Context Effects in Evaluative Conditioning of Implicit Evaluations

Riccardo Zanon, Jan De Houwer, and Anne Gast

Ghent University

WORD COUNT: 8.310

Author Note

Riccardo Zanon, Jan De Houwer, and Anne Gast, Department of Experimental Clinical and

Health Psychology, Faculty of Psychology and Educational Sciences, Ghent University.

Correspondence should be addressed to: Riccardo Zanon, Department of Experimental

Clinical and Health Psychology, H. Dunantlaan 2, 9000 Ghent (Belgium). E-mail:

Riccardo.Zanon@UGent.be

Abstract

Propositional models of evaluative conditioning postulate that the impact of stimulus pairings on liking should depend not on the pairings themselves but on what the pairings imply about the relation between stimuli. Hence, context manipulations that change the implications of stimulus pairings should moderate evaluative conditioning. We manipulated context by varying the way in which context cues were paired with affective outcomes while keeping the pairings between target cues and affective outcomes constant. All participants saw one target cue compound that was followed by a positive outcome (XF+) and another target cue compound that was followed by a negative outcome (YG-). In condition Same, each context cue was consistently paired with a positive or negative outcome, regardless of whether it was presented alone or in compound with another cue (A+, B+, AB+; C-, D-, CD-). In condition Opposite, however, a context cue was paired with a certain outcome when presented alone and with an outcome of the opposite valence when presented in a compound with another cue (A+, B+, AB-; C-, D-, CD+). Employing several implicit measures, we assessed the implicit evaluations of the target cues X and Y. In all three studies, the outcome of the measurement procedure differed between conditions. In condition Same, the positively paired cue X was evaluated more positively than the negatively paired cue Y. In condition Opposite, however, this preference was not present. This pattern of results suggests that EC is determined not only by the objective pairings but also by the context in which these pairings occur. Implications for models of evaluative conditioning are discussed.
(262 words)

Keywords: evaluative conditioning, learning, implicit attitudes

Context Effects in Evaluative Conditioning of Implicit Evaluations

Because preferences drive behavior, cognition, and emotion, it is important to understand how those likes and dislikes are formed and how they can be influenced. Research has shown that people do not necessarily evaluate stimuli in a conscious, controlled and intentional manner. As Zajonc (1980) argued in his seminal paper, evaluations can arise also in a spontaneous, uncontrolled, unconscious, efficient and fast manner. We refer to this type of automatic preferences as implicit evaluations, whereas we refer to non-automatic preferences as explicit evaluations.[1] Implicit evaluations have been shown to play a crucial role in many important psychological phenomena including psychopathology (see Roefs et al., 2011), addiction (Wiers & Stacy, 2006), and social interactions (Fazio & Olson, 2003).

Implicit evaluations often arise as the result of repeated experiences. For example, the implicit evaluation of a neutral stimulus (conditioned stimulus, CS) might be changed by repeatedly pairing it with an affectively relevant (positive or negative) stimulus (unconditioned stimulus, US). A change in liking that occurs as the result of pairings of stimuli is typically referred to as evaluative conditioning (EC; De Houwer, 2007; De Houwer, Thomas, & Baeyens, 2001; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). EC can involve changes in both explicit and implicit evaluations. Hence, implicit evaluations that result from the repeated pairing of stimuli are a subset of EC effects, namely those EC effects that involve changes in automatic rather than non-automatic evaluations.

At a mental process level, it is often assumed that EC of both implicit and explicit evaluations is due to the slow and gradual formation of associations in mind, that is, of unqualified links between mental representations. This view is endorsed by several associative (e.g., Baeyens, Eelen, Crombez, & Van den Bergh, 1992) and dual process models (e.g., Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006). Dual process models of evaluations postulate the existence of two separate but interacting mental systems of processing: an associative, impulsive system which is responsible for the formation of implicit evaluations and a propositional, reflective

system which is responsible for the formation of explicit evaluations (e.g., Rydell & McConnell, 2006). The associative system consists of a network of mental representations that are linked via unqualified associations. Associations are typically assumed to result from the direct experience of stimulus pairings. Once an association has been formed between the CS and US representations as the result of CS-US pairings, presentation of the CS not only results in the activation of the CS representation but also, via automatic spreading of activation, in the automatic activation of the US representation. As a result, the originally neutral CS will evoke implicit evaluations that are in line with the valence of the US. Under certain conditions, the automatic activation of the US representation can bias explicit CS evaluations as well and thus lead to EC of explicit evaluations (see Gawronski & Bodenhausen, 2006, for details). In sum, dual process models typically locate the source of EC effects, particularly those involving implicit evaluations, within an associative system.

