On the Intentional Control of Conditioned Evaluative Responses

Robert Balas
Warsaw School of Social Sciences and Humanities

Bertram Gawronski
The University of Western Ontario

The evaluative conditioning (EC) effect is defined as a change in the evaluation of a conditioned stimulus (CS) due to its pairing with a valenced unconditioned stimulus (US). The current research investigated the controllability of EC effects by asking participants to either promote or prevent the influence of CS-US pairings before they provided evaluative ratings of the CS. Experiment 1 showed that instructions to maximize or minimize the influence of CS-US pairings moderated EC effects in line with task instructions. However, this moderation was observed only when participants were able to recall the valence of the US that had been paired with a given CS. When participants failed to remember the valence of the US, significant EC effects emerged regardless of control instructions. Experiment 2 tested whether the influence of CS-US pairings on CS evaluations can be intentionally reversed. The results showed that reversal instructions led to a reverse EC effect when participants were able to recall the valence of the US that had been paired with a given CS, but not when they were unable to recall the valence of the US. Taken together, these results suggest that US valence memory is a necessary precondition for controlling the expression of a conditioned evaluative response, but it is not a necessary precondition for the emergence of EC effects per se.

Keywords: attitude expression; contingency memory; evaluative conditioning; judgmental correction

Evaluative processes play a major role in virtually all aspects of human life. Examples include the choice of friends and romantic partners, prejudice against social groups, and preferences for food and consumer products. Thus, an important question is how people acquire their likes and dislikes. This question is a central theme in research on evaluative conditioning (EC; for a review, see De Houwer, Thomas, & Baeyens, 2001), which is defined as the change in the evaluation of a conditioned stimulus (CS) due to its pairing with a valenced unconditioned stimulus (US).

Although EC effects are well-established in the psychological literature (for a meta-analysis, see Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010), the processes underlying EC effects are still the subject of ongoing debates (see De Houwer, 2009; Gawronski & Bodenhausen, 2011; Jones, Olson, & Fazio, 2010; Kruglanski & Dechesne, 2006; Mitchell, De Houwer, & Lovibond, 2009). An important question in these debates concerns the functional properties of the mechanisms that are responsible for EC effects. For example, it has been argued that EC is distinct from Pavlovian conditioning in the sense that it is (a) independent of participants' awareness of CS-US contingencies, (b) related to mere spatio-temporal contiguity of CS-US co-occurrences instead of their statistical contingencies, and (c) resistant to extinction (see De Houwer et al., 2001). These functional properties have been related to a process of automatic link formation, in which the mental representation of the CS is automatically associated with either the mental representation of the US (e.g., Walther, Gawronski, Blank, & Langer, 2009) or the affective response elicited by the US (e.g., Sweldens, Van Osselaer, & Janiszewski, 2010).

Counter to the assumption that EC effects are the product of automatic processes, however, recent research suggests that EC is characterized by various features of controlled processing. This research includes demonstrations that EC effects depend on conscious memory for CS-US contingencies (Bar-Anan, De Houwer, & Nosek, 2010; Dawson, Rissling, Schell, & Wilcox, 2007; Pleyers, Corneille, Luminet, & Yzerbyt, 2007; Stahl & Unkelbach, 2009; Stahl, Unkelbach, & Corneille, 2009). In addition, several studies have shown that EC depends on momentary processing goals (Corneille, Yzerbyt, Pleyers, & Mussweiler, 2009; Fiedler & Unkelbach, 2011; Gast & Rothermund, 2011) and the availability of attentional resources (Dedonder, Corneille, Yzerbyt, & Kuppens, 2010; Pleyers, Corneille, Yzerbyt, & Luminet, 2009) during the encoding of CS-US pairings. These results challenge earlier claims that EC effects are the product of automatic processes. Instead, they suggest that controlled processes may play a major role in the emergence of EC effects.

The main goal of the current research was to investigate another feature of controlled processing: the controllability of EC effects (see Moors & De Houwer, 2006). More specifically, we were interested in whether the *expression* of a conditioned evaluative response can be intentionally enhanced or reduced, and if so, what conditions have to be met for the intentional modulation of EC effects. Social-cognitive theories suggest that successful correction of biasing influences on overt behavior depends on at least three factors: (a) people have to be *motivated* to control their behavior for biasing influences; (b) people have to be *able* to engage in behavioral control; and (c) people have to be *aware* of the biasing influence (Strack & Hannover, 1996; Wegener & Petty, 1997; Wilson & Brekke, 1994). With regard to the expression of a conditioned evaluative response, one could argue that people generally have the ability to control their responses on traditional self-report measures of evaluation. Thus, to the extent that they are motivated to follow instructions to either promote or prevent the impact of CS-US pairings on their self-reported evaluations, an important boundary condition may be people's awareness of the biasing influence, namely whether a given CS had been paired with either a positive or a negative US. In other words, although it seems likely that the influence of CS-US pairings on self-reported evaluations can be intentionally controlled, successful control may depend on people's memory for the valence of the US that had been paired with a given CS.

To investigate the controllability of conditioned evaluative responses, participants in the current studies were presented with pairings of neutral CSs and valenced USs. Immediately after the conditioning task, participants were asked to rate all CSs on a self-report measure of evaluation. In Experiment 1, participants were asked to either maximize or minimize the influence of CS-US pairings on their evaluative judgments. In Experiment 2, participants were asked to either maximize or reverse the influence of CS-US pairings. Both experiments additionally included a control condition in which participants received no additional instructions before they were asked to evaluate the CSs. Drawing on theories of bias correction (Strack & Hannover, 1996; Wegener & Petty, 1997; Wilson & Brekke, 1994), we expected that instructions to maximize, minimize, or reverse the influence of CS-US pairings on CS evaluations moderate the size of EC effects in line with task instructions. However, this moderation was expected to depend on participants' memory for the valence of the US that had been paired with a given CS.

Experiment 1

Method

Participants. 164 students (101 female, 63 male) from various faculties of higher-education institutions in Warsaw volunteered to participate in the current study without compensation. Participants' age ranged from 18 to 30 years (M = 23.1, SD = 2.68). Each participant was randomly assigned to one of three experimental conditions and tested individually. All participants provided informed consent before the experiment and were debriefed after they completed the study.

