing Mode) mixed-model ANOVA on these scores revealed a significant main effect of US Valence, $F(1, 280) = 17.81$, $p < .001$, $\eta_p^2 = .060$, indicating that CSs that had been paired with positive USs elicited more favorable responses than CSs that had been paired with negative USs (*M*s = 5.94 vs. -15.06). Replicating the absence of extinction effects on self-reported evaluations, US Valence did not show any significant interactions with Time of Measurement (all *F*s < 1.81, all *p*s > .18). The two-way interaction of US Valence and Time of Measurement failed to reach statistical significance for any of the four kinds of CS-US pairings (all *F*s < 1, all *p*s > .39), indicating that EC effects on the evaluative priming measure were generally unaffected by unreinforced CS presentations.

Discussion

Counter to the findings of Experiment 1, the current study failed to obtain significant reductions in EC effects resulting from unreinforced CS presentations. This resistance to extinction generalized across EC effects that resulted from simultaneous versus sequential pairings and pairings with single versus multiple USs. Moreover, there was no significant reduction in EC effects regardless of whether CS evaluations were measured by self-report or an evaluative priming task. Together with the findings of Experiment 1, these results suggest that the obtained reductions in EC effects are the result of judgment-related processes elicited by prior judgments of the CSs rather than genuine changes in the underlying evaluative representations.

Experiment 3

The main goal of Experiment 3 was to provide a more stringent test of our conclusion that the different patterns in Experiments 1 and 2 reflect systematic effects of judgment-related processes rather than incidental characteristics of the two studies. Toward this end, Experiment 3 compared EC effects on post-extinction measurements as a function of whether participants rated the CSs after the initial presentation of the CS-US pairings. Based on the results of Experiments 1 and 2, we expected post-extinction EC effects on self-reported evaluations to be smaller when participants rated the CSs after acquisition than when they did not rate the CSs after acquisition. Yet, post-extinction EC effects on the evaluative priming measure should be unaffected by prior ratings of the CSs.

Method

Participants and design. Two-hundred-and-ninety-five undergraduate students (186 women, 109 men) at the University of Western Ontario were recruited for a one-hour battery that included the current study and two unrelated studies. Participants completed the current study as the third one in this battery. One-hundred participants received $10. One-hundred-and-ninety-five participants received research credit for an introductory psychology course. The study included a 2 (US Valence: positive vs. negative) × 2 (Post-acquisition Rating: CSs vs. unfamiliar stimuli) × 2 (Number of USs: single vs. multiple) × 2 (Pairing Mode: simultaneous vs. sequential) mixed-model design with the first variable as within-subjects factor and the other three as between-subjects factors. Due to a computer malfunction, evaluative priming data from one participant were not recorded. In addition, we excluded data from one participant who showed close-to-random responses on the dichotomous categorization task of the evaluative priming measure (i.e., error rate of 33%).

Procedure. The procedure and all materials were identical to Experiment 2, the only differences being that (1) all participants completed the two evaluation measures after the unreinforced CS presentations, and (2) half of the participants were asked to rate the CSs after the initial presentation of CS-US pairings whereas the remaining half were asked to rate five unfamiliar stimuli from the same set of computer-generated drawings. Thus, the contextual conditions of post-extinction evaluations were equal in terms of time and number of post-acquisition ratings, differing only with regard to whether participants had rated the CSs or unfamiliar stimuli before.