Recently, single process models have been proposed that reject the existence of an associative system. More specifically, propositional models attribute all instances of EC (and other types of associative learning[2]) to the non-automatic formation of propositions about stimulus relations (e.g., De Houwer, 2007, 2009b; Mitchell, De Houwer & Lovibond, 2009). Propositions are statements about the world that can be true or untrue. They can specify not only that two stimuli are related, but also the way in which the stimuli are related (e.g., that the stimuli simply co-occur or that one stimulus causes another stimulus). Because the impact of stimulus pairings on evaluative and other responses is, according to propositional models, mediated by the formation of propositions, EC can occur only after a proposition about the stimulus pairings has been formed and will depend on the content of the proposition that is formed.

Propositional models also incorporate several other assumptions that generate a number of interesting predictions. First, the formation of propositions is assumed to be a non-automatic process that requires awareness and cognitive resources. Hence, propositional models postulate that EC and other types of associative learning should depend on awareness of the stimulus pairings. Although there have been reports of unaware EC, questions have been raised about the validity of

these findings (e.g., Lovibond & Shanks, 2002; Mitchell et al., 2009), and many studies have failed to find EC effects in the absence of contingency awareness (e.g., Pleyers, Corneille, Luminet, & Yzerbyt, 2007; Stahl, Unkelbach, & Corneille, 2009). Moreover, a recent meta-analysis showed that awareness of the stimulus pairings was by far the biggest moderator of the size of EC effects (Hofmann et al., 2010). Second, it is assumed that propositions can be formed not only on the basis of direct experience but also on the basis of instructions or inferences (De Houwer, 2009). In line with this assumption, De Houwer (2006) showed that merely informing participants about stimulus pairings without actually presenting these pairings is enough to induce EC effects (see also Gast & De Houwer, in press). Third, it has been proposed that after the non-automatic formation of a proposition, the proposition can be stored in memory from which it can be retrieved automatically (e.g., Bar-Anan, De Houwer, & Nosek, 2010; Hughes, Barnes-Holmes, & De Houwer, 2011). This latter assumption allows propositional models to explain EC of implicit evaluations.

In the present paper, we report studies that were inspired by the central idea of propositional models, namely that the effect of CS-US pairings on liking should depend on the content of the proposition about the CS-US relation rather than on the pairings themselves. Importantly, the same objective CS-US pairings could lead to propositions about the CS-US relation with different content depending on the context in which those pairings are presented. For instance, if a neutral stimulus co-occurs with a positive stimulus (e.g., a neutral nonword and a positive word) in a context in which this implies that the neutral stimulus actually has negative properties (e.g., when the nonwords are said to be antonyms of the existing words), such co-occurrences should lead to the formation of negative propositions about the initially neutral stimulus (e.g., the nonword has a negative meaning). As a result, the originally neutral stimulus should become negative even though it was paired with positive stimuli. Recent studies support this prediction. For instance, Förderer and Unkelbach (2011; see also Fiedler & Unkelbach, 2010) showed that neutral faces paired with positive pictures were rated *less* positive than neutral faces paired with negative pictures when participants were told that the depicted people loathed the pictures they were paired with. At a

functional level, studies such as these are important because they show that the effect of stimulus pairings on liking depends on the (verbal) context in which they are presented. At the level of mental process theories, they provide support for propositional models of EC.

It is a different question, however, whether this type of context effects can be found also for EC of implicit evaluations. Assuming that in particular implicit evaluations are determined by associations in memory and that associations in memory are determined by actual stimulus pairings (e.g., Rydell & McConnell, 2006), one could predict that EC of implicit evaluations will reflect stimulus pairings independent of what the context implies about the way in which the CS and US are related. Therefore, observing context effects on EC of implicit evaluations would constrain theories about the determinants of implicit evaluations more than observing context effects on EC of explicit evaluations.