Materials. The study was run on a standard PC laptop computer using Inquisit 3.0 (Draine, 2011). Six neutral pronounceable non-words were used as CSs (e.g., *pumata*, *hatuka*, *sumira*). All of these non-words had been pre-tested in previous experiments and were selected on the basis of their neutral valence and low variability of evaluations. As USs we used three positive and three negative images from the International Affective Picture System (Lang, Bradley, & Cuthbert, 2008). The three positive images depicted a toddler (Image 2070), a kitten (Image 1460), and a happy family (Image 2340); the three negative images depicted a spider (Image 1201), an aggressive dog (Image 1301), and a crying child (Image 2900). The USs were selected on the basis of a pretest in which 40 participants were asked to rate a larger sample of positive and negative pictures on 21-point scales raging from 1 (*extremely negative*) to 21 (*extremely positive*) with 11 serving as a neutral reference point. In this pretest, the three positive USs were rated higher (M = 16.21) than the three negative USs (M = 6.56), $F(1, 39) = 24.32$, $p < .001$, $\eta_p^2 = .46$. There was no difference between the positive and negative USs in terms of their valence strength, which is reflected in their absolute difference to the neutral reference point (Ms = 4.21 vs. 4.44), $F(1, 39) = .48$, $p < .86$, $\eta_p^2 = .01$. Each CS was paired with the same US for each participant; the particular pairings of CSs and USs were counterbalanced across participants.

Procedure and Measures. Participants' first task was to evaluate all CSs on a 21-point rating scale ranging from 1 (*extremely negative*) to 21 (*extremely positive*). Each CS was presented in black 22pt Arial font against a white background. A rating scale appeared simultaneously below the CS until participants gave their response (ITI = 1000 ms).

After the pre-conditioning ratings of the CSs, participants were asked to carefully watch words and pictures on the computer screen. The instructions informed participants that the study concerns language acquisition and that they would be presented with words (CS) together with pictures (US) illustrating their meanings (see Mitchell, Anderson, & Lovibond, 2003). CS-US pairs were simultaneously displayed on a white background for 2000 ms followed by a 2000 ms blank screen. Each CS-US pair was presented seven times in randomized order.

After the presentation of the CS-US pairings, participants were asked to evaluate the CSs a second time using the same scale that was used for the pre-conditioning ratings. To investigate the controllability of EC effects, we manipulated the instructions that participants received before they evaluated the CSs. One third of the participants were simply asked to evaluate the CSs (*neutral instructions condition*). Another third was informed about the influence of CS-US pairings on evaluations of the CSs and instructed to rate the words so as to maximize the impact of CS-US pairings on CS evaluations (*proactive control condition*). The final third was informed about the influence of CS-US pairings on evaluations of the CSs and instructed to rate the words so as to minimize the impact of CS-US pairings on CS evaluations (*counteractive control condition*). The exact wording of the instructions is provided in the Appendix.

After the post-conditioning ratings, we used a force-choice task to measure participants' memory for the valence of the US that had been paired with a given CS. Toward this end, each CS was presented individually on the screen with the evaluative labels *negative* and *positive*. Participants' task was to indicate the valence of the US that had been paired with the presented CS. The ITI was 1000 ms.

Results

EC Effects. To investigate the controllability of EC effects, we calculated change scores by subtracting pre-conditioning ratings of a given CS from post-conditioning ratings of the same CS. Thus, negative values indicate less favorable evaluations as a result of CS-US pairings, whereas positive values indicate more favorable evaluations. The difference scores were then aggregated by averaging the scores of the three CSs that had been paired with a US of the same valence. Submitted to a 3 (Task Instructions: neutral instructions vs. proactive control vs. counteractive control) × 2 (US Valence: positive vs. negative) mixed-model ANOVA, these scores revealed a significant main effect of US valence, $F(1, 161) = 95.83$, $p < .001$, $\eta_p^2 = .48$, indicating that evaluations became more favorable for CSs that had been paired with a positive US ($M = 4.02$) and less favorable for CSs that had been paired with a negative US ($M = -2.78$). This effect was qualified by a significant two-way interaction between task instructions and US valence, $F(2, 161) = 10.42$, $p < .001$, $\eta_p^2 = .12$. Although the difference between positively and negatively conditioned CSs was statistically significant in all three groups, EC effects were less pronounced in the counteractive control condition (*M*s = 1.94 vs. -1.42), $F(1, 55) = 21.27$, $p < .001$, $\eta_p^2 = .28$, compared with the neutral instructions condition (*M*s = 4.46 vs. -3.11), $F(1, 55) = 59.88$, $p < .001$, $\eta_p^2 = .52$, and the proactive control condition (*M*s = 5.64 vs. -3.81), $F(1, 51) = 65.82$, $p < .001$, $\eta_p^2 = .56$. The size of EC effects, defined as the difference between positively and negatively conditioned CSs, differed between the counteractive control condition and the proactive control condition, $F(1, 106) = 20.21$, $p < .001$, $\eta_p^2 = .16$, as well as the counteractive control condition and the neutral instructions condition, $F(1, 110) = 11.89$, $p < .001$, $\eta_p^2 = .10$, but not between the proactive control condition and the neutral instructions condition, $F(1, 106) = 1.53$, $p = .22$, $\eta_p^2 = .02$.

US Valence Memory. Memory for US valence was significantly above the chance level of .50 with a mean accuracy of .63, $t(163) = 12.58$, $p < .001$. Memory for US valence did not differ as a function of task instructions or US valence (all *F*s < 1, all *p*s > .50). To investigate the role of US valence memory in the control of EC effects, we calculated four separate CS evaluation scores for each participant depending on whether (a) a given CS was paired with a positive or negative US and (b) participants did or did not correctly recall the valence of the US a given CS had been paired with. Because this data analytic strategy produces missing data in the evaluation measure when participants show either 100% or 0% correct responses on the memory measure within a given cell of our research design (see Table 1), such missing data were replaced with values estimated by linear trends based on the existing values within the same cell (see Howell, 2007). For example, if a participant in the counteractive control condition failed to correctly recall the valence of all positive USs, this participant would have a missing value in the evaluation measure in this particular cell. Thus, to avoid a loss of statistical power resulting from the exclusion of this participant, the missing data point would have been replaced by the mean evaluation of the CSs that had been paired with a positive US by all other participants in the counteractive control condition who did not have a missing value in this particular cell.[1]

Submitted to a 3 (Task Instructions) × 2 (US Valence) × 2 (Memory for US Valence) mixed-model ANOVA, the resulting evaluation scores replicated the main effect of US valence, $F(1, 161) = 102.94$, $p < .001$, $\eta_p^2 = .50$, and the interaction between US valence and task instructions, $F(2, 161) = 8.26$, $p < .001$, $\eta_p^2 = .09$. Memory for US valence further revealed a significant two-way interaction with US valence, $F(1, 161) = 13.69$, $p < .001$, $\eta_p^2 = .08$, such that the difference between positively and negatively conditioned CSs was smaller in the absence of US valence memory (*M*s = 2.91 vs. -1.44) than when participants did remember the valence of the US a given CS had been paired with (*M*s = 4.23 vs. -3.04). Nevertheless, EC effects were statistically significant regardless of whether participants did or did not remember the valence of the US a particular CS had been paired with, $t(163) = 10.27$, $p < .001$ and $t(163) = 7.86$, $p < .001$, respectively. More importantly, there was a significant three-way interaction between US valence, task instructions, and memory for US valence, $F(2, 161) = 6.32$, $p < .01$, $\eta_p^2 = .07$ (see Figure 1).