Results

Evaluative ratings. Baseline-corrected scores of self-reported post-extinction evaluations were calculated according to the procedures in Experiment 1. The resulting difference scores were then aggregated by averaging the baseline-corrected scores of the two CSs that had been paired with a US of the same valence (Cronbach's $\alpha = .72$ and .77, respectively). Submitted to a 2 (US Valence) × 2 (Post-acquisition Rating) × 2 (Number of USs) × 2 (Pairing Mode) mixed-model ANOVA, these scores revealed a significant main effect of US Valence, $F(1, 285) = 101.41$, $p < .001$, $\eta_p^2 = .262$, indicating that CSs that had been paired with positive USs were evaluated more favorably than CSs that had been paired with negative USs (*M*s = 0.59 vs. -0.54). In addition, the ANOVA revealed a significant two-way interaction of US Valence and Number of USs, $F(1, 285) = 6.44$, $p = .01$, $\eta_p^2 = .022$, indicating that EC effects were again more pronounced when the CSs had been paired with multiple USs of the same valence, $F(1, 143) = 67.74$, $p < .001$, $\eta_p^2 = .321$, than when they had been paired with a single US, *F*(1,

142) = 34.36, $p < .001$, $\eta_p^2 = .195$. More important for the current investigation, the main effect of US Valence was qualified by a significant two-way interaction with Post-acquisition Rating, $F(1, 285) = 4.03$, $p = .046$, $\eta_p^2 = .014$ (see Figure 2, left panel). This interaction indicated that post-extinction EC effects were less pronounced when participants had rated the CSs after the presentation of the CS-US pairings, $F(1, 134) = 41.34$, $p < .001$, $\eta_p^2 = .236$, than when participants had rated unfamiliar stimuli after acquisition, $F(1, 151) = 63.07$, $p < .001$, $\eta_p^2 = .295$. The interaction of US Valence and Post-acquisition Rating was unqualified by higher-order interactions with Number of USs and Presentation Mode (all *F*s < 1, all *p*s > .61).

Evaluative priming. Before we aggregated the evaluative priming data, we excluded latencies from trials with incorrect responses (6.4%) and truncated latencies higher than 1500 ms (2.2%). Priming scores were calculated for each CS according to the procedure in Experiment 2. The resulting scores were aggregated by averaging the priming scores of the two CSs that had been paired with a US of the same valence (Cronbach's $\alpha = .65$ and .57, respectively). Submitted to a 2 (US Valence) × 2 (Post-acquisition Rating) × 2 (Number of USs) × 2 (Pairing Mode) mixed-model ANOVA, these scores revealed a theoretically uninteresting two-way interaction of Number of USs and Pairing Mode, $F(1, 285) = 3.92$, $p = .05$, $\eta_p^2 = .014$, indicating that, regardless of US Valence, CSs that had been paired with multiple USs elicited more favorable responses than CSs that had been paired with a single US for sequential pairings, $F(1, 140) = 6.14$, $p = .01$, $\eta_p^2 = .042$, but not for simultaneous pairings, $F(1, 145) = 0.06$, $p = .80$, $\eta_p^2 < .001$. More important for the current investigation, a significant main effect of US Valence indicated that CSs that had been paired with positive USs elicited more favorable responses than CSs that had been paired with negative USs, $F(1, 285) = 6.84$, $p = .009$, $\eta_p^2 = .060$ (see Figure 2, right panel). This main effect was not qualified by any significant interactions with Post-acquisition Rating (all *F*s < 1, all *p*s > .37), indicating that EC effects on the evaluative priming measure were unaffected by whether participants had rated the CSs prior to the extinction phase.

Discussion

Consistent with the proposed interpretation of reduced EC effects in terms of judgment-related processes, post-extinction EC effects on self-reported evaluations were less pronounced when participants had rated the CSs after the presentation of CS-US pairings than when they had rated unfamiliar stimuli before. However, post-extinction EC effects on an evaluative priming measure remained unaffected by prior ratings of the CSs. Together with the results of Experiments 1 and 2, these findings indicate that reduced EC effects resulting from unreinforced CS presentations are the product of judgment-related processes elicited by prior ratings of the CSs rather than genuine changes in the underlying evaluative representations.