A recent study suggests that contextual information about the way CS and US are related can also moderate EC of implicit evaluations. Peters and Gawronski (2011) presented pictures of four neutral faces (CSs), two paired with a number of positive descriptions (USs+), and two paired with a number of negative descriptions (USs-). After each trial, participants were explicitly told whether the pairing they had just seen was true or false. If the pairing was said to provide false information, participants were asked to "mentally reverse" the pairing (e.g., when someone was described as SAD, they should infer that this person was HAPPY). Employing an affective misattribution procedure (AMP, Payne, Cheng, Govorun & Stewart, 2005) and an affective priming task (Fazio, Jackson, Dunton, & Williams, 1995) to capture implicit evaluations of the CSs, they found that EC effects depended not only on the valence of the USs but also on the relational information (true vs. false) that was presented right after each pairing. In particular, a standard EC effect (i.e., preference for CSs paired with positive USs over CSs paired with negative USs) was found when the pairings were said to provide correct information. However, when pairings were said to provide false information, a reversed EC effect was observed (i.e., more negative rating for CSs paired with positive USs than for CSs paired with negative USs).

In the current research, we wanted to investigate further whether and when EC of implicit evaluations is moderated by the context in which pairings are presented. Unlike earlier studies (Förderer & Unkelbach, 2011; Peters & Gawronski, 2011), we manipulated the context in an indirect way, that is, without verbally instructing participants about how the CS and US are related. Moreover, our context manipulation was embedded in a traditional EC procedure that involved pairings of single stimuli as CSs and USs rather than statements or personality trait descriptions as used by Peters and Gawronski (2011). More specifically, we manipulated relational information by presenting the target CS-US pairings in the context of other pairings that followed a certain rule. A first CS X was always presented in compound with cue F and was always paired with a positive US (i.e., winning a game; XF+). A second CS Y always occurred in compound with cue G and was always paired with a negative US (i.e., losing a game; YG-). In addition to these target compounds, we also presented context cues and manipulated the rule that determined when a context cue was paired with a win or a loss. In condition Same, a context cue was always followed by the same US regardless of whether it was presented alone or in compound with another cue (i.e., A+, B+, AB+, C-, D-, CD-). In condition Opposite, however, a context cue was followed by a different US when it was presented on its own than when it was presented in compound with another cue (i.e., A+, B+, AB-, C-, D-, CD+). Our dependent measure was the implicit and explicit evaluation of the two target stimuli X and Y when presented on their own (and thus in the absence of the cues F and G with which they appeared on compound trials).

Different mental process models of EC on implicit measures make different predictions regarding the outcome of our studies. Regarding the condition Same, all currently available models make the same predictions. Although postulating different mediating processes, in functional terms they would all expect a standard EC effect, that is, a change in implicit evaluation of the CSs that reflects the valence of the USs they co-occurred with. The models do, however, make different predictions about whether this effect will also be found in condition Opposite. A purely associative account of EC of implicit evaluations implies that the change in implicit evaluation is driven simply

by associations that are formed between stimuli that co-occurred. Importantly, in both conditions, the target stimulus X was paired with a positive outcome (XF+) whereas target stimulus Y was paired with a negative outcome. Because the pairings involving X and Y were identical in both conditions, the implicit evaluations of the target stimuli X and Y should not differ between the conditions Same and Opposite. In both cases, X should be evaluated more positively than Y.

A propositional account, on the other hand, would predict different EC effects in condition Same than in condition Opposite. In condition Same, the context pairings (A+, B+, AB+, C-, D-, CD-) imply that a stimulus is paired with the same US when presented on its own and when presented in compound with another stimulus. Based on this information, participants can infer from the XF+ and YG- trials that X on its own will be followed by a positive outcome and that Y on its own would be followed by a negative outcome. Hence our dependent measure (i.e., the evaluation of X and Y in the absence of their paired compound cues F and G) should reflect the valence of the paired USs. In condition Opposite, however, the context pairings (A+, B+, AB-, C-, D-, CD+) imply that a cue is paired with different outcomes when presented alone than when presented in compound. Hence, the XF+ pairings imply that X on its own will be followed by a negative outcome whereas Y would be followed by a positive outcome. Therefore in this case, the evaluation of X and Y when presented on their own, should reflect the opposite of the valence of the US with which they were paired. Provided that participants indeed use the context trials to form propositions about the X-US and Y-US relations, and transfer the rule from the context pairings to the target pairings, propositional models would thus predict that participants will prefer X over Y in condition Same but will prefer Y over X in condition Opposite.