[1] To avoid confusion, it is important to point out that the data substitution was done on the basis of the mean values that were observed in a given cell (e.g., CSs paired with a positive US without US valence memory in the proactive control group), rather than on the basis of values from single CSs. For example, if a participant in the proactive control group correctly remembered the valence of the US for two out of three CS-US pairs and failed to remember the valence of the US for the third CS-US pair, there would be no empty cells for CSs with US valence memory (i.e., mean evaluation of two CSs) and without US valence memory (i.e., evaluation of a single CS). The only situation that generated a missing value for a given condition was when a participant remembered the valence of all three USs within a given cell or failed to remember the valence of all three USs.

Whereas the interaction between task instructions and US valence was statistically significant when participants remembered the valence of the US a given CS had been paired with, $F(2, 161) = 11.62$, $p < .001$, $\eta_p^2 = .13$, this interaction failed to reach statistical significance when they did not the remember the valence of the US, $F(2, 161) = 1.39$, $p = .25$, $\eta_p^2 = .02$. This result indicates that task instructions influenced EC effects only when participants correctly remembered the valence of US a given CS had been paired with, but not when they failed to remember the valence of the US. To further specify the three-way interaction in terms of our hypotheses, we also conducted separate 2 (US Valence) × 2 (Memory for US Valence) ANOVAs for each of the three task instruction conditions.

In the neutral instructions condition, the ANOVA revealed a significant main effect of US valence, $F(1, 55) = 49.62$, $p < .001$, $\eta_p^2 = .47$, which was qualified by a significant two-way interaction between US valence and memory for US valence, $F(1, 55) = 14.02$, $p < .001$, $\eta_p^2 = .20$. Although the difference between positively and negatively conditioned CSs was statistically significant regardless of US valence memory, EC effects were more pronounced when participants were able to remember the valence of the US a given CS had been paired with, $t(55) = 6.74$, $p < .001$, than when they were unable to remember the valence of the US, $t(55) = 4.08$, $p < .01$.

In the proactive control condition, the ANOVA also revealed a significant main effect of US valence, $F(1, 51) = 77.07$, $p < .001$, $\eta_p^2 = .60$, and a significant two-way interaction between US valence and memory for US valence, $F(1, 51) = 10.45$, $p < .01$, $\eta_p^2 = .17$. Post-hoc analyses specified this interaction by showing a larger difference between positively and negatively conditioned CSs when participants were able to remember the valence of the US a given CS had been paired with, $t(51) = 9.09$, $p < .001$, than when they were unable to remember the valence of the US, $t(51) = 4.53$, $p < .01$.

In the counteractive control condition, the ANOVA also revealed a significant main effect of US valence, $F(1, 55) = 31.47$, $p < .001$, $\eta_p^2 = .36$, and a significant two-way interaction between US valence and memory for US valence, $F(1, 55) = 5.03$, $p < .05$, $\eta_p^2 = .09$. However, the particular pattern of the two-way interaction was remarkably different from the one obtained in the neutral instructions and the proactive control conditions. Specifically, participants in the counteractive control condition showed a significant difference between positively and negatively conditioned CSs only when they were *unable* to recall the valence of the US a given CS had been paired with, $t(55) = 5.42$, $p < .01$, but not when they were able to recall the valence of the US, $t(55) = 1.47$, $p = .19$.[2]

Discussion

The results of Experiment 1 suggest that EC effects on self-reported evaluations can be intentionally enhanced or reduced. However, intentional control of EC effects depended on US valence memory. To the extent that participants correctly recalled the valence of the US a given CS had been paired with, EC effects were enhanced when participants were instructed to maximize the impact of CS-US pairings and reduced when they were instructed to minimize the impact of CS-US pairings. However, EC effects remained unaffected by control instructions when participants failed to remember the valence of the US that had been paired with a given CS. In this case, we found significant EC effects regardless of the particular instructions participants received before evaluating the CSs. The latter finding suggests that US valence memory may not be a necessary precondition for EC effects and that EC effects may still emerge in spite of faded memory for the CS-US pairings that are responsible for these effects (see also Hütter, Sweldens, Stahl, Unkelbach, & Klauer, in press). Nevertheless, US valence memory seems to be a necessary precondition for controlling the expression of a conditioned evaluative response.

Experiment 2 expanded on these findings by examining whether participants are able to reverse the influence of CS-US pairings on self-reported evaluations. We expected that reversal instructions would lead to a reversed EC effect, such that CS-pairings with a positive US lead to less favorable evaluations and CS-pairings with a negative US lead to more favorable evaluations. Yet, in line with the findings of Experiment 1, the emergence of a reverse EC effect was expected to depend on participants' memory for the valence of the US that had been paired with a given CS.

Experiment 2

Method

Participants. 135 participants (75 female, 60 male) from various faculties of Warsaw University volunteered to participate in the current study without compensation. Participants' age ranged from 18 to 32 years ($M = 24.27$, $SD = 4.65$). Each participant was randomly assigned to one of three experimental conditions and tested individually. All participants

[2] Although statistical power for testing the predicted effects was substantially reduced when participants with missing values in the memory measure were excluded (24.4% of the total sample), the basic pattern of means replicated without the employed data replacement. Importantly, the reduced sample still revealed a significant EC effect when participants failed to remember the valence of the US, $t(92) = 6.32$, $p < .01$.

provided informed consent before the experiment and were debriefed after they completed the study.

Materials and Procedure. The materials and procedure were identical to Experiment 1, the only difference being that participants in one of the three experimental conditions were instructed to reverse (rather than minimize) the influence of CS-US pairings on CS evaluations (see Appendix). The instructions in the other two conditions (i.e., neutral instructions, proactive control) were identical to Experiment 1.