General Discussion

The main goal of the current research was to gain deeper insights into the conditions under which EC is resistant to extinction. Counter to our initial speculation, we found no evidence for the hypothesis that extinction depends on specific features of CS-US pairings. Although this does not exclude the possibility that extinction effects depend on other procedural factors, the generality of the obtained results is important, because the procedural factors included in the current studies have been shown to impact other functional properties of EC, such as their dependence on recollective memory (e.g., Hütter & Sweldens, 2012) and susceptibility to US revaluation (e.g., Sweldens et al., 2010). In the current studies, extinction did not depend on whether EC effects were due to simultaneous versus sequential pairings or pairings with single versus multiple USs. Instead, the most important determinant of extinction was whether participants had rated the CSs before. Specifically, we found that unreinforced CS presentations reduced EC effects on self-reported evaluations only when participants had rated the CSs after the presentation of the CS-US pairings. When participants had not rated the CSs before, EC effects on self-reported evaluations were unaffected by unreinforced CS presentations. There was no influence of unreinforced CS presentations and prior ratings on EC effects measured with an evaluative priming task. Taken together, these results suggest that reduced EC effects resulting from unreinforced CS presentations are due to judgment-related processes elicited by prior ratings of the CSs rather than genuine changes in the underlying evaluative representations.

The current research makes an important contribution to the EC literature, because it qualifies one of the most significant conclusions of a recent meta-analysis on the functional properties of EC. Aggregating data from 35 years of research, Hofmann et al. (2010) found that EC effects were less pronounced for post-extinction measurements than post-acquisition measurements. Based on this finding, the authors concluded that EC is reduced by unreinforced presentations of the CS without the US, which stands in contrast to the findings of numerous studies suggesting that EC is resistant to extinction (e.g., Baeyens et al., 1988; Díaz et al., 2005; Dwyer et al., 2007; Vansteenwegen et al., 2006). To explain this inconsistency, Hofmann et al. argued that earlier failures to identify significant reductions in EC effects may be due to low statistical power of individual studies, which is overcome when the available data are aggregated across studies. The current findings suggest a different conclusion, in that reduced EC effects after unreinforced presentations of the CS are the result of judgment-related processes during the verbal expression of CS evaluations rather than genuine changes in the underlying evaluative representations. This conclusion is consistent with the results of several studies that failed to identify significant reductions in EC when extinction effects were investigated with between-subjects designs or unobtrusive measures of evaluation (e.g., Díaz et al., 2005; Vansteenwegen et al., 2006).

Nevertheless, the current findings are also consistent with research showing extinction effects in within-subjects designs using self-report measures. One example is a study by Lipp and Purkis (2006), who argued that prior ratings influence whether participants integrate all available information about the CSs or instead rely on the most recent information when making an evaluative judgment. Based on the assumption that the second strategy is more likely when participants rated the CSs before, Lipp and Purkis proposed that previously undetected extinction effects can be uncovered by including a measure of self-reported evaluations prior to unreinforced CS presentations. Consistent with this assumption, the authors found reduced EC effects when extinction was investigated on a within-subjects basis, but not when it was tested on a between-subjects basis. However, the current studies suggest a different interpretation of this finding, in that reduced EC effects in within-subjects designs do not signify genuine changes in the underlying evaluative representations, but instead reflect adjustments in the verbal expression of CS evaluations. This conclusion is supported by the finding that EC effects on the evaluative priming measure remained unaffected by unreinforced presentations of the CSs regardless of whether participants had rated the CSs before. Thus, counter to Lipp and Purkis's conclusion, the present work corroborates the hypothesis that EC is indeed distinct from other forms of conditioning (see De Houwer et al., 2001; Walther et al., 2005). Whereas most conditioned responses are attenuated by subsequent unreinforced presentations of the CS without the US, repeated pairings of a CS with a valenced US seem to create evaluative representations of the CS that are resistant to extinction.