Finally, a third possible scenario for condition Opposite would be a significant reduction of the EC effect compared to condition Same but not a reversal. This result would allow for two theoretical interpretations. A propositional account of such a result would entail that participants form two competing propositions, one based on the experienced co-occurrences and another based on the rule implied by the context pairs. Whereas these two propositions would lead to the same

preferences in condition Same (e.g. "X goes with F and with win" and "X on its own goes with win") this would not be the case for condition Opposite (e.g., "X goes with F and with win" and "X on its own goes with loss"). This might result in a preference for X over Y in condition Same but not in condition Opposite. Such a pattern of results, however, would also be in line with a hybrid account that assumes that both associative and propositional processes contribute to implicit EC effects. Whereas both processes would result in a preference for X over Y in condition Same, they would oppose each other in condition Opposite. For instance, the XF+ pairings would result in an association between X and winning (and thus a liking of X) but in the proposition that X on its own will be followed by a loss (and thus a disliking of X). Hence, in condition Opposite, X will not be liked more than Y.

In sum, if we observe that X and Y are evaluated differently in condition Same than in condition Opposite (either reversed or reduced effects in condition Opposite), this would support the idea that propositional processes (co-)determine EC of implicit evaluations. In addition, it would go beyond the study of Peters and Gawronski (2011) by showing that context effects in EC of implicit evaluations are not limited to the rather atypical EC procedure that they used in which USs were explicitly said to provide a true or false description of the personality of the CSs.

Experiments 1, 2, and 3

In three studies, we examined context effects in the following manner (see De Houwer & Vandorpe, 2010, for a similar approach in research on causal learning). Participants were informed that on each trial of a learning phase, they would see the picture of a slot machine frame on the screen. The slot machine had two displays, one at its top and one at its bottom. On each trial, the top one displayed one or two neutral nonwords, whereas the bottom one displayed an outcome that could either be a win or a loss. Participants were instructed that each nonword or compound of nonwords would be the cause of the respective outcome, that could either be a win or a loss. We refer to different nonwords with different letters (A-G, X and Y), to a win with "+", and to a loss with "-". In condition Opposite, participants experienced the pairings A+, B+, AB-, C-, D-, CD+,

XF+, YH-. In condition Same, participants experienced the pairings A+, B+, AB+, C-, D-, CD-,

XF+, YH-. The stimuli X and F which are in the learning phase always presented in compound with

the stimuli, F and H, respectively, serve as CSs. After the repeated presentation of these pairings,

we measured the implicit evaluations of the single stimuli X and Y, employing a range of implicit

measures, that is, measures of implicit evaluations. More specifically, in Experiment 1, we used the

implicit association test (IAT; Greenwald, McGhee, and Schwartz, 1998), in Experiment 2 the

personalized IAT (Olson & Fazio, 2004), and in Experiment 3 the affective priming task (Fazio,

Jackson, Dunton, & Williams, 1995). We employed a variety of implicit measures to ensure that our

conclusions were not specific to one particular measure.

Method

We report all three studies in the same section because in all three studies we employed the

same procedure during the learning phase. The only major difference between experiments

concerned the nature of the measure of implicit evaluations.

Participants. All participants were students at Ghent University. Fifty-six students

participated in Experiment 1 (mean age = 18.64, *SD* = 1.77; 68% women), 46 in Experiment 2

(mean age = 19.63, *SD*= 2.86; 83% women); and 50 in Experiment 3 (mean age = 19.64, *SD*= 2.75;

76% women). All participants were native Dutch speakers. For their participation in the

experiments, they were given either course credits or four Euros.

Materials.

Materials Learning phase. Each cue and outcome appeared on one of two displays of a slot-

machine picture, which was presented on a computer screen in the size of 26 (height) x 17 (width)

cm. The cues appeared on a display at the top, whereas the outcomes appeared on a display at the

bottom. In all three experiments, eight nonsense words ("BAYRAM", "ENANWAL",

"UDIBNON", "KADIRGA", "LOKANTA", "SARICIK", "FEVKANI" and "NIJARON") were

used as stimuli during the learning phase. The nonwords "LOKANTA" and "FEVKANI" were the

target stimuli, whereas the other nonwords constituted the set of context stimuli. All context and

target stimuli were written in black , upper case letters, and in the font "Arial Black", font size 40. We used nonsense words to avoid any potential pre-experimental association between our stimuli and the outcomes. The positive outcome was the Dutch word for "win" ("winst"), presented in green color, upper case letters, and in the font "Britannic Bold", font size 100.The negative outcome was the Dutch word for "loss" ("verlies"), presented in red color, upper case letters, and in the font "Haettenschweiler", font size 100. The positive outcome was presented together with a pleasant sound (a soft melody) and several pictures of 2 euro coins. The negative outcome was presented together with an unpleasant sound (a loud buzzer) and two cartoon pictures of a sad face.