Results

EC Effects. Evaluation scores were aggregated according to the procedures described for Experiment 1. A 3 (Task Instructions) × 2 (US Valence) mixed-model ANOVA revealed a significant main effect of US valence, $F(1, 132) = 44.48$, $p < .001$, $\eta_p^2 = .25$, indicating that evaluations became more favorable for CSs that had been paired with positive USs ($M = 2.91$) and less favorable for CSs that had been paired with negative USs ($M = -1.89$). This main effect was again qualified by a significant two-way interaction between US valence and task instructions, $F(2, 132) = 24.53$, $p < .001$, $\eta_p^2 = .27$. Post-hoc analyses revealed that positively conditioned CSs were evaluated more favorably than negatively conditioned CSs in the proactive control condition (*M*s = 5.56 vs. -4.34), $t(44) = 10.06$, $p < .001$, and the neutral instructions condition (*M*s = 3.73 vs. -2.78), $t(44) = 4.93$, $p < .001$, but not in the reversal condition (*M*s = -.62 vs. 1.45), $t(44) = -1.48$, $p = .14$. The size of EC effects, defined as the difference between positively and negatively conditioned CSs, differed between the reversal condition and the proactive control condition, $F(1, 88) = 49.43$, $p < .001$, $\eta_p^2 = .36$, the reversal control condition and the neutral instructions condition, $F(1, 88) = 20.09$, $p < .001$, $\eta_p^2 = .19$, as well as the proactive control condition and the neutral instructions condition, $F(1, 88) = 4.13$, $p < .05$, $\eta_p^2 = .05$.

US Valence Memory. Memory for US valence was significantly above the chance level of .50 with a mean accuracy level of .74, $t(134) = 18.83$, $p < .001$. Memory for US valence did not differ as a function of task instructions or US valence (all *F*s < 1, all *p*s > .40). Missing data were again replaced with values estimated by means of linear trends based on existing values within each of the three task instruction conditions (see Table 1). A 3 (Task Instructions) × 2 (US Valence) × 2 (Memory for US Valence) mixed-model ANOVA replicated the main effect of US valence, $F(1, 132) = 54.39$, $p < .001$, $\eta_p^2 = .29$, and the interaction between US valence and task instructions, $F(2, 132) = 28.06$, $p < .001$, $\eta_p^2 = .29$. In addition, the analysis revealed a significant two-way interaction between US valence and memory for US valence, $F(1, 132) = 12.69$, $p < .001$, $\eta_p^2 = .08$, showing that the difference between positively and negatively conditioned CSs was statistically significant when participants remembered the valence of the US a given CS had been paired with (*M*s = 2.83 vs. -2.28, respectively), $t(134) = 5.71$, $p < .001$, but not when they failed to remember the valence of the US (*M*s = .56 vs. -.38, respectively), $t(134) = 1.76$, $p = .12$. Again, these effects were qualified by a significant three-way interaction between US valence, task instructions, and memory for US valence, $F(2, 132) = 10.82$, $p < .001$, $\eta_p^2 = .14$ (see Figure 2). Replicating the pattern obtained in Experiment 1, the interaction between task instructions and US valence was statistically significant when participants were able to remember the valence of the US that had been paired with a given CS, $F(2, 132) = 26.63$, $p < .001$, $\eta_p^2 = .28$, but not when they failed to remember the valence of the US, $F(2, 132) = 2.32$, $p = .10$, $\eta_p^2 = .03$. This result indicates that task instructions influenced EC effects only when participants correctly remembered the valence of US a given CS had been paired with, but not when they failed to remember the valence of the US. To further specify the three-way interaction in terms of our hypotheses, we again conducted separate 2 (US Valence) × 2 (Memory for US Valence) ANOVAs for each of the three task instruction conditions.

In the neutral instructions condition, the ANOVA revealed a significant main effect of US valence, $F(1, 44) = 32.33$, $p < .001$, $\eta_p^2 = .42$, which was qualified by a significant two-way interaction between US valence and memory for US valence, $F(1, 44) = 4.23$, $p < .05$, $\eta_p^2 = .11$. Although the difference between positively and negatively conditioned CSs was statistically significant regardless of US valence memory, EC effects were more pronounced when participants were able to remember the valence of the US a given CS had been paired with, $t(44) = 5.37$, $p < .01$, than when they were unable to remember the valence of the US, $t(44) = 3.04$, $p < .05$.

In the proactive control condition, the ANOVA also revealed a significant main effect of US valence, $F(1, 44) = 76.53$, $p < .001$, $\eta_p^2 = .74$, and a significant two-way interaction between US valence and memory for US valence, $F(1, 44) = 9.47$, $p < .01$, $\eta_p^2 = .22$. As with the neutral instructions condition, EC effects were larger when participants were able to remember the valence of the US a given CS had been paired with, $t(44) = 10.79$, $p < .001$, than when they were unable to remember the valence of the US, $t(44) = 4.58$, $p < .01$.

In the reversal condition, the ANOVA revealed a marginally significant main effect of US valence, $F(1, 44) = 3.07$, $p = .09$, $\eta_p^2 = .06$, and a significant two-way interaction between US valence and memory for US valence, $F(1, 44) = 5.43$, $p < .05$, $\eta_p^2 = .13$. However, the particular pattern of this interaction was again different from the one obtained in the neutral instructions and the proactive control conditions. Specifically, participants showed a reverse EC effect when they were able to remember the valence of the US that had been paired with a given CS, $t(44) = -2.98$, $p < .05$, but this effect failed to reach statistical significance when they were unable to recall the valence of the US, $t(44) = -0.68$, $p = .45$.[3]

Discussion

Experiment 2 replicated and extended the results of Experiment 1 by showing that intentional control of conditioned evaluative responses depended on participants' memory for the valence of the US that had been paired with a given CS. Specifically, participants were able to enhance or reverse the influence of CS-US pairings on self-reported CS evaluations, but only when they were able to remember the valence of the US that had been paired with a given CS. Interestingly, EC effects were eliminated (though not reversed) by reversal instructions when participants failed to correctly remember the valence of the US. A potential explanation of this finding is that these participants tried to identify alternative cues that they could use as an anchor for the reversal of their evaluative ratings, which should attenuate EC effects if these cues are unrelated to the valence of the US a given CS had been paired with (e.g., contingent evaluative features of the nonwords used as CSs). Another interesting aspect of Experiment 2 is that EC effects again emerged in the absence of US valence memory in the conditions with neutral and proactive control conditons. This finding supports the assumption that EC effects may emerge in spite of faded memory for the CS-US pairings that are responsible for these effects (see also Hütter et al., in press). Yet, US valence memory seems to be a necessary precondition for reversing the expression of a conditioned evaluative response, although reversal instructions seem to be sufficient to attenuate EC effects in the absence of US valence memory.