In addition to providing deeper insights into the processes and representations underlying extinction effects in EC, another interesting aspect of the current findings is that pairings with multiple USs of the same valence produced larger EC effects than pairings with a single USs. This pattern stands in contrast to earlier findings by Stahl and Unkelbach (2009) who found larger EC effects for pairings with single USs than multiple USs. A similar inconsistency occurred for the effect of sequential pairings with multiple USs of the same valence, which produced significant EC effects in the current experiments, although earlier research by Sweldens et al. (2010) failed to obtain significant EC effects for such kinds of CS-US pairings. We believe that the inconsistent outcomes are most likely due to differences in the procedural parameters of the CS-US pairings (cf. Gast et al., 2012). For example, whereas sequential pairings in Sweldens et al.'s studies included an interval of 500 ms between the CS and the US, sequential pairings in the current research involved immediate successions of the two stimuli without any delay. Thus, it is possible that sequential pairings with multiple USs produce EC effects only for immediate, but not delayed, CS-US successions. This speculation is consistent with De Houwer's (2007) argument that many inconsistencies in the EC literature may stem from differences in the procedural parameters of the CS-US pairings, which may moderate the contribution of functionally distinct processes to EC effects (see also Sweldens, Corneille, & Yzerbyt, 2014). Future research may help to further clarify the moderators of EC effects resulting from simultaneous versus sequential pairings with single versus multiple USs as well as their underlying mental processes. Although the current findings generalized across different types of CS-US pairings, these investigations may also explore the moderating role of other procedural parameters on the obtained resistance to extinction.

A potential concern about our main conclusion is that it is partially based on a null effect of unreinforced CS presentations on EC effects measured with an evaluative priming task. Because priming measures often show low reliability (Gawronski & De Houwer, 2014), it is possible that the obtained absence of extinction effects on the priming measure is due to measurement error rather than genuine resistance to extinction. In response to this concern, it is worth noting that the priming measure in the current research showed (1) significant effects of our EC manipulation and (2) reliability estimates that were comparable to the ones of the self-report measure (see also Gawronski et al., 2014). Moreover, even self-reported evaluations were unaffected by unreinforced CS presentations when post-acquisition and post-extinction measurements were compared in a between-subjects design. Together, these results suggest that low reliability of the priming measure does not account for the obtained pattern of results. Nevertheless, future research may provide further evidence for our conclusions by showing similar effects on unobtrusive measures with higher reliability (cf. Gawronski & De Houwer, 2014).

The finding that the evaluative representations resulting from repeated CS-US pairings are resistant to extinction has important implications for basic and applied research on EC. Counter to the widespread assumption that EC is different from other forms of conditioning (De Houwer et al., 2001; Walther et al., 2005), some researchers have raised doubts as to whether this conclusion is justified on the basis of the currently available evidence (e.g., Mitchell et al., 2009). One example in this regard is the claim that EC differs from other forms of conditioning in terms of its resistance to extinction. This claim has been challenged by Hofmann et al.'s (2010) meta-analysis, showing that EC effects were less pronounced for post-extinction measurements than post-acquisition measurements (see also Gawronski & Mitchell, 2014). The current experiments suggest that such reductions are most likely due to judgment-related processes elicited by repeated ratings of the CSs rather than genuine changes in the underlying evaluative representations. As such, our findings corroborate earlier claims that EC differs from other forms of conditioning, in that repeated pairings of a CS with a valenced US create evaluative representations of the CS that are resistant to extinction. This conclusion imposes significant constraints on theories of the mental processes and representations underlying EC (e.g., Baeyens et al., 1992; Field & Davey, 1999; Gawronski & Bodenhausen, 2006; Jones et al., 2009; Martin & Levey, 1978; Mitchell et al., 2009; for a review, see Jones et al., 2010), which have to explain why EC is resistant to extinction. In addition, our findings have important implications for applications of EC in real-world settings, in that EC effects seem to be unaffected by encounters of a CS without the US. For example, if a consumer product is repeatedly paired with a pleasant stimulus in a commercial advertisement, subsequent encounters of that product without the pleasant stimulus (e.g., in a store) would have the potential to attenuate the initial effect of the advertisement to the extent that EC effects are reduced by unreinforced presentations of the CS. The current findings suggest that such reductions are in fact unlikely, in that subsequent encounters of the consumer product without the pleasant stimulus do not change the positive representation that has been created by the advertisement. Although unreinforced CS presentations may reduce EC effects on self-reported evaluations as a result of judgment-related processes, the underlying evaluative representations seem to be resistant to extinction.