    Materials Measurement phase. In Experiment 1, we employed an implicit association test designed to measure the evaluation of the target words (IAT, Greenwald et al., 1998). In this IAT, the attribute stimuli referring to the category positive were the Dutch words for "happy" ("gelukkig"), "honest" ("eerlijk"), "pleasant" ("prettig") and "sincere" ("oprecht"). The attribute stimuli for the category negative were the Dutch words for "mean" ("gemeen"), "rude" ("brutaal"), "aggressive" ("agressief") and "deceptive" ("bedrieglijk"). The target stimuli were the cues X and Y ("LOKANTA" and "FEVKANI"). As we were interested in the single nonwords X and Y, we deviated from the standard IAT procedure by using only one stimulus for each of the target categories. To avoid that stimuli were classified only on the basis of simple perceptual features, each target stimulus was presented in four different fonts (lower case Arial Black, upper case Arial Black, lower case Fixedsys, and upper case Fixedsys), resulting in 8 different target stimuli. De Houwer (2006) showed the reliability of this procedural modification and its suitability to capture recent learning.

    In Experiment 2, we employed a personalized IAT (Olson & Fazio, 2004) which differed from the standard IAT in two important ways. First, the attribute labels were "I like" ("Heb ik graag") and "I don't like" ("Heb ik niet graag"), instead of "positive" and "negative", and second, there was no error feedback for the stimuli of the attribute categories. In addition to that, the attribute stimuli were selected in such a way that their valence could easily be judged in a personal

(e.g., non-normative) manner. We used the Dutch words for "holiday" ("vakantie"), "summer" ("zomer"), "gift" ("cadeau"), "party" ("feest"). The attribute stimuli for the negative category were the Dutch words for "war" ("oorlog"), "vomit" ("braaksel"), "accident" ("ongeluk"), "divorce" ("scheiding"). As in Experiment 1, the target stimuli were different perceptual instantiations of the nonwords "lokanta" and "fevkani".

In Experiment 3, we applied an affective priming procedure (Fazio, et al., 1995). As targets we used the positive and negative attributes from Experiment 1. The nonwords X and Y were used as primes.

All tasks were presented on an Intel Core2 Duo PC with a 19" 100 Hz monitor, screen resolution 1280 by 1024 pixels, and implemented using custom made Inquisit 2.0 programs. For each implicit measure, the responses were given by pressing the key A (left key) or the key P (right key) on a standard AZERTY keyboard.

Procedure. At the beginning of each experiment, participants were informed that they would see a series of trials in a slot machine game. They were told that each trial consisted of the presentation of one or two slot-machine words, followed by a winning or losing outcome. Their task was to identify which of the presented slot-machine words predicted a win and which predicted a loss. They were asked to detect and remember this information because it was important for another upcoming task. Finally, we took measures to ensure that participants would register the rule that determined the outcome of the context cues. From the perspective of propositional models, detection of the context rule is a precondition for finding an impact of this context element on EC. In all experiments, participants were therefore told that a hidden rule regulated the presentation of the slot-machine words and their outcome and were asked to try to discover the rule from the series of stimuli. In Experiment 2, we implemented an additional measure by actually informing the participants about the nature of the rule. That is, in condition Opposite, participants were told that words presented in compound result in the opposite outcome than when presented alone. In condition Same, they were told that words presented in compound produce the same outcome as

when presented alone. Although these measures imply that we cannot draw strong conclusions about whether participants can learn these kinds of rules in a spontaneous manner, they are instrumental for determining whether the rules that operate in a certain context can moderate EC of implicit evaluations. In any case, unlike the case in previous related studies (e.g., Förderer & Unkelback, 2011; Peters & Gawronski, 2011), participants were not directly instructed about how CS-US pairings should be interpreted.