General Discussion

The present research addressed the question of whether the expression of conditioned evaluative responses can be intentionally controlled. We investigated this issue by instructing participants to maximize, minimize, or reverse the influence of CS-US pairings before they completed a self-report measure of CS evaluations. The main finding is that the influence of CS-US pairings on self-reported evaluations can be intentionally controlled, such that EC effects were moderated in line with instructions to maximize, minimize, or reverse the influence of CS-US pairings. However, the impact of task instructions depended on participants' memory for the valence of the US that had been paired with a given CS, such that task instructions moderated EC effects only when US valence memory was present, but not when it was absent. Drawing on social-cognitive theories of bias correction (Strack & Hannover, 1996; Wegener & Petty, 1997; Wilson & Brekke, 1994), one could argue that US valence memory is an important aspect of participants' awareness of the influence of CS-US pairings, which represents a necessary precondition for successful judgmental correction over and above motivation and ability. Thus, even when people are motivated and able to control the influence of CS-US pairings on self-reported CS evaluations, their attempts to control for this influence may remain ineffective when they lack memory for the valence of the US that had been paired with a given CS.

Another interesting finding is that participants showed significant EC effects even when they failed to correctly remember the valence of the US that had been paired with a given CS. The only condition in which did not find a significant EC effect in the absence of US valence memory was when participants were instructed to reverse the influence of CS-US pairings. Although this finding should be replicated before drawing strong conclusions, one could argue that reversal instructions led participants to reverse their evaluative judgments on the basis of whatever cue they considered a valid anchor for making an evaluative judgment. For example, if contingent features of a given CS (e.g., pleasant pronunciation) led participants to assume that this CS must have been paired with a US of a particular valence (e.g., pleasant US), they might reverse their response by providing an evaluative judgment of the opposite valence (e.g., a negative evaluation). To the extent that these features are unrelated to the actual CS-US contingencies, such a reversal strategy should attenuate EC effects in the absence of US valence memory. Nevertheless, US valence memory does not seem to be a necessary precondition for the emergence of EC effects per se, given that significant EC effects emerged in the absence of US valence memory in all other conditions. Although this finding stands in contrast to several studies showing EC effects only when participants correctly remembered the valence of the US that had been paired with a given CS (e.g., Bar-Anan et al., 2010; Dawson et al., 2007; Pleyers et al., 2007; Stahl & Unkelbach, 2009; Stahl et al., 2009), it is consistent with recent evidence by Hütter et al. (in press) who demonstrated reliable EC effects in the absence of US valence memory using multinomial modeling to

[3] As with Experiment 1, the basic pattern of means replicated without the employed data replacement, although statistical power was substantially reduced due to the exclusion of a considerable number of participants (22.6% of the total sample). Nevertheless, EC effects emerged in the absence of US valence memory for participants in the neutral instructions condition, $t(29) = 2.98$, $p < .05$, as well as for participants in the proactive control condition, $t(32) = 3.43$, $p < .05$.

disentangle the contribution of multiple distinct processes to memory judgments.

Although Hütter et al.'s (in press) research was not directly concerned with the question of whether the expression of conditioned attitudes can be intentionally controlled, their paradigm shares some resemblance to the one employed in the current studies. To determine the relation between US valence memory and EC effects, Hütter et al. used a memory test that involved either an inclusion instruction or an exclusion instructions. In the inclusion condition, participants were asked to identify the actual valence of the US that had been paired with a given CS. In the exclusion condition, participants were instructed to identify the valence that was opposite to the valence of the paired US. In addition, participants were asked to respond on the basis of either their memory for CS-US pairings or their attitude towards the CSs. Using multinomial modeling to disentangle the contribution of memory-related and attitudinal processes to memory judgments, Hütter et al. obtained evidence for the impact of conditioned attitudes to memory judgments that were independent of participants' memory for US valence.

Hütter et al.'s exclusion instructions is similar to the reversal instructions in Experiment 2, in that both requested participants to respond in a manner that is opposite to their judgmental inclination. Nevertheless, there are several important differences between the two studies. The most significant difference is that Hütter et al.'s exclusion instructions referred to participants' responses in a memory task, whereas reversal instructions in the current research referred to evaluative ratings of the CSs. Moreover, whereas the memory task in Hütter et al.'s research was administered before participants were asked to rate the valence of the CS, evaluative ratings preceded the memory task in the current research. Despite these methodological differences, the evidence provided by the two studies leads to the same conclusion that EC effects do not depend on US valence memory.[4] However, the current findings go beyond Hütter et al.'s research by showing that, although US valence memory is not a necessary precondition for EC effects, it is a necessary precondition for controlling the expression of a conditioned evaluative response.

Despite the converging evidence for EC effects in the absence of US valence memory, it is important to note that these findings do not necessarily indicate that EC effects occur when participants are unaware of the relevant CS-US contingencies during encoding. As Gawronski and Walther (in press) pointed out, memory measures are not suitable to resolve the ongoing debate about the role of contingency awareness in EC, because (a) memory measures conflate encoding and retrieval processes and (b) the relation between memory and evaluation measures in the standard paradigm is merely correlational. These ambiguities imply that any possible outcome can be interpreted in at least two possible ways: one suggesting that contingency awareness during encoding is a necessary precondition for EC effects and one suggesting that EC effects do not require contingency awareness during encoding. In the current studies, for example, participants may have been aware of relevant contingencies during the encoding of CS-US pairings, but they may have failed to remember the valence of USs in the subsequent memory task due to retrieval-related processes. Thus, although the current findings suggest that memory for US valence is not required for EC effects to occur (see also Hütter et al., in press), they do not speak to the question of whether EC effects require conscious awareness of CS-US contingencies during the encoding of CS-US pairings.