References

Baeyens, F., Crombez, G., Van den Bergh, O., & Eelen, P. (1988). Once in contact, always in contact: Evaluative conditioning is resistant to extinction. *Advances in Behaviour Research and Therapy, 10*, 179-199.

Baeyens, F., Eelen, P., Crombez, G., & Van den Bergh, O. (1992). Human evaluative conditioning: Acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behaviour Research and Therapy, 30*, 133-142.

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1992). The content of learning in human evaluative conditioning: Acquired valence is sensitive to US-revaluation. *Learning and Motivation, 23*, 200-224.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology, 10*, 230-241.

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

Díaz, E., Ruiz, G., & Baeyens, F. (2005). Resistance to extinction of human evaluative conditioning using a between-subjects design. *Cognition and Emotion, 19*, 245–268.

Dwyer, D. M., Jarratt, F., & Dick, K. (2007). Evaluative conditioning with foods as CSs and body shapes as USs: No evidence for sex differences, extinction, or overshadowing. *Cognition and Emotion, 21*, 281–299.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013-1027.

Field, A. P., & Davey, G. C. L. (1999). Reevaluating evaluative conditioning: A nonassociative explanation of conditioning effects in the visual evaluative conditioning paradigm. *Journal of Experimental Psychology: Animal Behavior Processes, 25*, 211-224.

Fulcher, E. P., & Cocks, R. P. (1997). Dissociative storage systems in human evaluative conditioning. *Behaviour Research and Therapy, 35*, 1-10.

Gast, A., Gawronski, B., & De Houwer, J. (2012). Evaluative conditioning: Recent developments and future directions. *Learning and Motivation, 43*, 79-88.

Gawronski, B., Balas, R., & Creighton, L. A. (2014). Can the formation of conditioned attitudes be intentionally controlled? *Personality and Social Psychology Bulletin, 40*, 419-432.

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.

Gawronski, B., Bodenhausen, G. V., & Becker, A. P. (2007). I like it, because I like myself: Associative self-anchoring and post-decisional change of implicit evaluations. *Journal of Experimental Social Psychology, 43*, 221-232.

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 edition, pp. 283-310). New York, NY: Cambridge University Press.

Gawronski, B., & Mitchell, D. G. V. (2014). Simultaneous conditioning of valence and arousal. *Cognition and Emotion, 28*, 577-595.

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

Hütter, M., & Sweldens, S. (2012). Implicit misattribution of evaluative responses: Contingency-unaware evaluative conditioning requires simultaneous stimulus presentations. *Journal of Experimental Psychology: General, 142*, 638-643.

Jones, C. R., Fazio, R. H., & Olson, M. A. (2009). Implicit misattribution as a mechanism underlying evaluative conditioning. *Journal of Personality and Social Psychology*, 96, 933-948.

Jones, C. R., Olson, M. A., & Fazio, R. H. (2010). Evaluative conditioning: The “how” question. *Advances in Experimental Social Psychology, 43*, 205-255.

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-7*. University of Florida, Gainesville, FL.

Lipp, O. V., & Purkis, H. M. (2006). The effects of assessment type on verbal ratings of conditional stimulus valence and contingency judgments: Implications for the extinction of evaluative conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 32,* 431-440.

Martin, I., & Levey, A. B. (1978). Evaluative conditioning. *Advances in Behaviour Research and Therapy, 1,* 57-102.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32*, 183-198.