A related caveat is that the current findings speak only to the mechanisms that are involved in the *expression* of a conditioned attitude; they do not speak to the mechanisms that are involved in the *formation* of a conditioned attitude. The latter question is particularly relevant for the ongoing debate between the proponents of associative and propositional accounts of EC. Whereas associative accounts argue that EC effects are driven by an associative process of automatic link formation (e.g., Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006), propositional accounts attribute EC effects to the non-automatic acquisition and validation of propositional knowledge about CS-US relations (e.g., De Houwer, 2009; Mitchell et al., 2009). Although this debate is primarily concerned with how conditioned attitudes are formed, the current findings suggest important limitations in participants' ability to control the expression of conditioned attitudes.[5] Specifically, our findings suggest that intentional control over the expression of a conditioned attitude is possible to the extent that people have accurate memory for the valence of the US that had been paired with a given CS. Thus, an important question for future research is

[4] A potential objection is that the current research assessed only US valence memory, but not memory for the specific US that had been paired with a given CS. Although we cannot rule out that a measure of US identity memory would have led to different conclusions, a recent study by Stahl et al. (2009) suggests that US identity memory does not provide any contribution to EC effects that is not accounted for by US valence memory.

[5] Another important caveat in this context is that participants in the current study were told that the pictures illustrated the meaning of the artificial words (see Mitchell et al., 2003). In a strict sense, this instruction goes beyond mere pairings between a CS and a US by including propositional information about their relation. Future research should investigate whether similar effects occur for mere pairings of CSs and USs without explicit information about their relation.

whether these findings generalize to the formation of conditioned attitudes, in particular whether the formation of conditioned attitudes can be intentionally controlled. Whereas propositional accounts suggest that the formation of conditioned attitudes is controllable, associative accounts would argue that conditioned attitudes may be formed even when participants are motivated to avoid being influenced by CS-US pairings.

A potential objection against the current studies concerns the nature of our memory measure, which required participants to select either the *positive* or the *negative* response option regardless of whether they did or did not remember the valence of the US. Thus, it is possible that some participants without US valence memory showed correct responses on the memory task by mere chance. Moreover, participants with inaccurate US valence memory might have used their subjective memory judgments to control their evaluative responses if they were convinced that their memory judgments were accurate. Despite this limitation of our memory measure, it is important to note that our treatment of the memory data is conservative, in that it counteracts the effects obtained in the current studies. On the one hand, *correct memory* classifications did not control for accuracy by mere chance, which reduces the probability of identifying the obtained instruction effects in the presence of US valence memory. On the other hand, *inaccurate memory* classifications should be associated with reversed EC effects if participants used their incorrect memories as a basis for their CS evaluations, which reduces the likelihood of significant EC effects in the absence of US valence memory.

The current findings also have important implications for the long-debated role of demand compliance in EC, particularly in studies using self-reported evaluations as the primary dependent measure. As Field (2005) pointed out, EC effects in many previous studies could be interpreted as reflecting participants' compliance in providing evaluations that are in line with the presumed hypothesis of the experimenter. However, such compliance effects seem possible only to the extent that participants are able to remember the relevant details of the CS-US pairings (Field & Moore, 2005). Although concerns about demand compliance have been raised for quite some time, they have been largely methodological without evidence for their validity. The current findings provide empirical support for these concerns, showing that participants can intentionally maximize or minimize the influence of CS-US pairings on self-reported evaluations, but only when they correctly remember the valence of the US that had been paired with a given CS. Thus, in line with EC studies using implicit measures as dependent variables (e.g., Dawson et al., 2007; Olson & Fazio, 2001; Pleyers et al., 2007; Stahl et al., 2009), the current findings indicate that findings obtained with self-report measures should ideally be replicated with measure that are less susceptible to strategic influences (cf. Gawronski & De Houwer, in press). In addition, demand compliance could be ruled out by using paradigms that reduce the likelihood that participants can remember the valence of the US that had been paired with a given CS. Although the latter recommendation stands in contrast with claims that EC effects depend on US valence memory (e.g., Bar-Anan et al., 2010; Dawson et al., 2007; Pleyers et al., 2007; Stahl et al., 2009), it seems appropriate from the perspective of the current studies which showed significant EC effects even when participants were unable to correctly recall the valence of the US that had been paired with a given CS.

References

Bar-Anan, Y., De Houwer, J., & Nosek, B. A. (2010). Evaluative conditioning and conscious knowledge of contingencies: A correlational investigation with large samples. *Quarterly Journal of Experimental Psychology*, *63*, 2313-2335. doi: 10.1080/17470211003802442

Corneille, O., Yzerbyt, V., Pleyers, G., & Mussweiler, T. (2009). Beyond awareness and resources: Evaluative conditioning may be sensitive to processing goals. *Journal of Experimental Social Psychology, 45*, 279-282. doi: 10.1016/j.jesp.2008.08.020

Dawson, M. E., Rissling, A. J., Schell, A. M., & Wilcox, R. (2007). Under what conditions can human affective conditioning occur without contingency awareness? Test of the evaluative conditioning paradigm. *Emotion, 7,* 755-766.

De Houwer, J. (2009). The propositional approach to learning as an alternative for association formation models. *Learning and Behavior, 37,* 1-20.

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. doi: 10.1037/0033-2909.127.6.853

Dedonder, J., Corneille, O., Yzerbyt, V., & Kuppens, T. (2010). Evaluative conditioning of high-novelty stimuli does not seem to be based on an automatic form of associative learning. *Journal of Experimental Social Psychology, 46,* 1118-1121.

Draine, S. (2011). *Inquisit (Version 3.0.3) [Computer Software]*. Seattle, WA: Millisecond Software. Retrieved from http://www.millisecond.com

Fiedler, K., & Unkelbach, C. (2011). Evaluative conditioning depends on higher order encoding processes. *Cognition & Emotion, 25,* 639-656.

Field, A. P. (2005). Learning to like (and dislike): Associative learning of preferences. In A. J. Wills

(Ed.), *New directions in human associative learning* (pp. 221-252). Mahwah, New Jersey: LEA.

Field, A. P., & Moore, A. C. (2005). Dissociating the effects of attention and contingency awareness on evaluative conditioning effects in the visual paradigm. *Cognition & Emotion, 19,* 217-243.

Gast, A., & Rothermund, K. (2011). What you see is what will change: Evaluative conditioning effects depend on a focus on valence. *Cognition & Emotion, 25,* 89-110. doi: 10.1080/02699931003696380

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., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology, 44,* 59-127.

Gawronski, B., & De Houwer, J, (in press). 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). New York: Cambridge University Press.

Gawronski, B., & Walther, E. (in press). What do memory data tell us about the role of contingency awareness in evaluative conditioning? *Journal of Experimental Social Psychology*.