Stahl, C., & Unkelbach, C. (2009). Evaluative learning with single versus multiple unconditioned stimuli: The role of contingency awareness. *Journal of Experimental Psychology: Animal Behavior Processes, 35,* 286–291.

Sweldens, S., Corneille, O., & Yzerbyt, V. (2014). The role of awareness in attitude formation through evaluative conditioning. *Personality and Social Psychology Review, 18,* 187-209.

Sweldens, S., Van Osselaer, S., & Janiszewski, C. (2010). Evaluative conditioning procedures and the resilience of conditioned brand attitudes. *Journal of Consumer Research, 37,* 473-489.

Vansteenwegen, D., Francken, G., Vervliet, B., De Clercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology. Animal Behavior Processes, 32,* 71-79.

Walther, E., Gawronski, B., Blank, H., & Langer, T. (2009). Changing likes and dislikes through the backdoor. The US revaluation effect. *Cognition and Emotion, 23,* 889-917.

Walther, E., Nagengast, B., & Trasselli, C. (2005). Evaluative conditioning in social psychology: Facts and speculations. *Cognition and Emotion, 19,* 175-196.

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95-116). New York: Guilford Press.

Author Note

The research reported in this article has been supported by grants from the Canada Research Chairs Program (215983), the Natural Sciences and Engineering Research Council of Canada (341601-2013), and the Research Council of Ghent University (VBO 011/01T01111) to the first author; and by Ghent University grant BOF09/01M00209 and the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office (IUAPVII/33) to the last author. Part of this research was conducted while the first author was a visiting scholar at Ghent University and the second author was a postdoctoral fellow at Ghent University. We thank Jasmine Desjardins, Nicole Dryburgh, Elizabeth Harris, Grace Ki, Nicole Miles, and Ting Zhao for their help in collecting the data.

Correspondence concerning this article should be sent to Bertram Gawronski, who is now at the Department of Psychology, University of Texas at Austin, 108 E. Dean Keeton A8000, Austin, TX 78712-1043, USA, email: gawronski@utexas.edu.

Table 1. Means and standard deviations of baseline-corrected CS evaluations on an evaluative rating measure as a function of US valence (positive vs. negative), number of USs (single vs. multiple), mode of CS-US pairings (simultaneous vs. sequential), and time of measurement (post-acquisition vs. post-extinction) using a between-subjects manipulation of measurement time, Experiment 2. Higher values indicate more favorable evaluations.

	Post-acquisition		Post-extinction	
	M	*SD*	*M*	*SD*
Single US – Simultaneous Pairings				
positive	1.25	1.39	0.69	1.53
negative	-0.44	1.77	-0.57	1.81
Single US – Sequential Pairings				
positive	0.59	2.05	0.63	1.69
negative	-0.64	1.29	-0.72	1.57
Multiple US – Simultaneous Pairings				
positive	1.14	1.42	1.81	1.67
negative	-1.45	1.83	-0.66	1.75
Multiple US – Sequential Pairings				
positive	1.41	1.75	1.09	1.73
negative	-1.02	2.05	-0.66	2.23

Table 2. Means and standard deviations of baseline-corrected CS evaluations on an evaluative priming measure as a function of US valence (positive vs. negative), number of USs (single vs. multiple), mode of CS-US pairings (simultaneous vs. sequential), and time of measurement (post-acquisition vs. post-extinction) using a between-subjects manipulation of measurement time, Experiment 2. Higher values indicate more favorable evaluations.