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

Howell, D. C. (2007). The analysis of missing data. In W. Outhwaite & S. Turner (Eds.), *Handbook of Social Science Methodology* (pp. 208-224). London: Sage.

Hütter, M., Sweldens, S., Stahl, C., Unkelbach, C., & Klauer, K. C. (in press). Dossociating contingency awareness and conditioned attitudes: Evidence of contingency-unaware evaluative conditioning. *Journal of Experimental Psychology: General*.

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

Kruglanski, A. W., & Dechesne, M. (2006). Are associative and propositional processes qualitatively distinct? Comment on Gawronski and Bodenhausen (2006). *Psychological Bulletin, 132,* 736-739.

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.

Mitchell, C. J., Anderson, N. E., & Lovibond, P. F. (2003). Measuring evaluative conditioning using the Implicit Association Test. *Learning & Motivation, 34,* 203-217.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*, 183-198. doi: 10.1017/s0140525x09000855

Moors, A., & De Houwer, J. (2006). Automaticity: A conceptual and theoretical analysis. *Psychological Bulletin*, *132*, 297-326.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12,* 413-147.

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: Item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 130-144. doi: 10.1037/0278-7393.33.1.130

Pleyers, G., Corneille, O., Yzerbyt, V., & Luminet, O. (2009). Evaluative conditioning may incur attentional costs. *Journal of Experimental Psychology: Animal Behavior Processes, 35*, 279-285. doi: 10.1037/a0013429

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology, 91,* 995-1008.

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. doi: 10.1037/A0013255

Stahl, C., Unkelbach, C., & Corneille, O. (2009). On the respective contributions of awareness of unconditioned stimulus valence and unconditioned stimulus identity in attitude formation through evaluative conditioning. *Journal of Personality and Social Psychology, 97,* 404-420. doi: 10.1037/A0016196

Strack, F., & Hannover, B. (1996). Awareness of the influence as a precondition for implementing correctional goals. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 579-596). New York: Guilford Press.

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.

Walther, E., Gawronski, B., Blank, H., & Langer, T. (2009). Changing likes and dislikes through the back door: The us-revaluation effect. *Cognition &*

Emotion, 23, 889-917. doi: 10.1080/02699930802212423

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naive theories of bias in bias correction. *Advances in Experimental Social Psychology, 29,* 141-208.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116,* 117-142.

Author Note

This research was supported by grant 557/N-COST/2009/0 from the Polish Ministry of Science and Higher Education.

Correspondence concerning this paper should be sent to: Robert Balas, Warsaw School of Social Sciences and Humanities, Chodakowska 19/31, 03-815, Warsaw, Poland, Email: rbalas@swps.edu.pl, Phone: +48 22 5179825.

Appendix:
Instructions Used in Experiments 1 and 2

Neutral Instruction Condition

Now we ask you to evaluate the words that you have seen with the pictures. The words will appear sequentially on the screen. A rating scale will be presented with each word and your task is to indicate your evaluation.

Proactive Control Condition

Now we ask you to evaluate the words that you have seen with the pictures. As you may know, research has shown that repeated pairings of a neutral stimulus (e.g., a neutral word) with a positive or negative stimulus (e.g., a positive or negative photograph) can influence people's responses to the neutral stimulus. Specifically, it has been shown that responses to a formerly neutral stimulus become more positive when it has been repeatedly paired with a positive stimulus and more negative when it has been repeatedly paired with a negative stimulus. In the current study, we are interested in whether the effects of such conditioning processes can be enhanced by people's intentional efforts to maximize the impact of affective stimuli on the conditioned meaning of neutral images. For this purpose, you will be again presented with the words that you have seen before with pictures. Your task is to evaluate the words and try your ABSOLUTE BEST to rate them so as to MAXIMIZE the influence of the positive/negative pictures that the words had been paired with.

Counteractive Control Condition

Now we ask you to evaluate the words that you have seen with the pictures. As you may know, research has shown that repeated pairings of a neutral stimulus (e.g., a neutral word) with a positive or negative stimulus (e.g., a positive or negative photograph) can influence people's responses to the neutral stimulus. Specifically, it has been shown that responses to a formerly neutral stimulus become more positive when it has been repeatedly paired with a positive stimulus and more negative when it has been repeatedly paired with a negative stimulus. In the current study, we are interested in whether the effects of such conditioning processes can be inhibited by people's intentional efforts to minimize the impact of affective stimuli on the conditioned meaning of neutral images. For this purpose, you will be again presented with the words that you have seen before with pictures. Your task is to evaluate the words and try your ABSOLUTE BEST to rate them so as to MINIMIZE the influence of the positive/negative pictures that the words had been paired with.

Reversal Instructions

Now we ask you to evaluate the words that you have seen with the pictures. As you may know, research has shown that repeated pairings of a neutral stimulus (e.g., a neutral word) with a positive or negative stimulus (e.g., a positive or negative photograph) can influence people's responses to the neutral stimulus. Specifically, it has been shown that responses to a formerly neutral stimulus become more positive when it has been repeatedly paired with a positive stimulus and more negative when it has been repeatedly paired with a negative stimulus. In the current study, we are interested in whether the effects of such conditioning processes can be inhibited by people's intentional efforts to reverse the impact of affective stimuli on the conditioned meaning of neutral images. For this purpose, you will be again presented with the words that you have seen before with pictures. Your task is to evaluate the words and try your ABSOLUTE BEST to rate them so as to REVERSE the influence of the positive/negative pictures that the words had been paired with.

Table 1. Percentages of missing values in the evaluation measure due to 100% or 0% correct responses on the memory task as a function of US valence memory (correct vs. incorrect), US valence (positive vs. negative), and instructional set (proactive control, neutral instructions, counteractive control, reversal instructions). Missing values were replaced with individual cell means estimated by linear trends based on existing values within the same cell.