	Post-acquisition		Post-extinction	
	M	*SD*	*M*	*SD*
Single US – Simultaneous Pairings				
positive	4	115	-18	106
negative	-15	83	-30	89
Single US – Sequential Pairings				
positive	10	107	-1	130
negative	12	126	-17	96
Multiple US – Simultaneous Pairings				
positive	23	94	13	87
negative	-8	80	-30	110
Multiple US – Sequential Pairings				
positive	14	94	3	130
negative	-18	95	-13	117

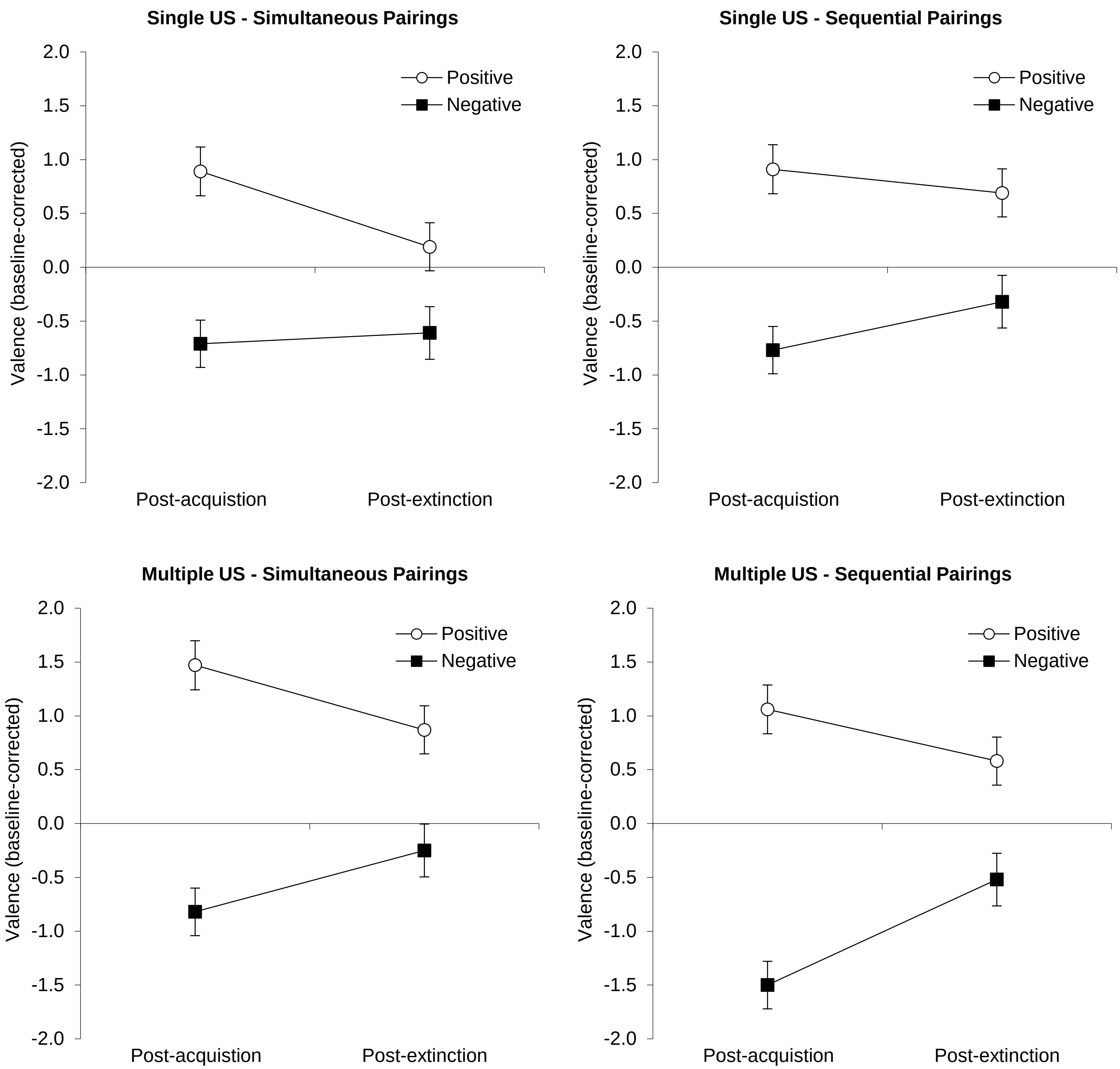

Figure 1. Baseline-corrected CS evaluations on an evaluative rating measure as a function of US valence (positive vs. negative), number of USs (single vs. multiple), mode of CS-US pairings (simultaneous