	Experiment 1				Experiment 2			
	US positive		US Negative		US positive		US Negative	
	Memory Correct	Memory Incorrect	Memory Correct	Memory Incorrect	Memory Correct	Memory Incorrect	Memory Correct	Memory Incorrect
Proactive	22.4%	28.1%	18.3%	27.1%	18.3%	21.1%	21.3%	24.2%
Neutral	20.9%	23.3%	23.2%	29.0%	21.3%	24.1%	22.2%	25.3%
Counteractive	25.3%	28.5%	21.3%	25.2%	-	-	-	-
Reversal	-	-	-	-	19.1%	23.1%	23.1%	27.6%

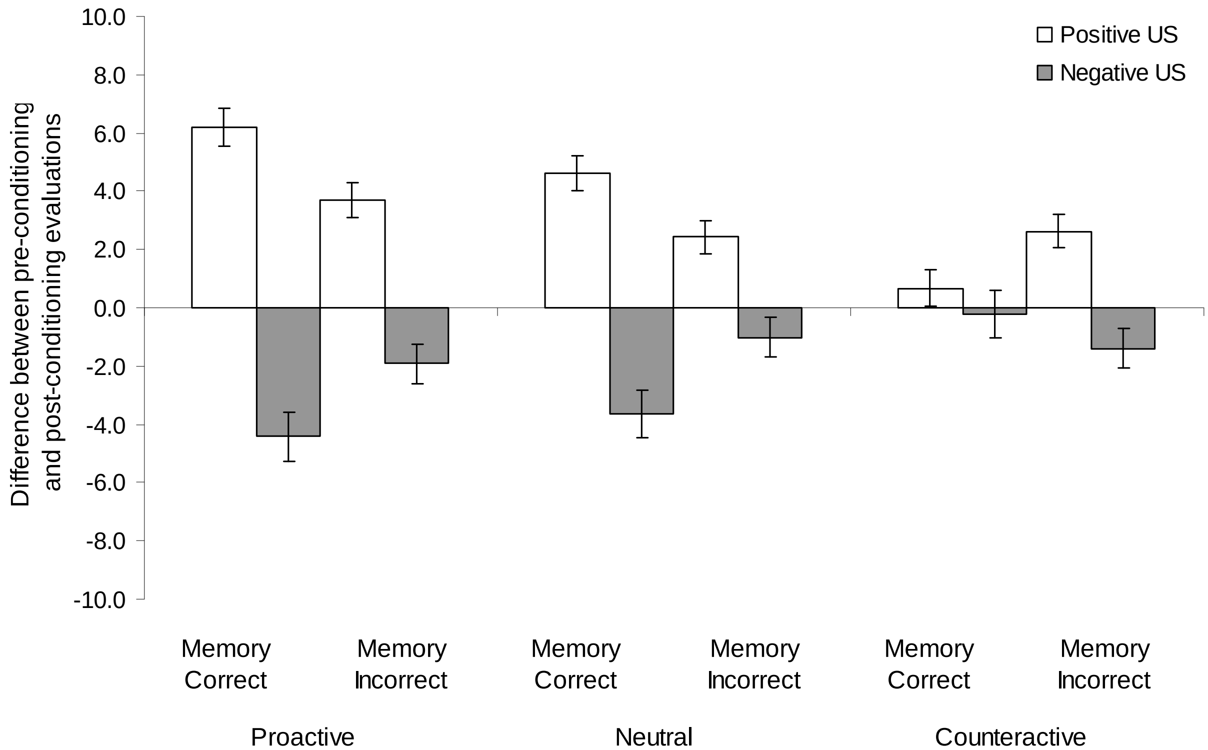

Figure 1. Mean differences between pre-conditioning and post-conditioning CS evaluations as a function of US valence (positive vs. negative), task instructions (proactive vs. neutral vs. counteractive), and item-based memory for US valence (correct vs. incorrect), Experiment 1. Error bars depict standard errors.

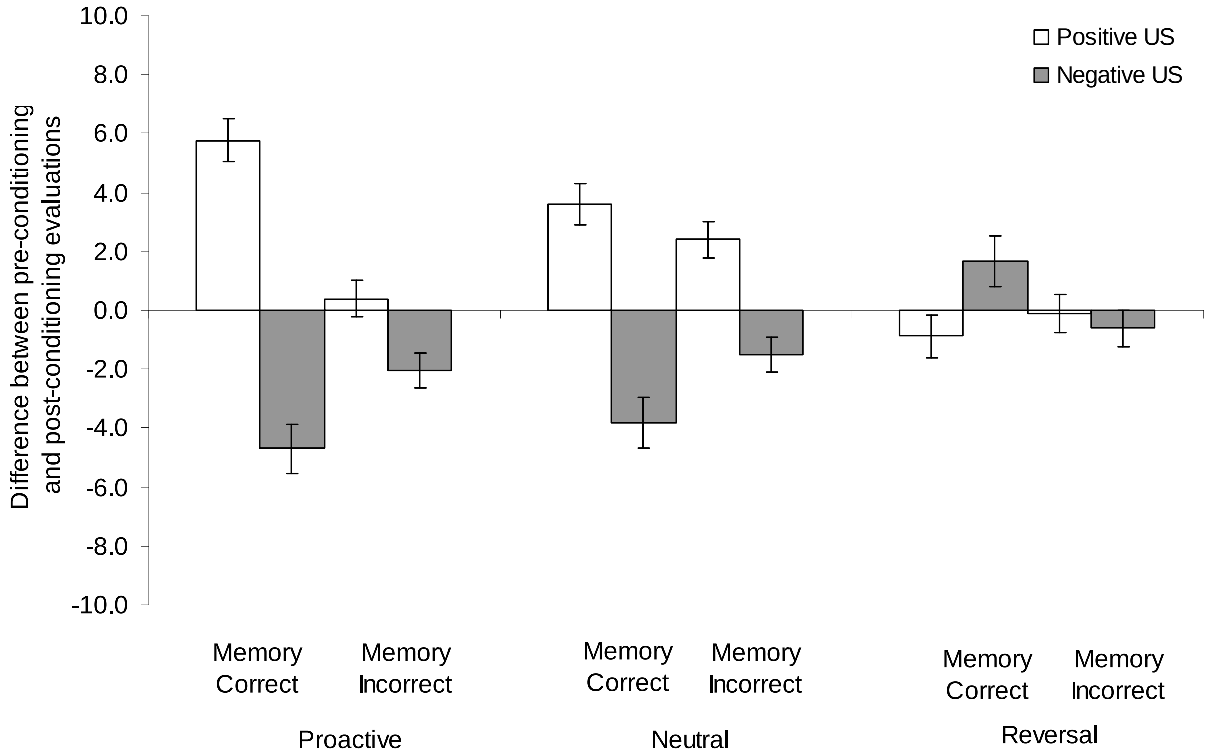

Figure 2. Mean differences between pre-conditioning and post-conditioning CS evaluations as a function of US valence (positive vs. negative), task instructions (proactive vs. neutral vs. reversal), and item-based memory for US valence (correct vs. incorrect), Experiment 2. Error bars depict standard errors.