vs. sequential), and time of measurement (post-acquisition vs. post-extinction) using a within-subjects manipulation of measurement time, Experiment 1. Higher values indicate more favorable evaluations. Error bars depict standard errors.

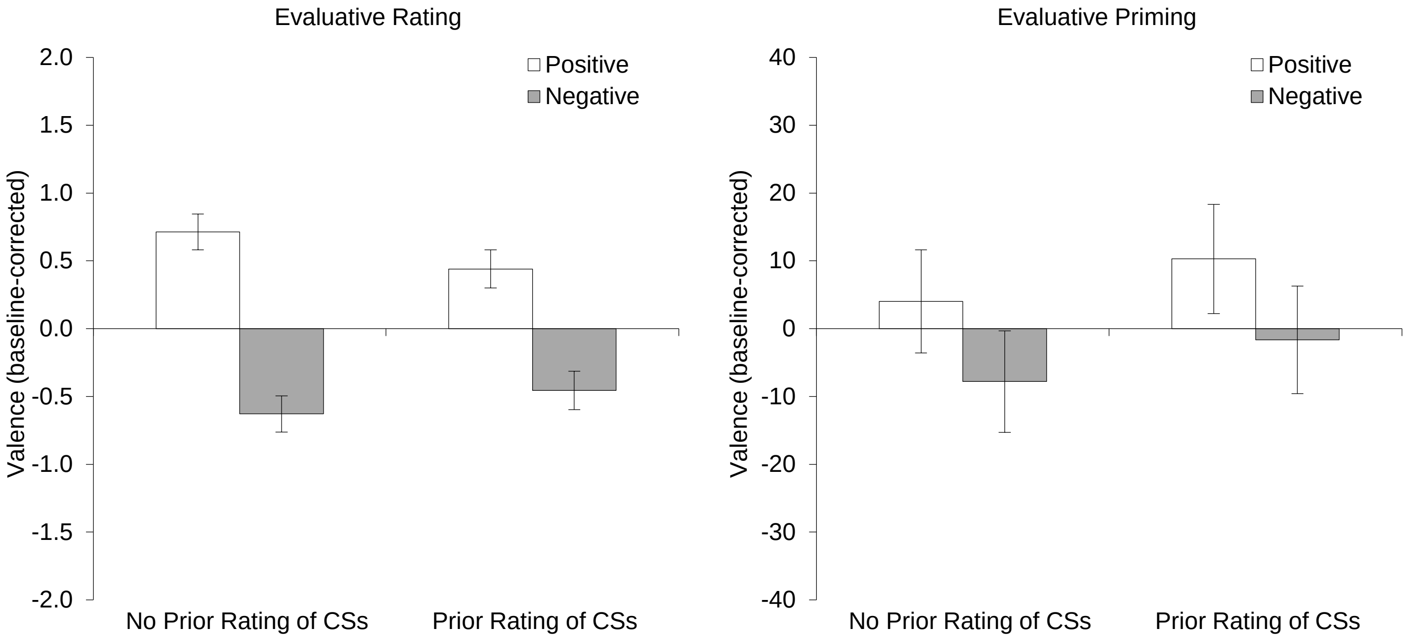

Figure 2. Baseline-corrected post-extinction CS evaluations as a function of US valence (positive vs. negative), prior rating of the CSs after acquistion (no prior rating vs. prior rating), and type of measure (evaluative rating vs. evaluative priming), Experiment 3. Higher values indicate more favorable evaluations. Error bars depict standard errors.