After the instruction page, participants could start the learning task by pressing a key. Each trial of the learning task represented a slot-machine game. A trial started with the presentation of the picture of a slot-machine, together with one or two nonword(s) which appeared at the top display position for 2000 ms. Then the outcome (the Dutch word for "win" or "loss") was added on a second display at the bottom of the slot-machine, together with the pleasant or unpleasant sound. The nonword(s) and the outcome remained on the screen together for 3000 ms during which the sound was also present. The next trial started after an inter trial interval of 3000 ms during which the screen was blank and no sounds were played. The sequence of trials presented was the following. First, A+, B+, and AB+ (Condition Same) or A+, B+, and AB- (Condition Opposite) trials were presented twice each, resulting in a block of six trials (Block 1). This block was followed by C-, D-, and CD- (Condition Same) or C-, D-, and CD+ (Condition Opposite) trials, each presented twice, resulting in another block of six trials (Block 2). The series of pairings of Block 1 and Block 2 constituted the context, that varied between conditions. Finally, for both condition Same and Opposite, the target trials XF+ and YH- were presented two times each, resulting in a final block of four trials (Block 3). Block 3 constituted the series of target pairings of our design. This sequence of three blocks of trials was presented five times, adding up to a learning phase of eighty trials. Exclusively in Experiment 2, the trials of Block 1 (the pairings of the stimuli A, B and AB) and of Block 2 (the pairings of the stimuli C, D and CD) were only presented one time per block, instead of twice (like in Experiment 1 and 3). This was an attempt to shorten the learning phase and thus keep participants focused during the whole learning phase. The trial order within

blocks was determined randomly for each participant and block separately, and the assignment of the cues X or Y to the positive or negative outcome was counterbalanced across participants.

After the presentation of all learning trials, participants performed an implicit measure with which we aimed to assess the implicit evaluation of the target stimuli X and Y. In Experiment 1, we applied an IAT (Greenwald et al, 1998). Our IAT consisted of seven phases: a target discrimination phase (B1), an attribute discrimination phase (B2), a combined (targets and attributes) practice phase (B3), a combined (targets and attributes) phase (B4), a reversed target discrimination phase (B5), a reversed combined practice phase (B6), and a reversed combined phase (B7). Phases (B1), (B2), and (B5) served only for practice, whereas Phases (B3), (B4), (B6) and (B7) were the combined test phases that served as measures of implicit evaluation. Before each phase, participants were informed about the assignment of the different categories to the left and right key. During the target and reversed target discrimination phase (B1 and B5), each target stimulus was presented four times, resulting in 32 trials. During the attribute discrimination phase (B2), each attribute stimulus was presented four times, also resulting in 32 trials. Finally, both combined phases consisted of two blocks (practice – B3 and B6 - and test phase – B4 and B7) of 32 trials each. In each block of these phases, each attribute and target stimulus was presented twice. During the reversed combined phase, the assignment of targets to response keys was opposite to the assignment during the other combined phase. The assignment of the target categories to the two response keys (e.g., during Phases B1, B3, and B4, press left for X and right for Y or vice versa) was counterbalanced across participants. Whether the X-positive / Y-negative task (e.g., X assigned to the same key as positive attributes and Y assigned to the same key as negative attributes) was presented before or after the X-negative / Y-positive IAT task (e.g., X assigned to the negative key and Y assigned to the positive key) was thus also counterbalanced across participants. The order of trials in each phase was determined randomly. On each trial, a word was presented in the center of the screen until a valid response (pressing key A or P) was registered. If the response was correct, the next word appeared after 400 ms. If the response was incorrect, a red cross was presented for

400 ms, also in the center of the screen. In this case, the next word was presented 400 ms after the red cross had disappeared.

In Experiment 2, we employed a personalized IAT (see Olson & Fazio, 2004). The procedure was identical to the standard IAT in Experiment 1, except for two modifications that are characteristic of the personalized IAT: Firstly, the attribute category labels were the Dutch words for "I like" instead of "Positive" and "I don't like" instead of "Negative". Secondly, there was no error feedback for the attribute stimuli. (see Materials section).

In Experiment 3, we applied an affective priming procedure. Each trial started with a 500-ms presentation of a fixation star in the center of the screen. Next, one of the cues was presented as prime for 200 ms, immediately followed by the target (stimulus onset asynchrony of 200 ms), both in the center of the screen. The task was to categorize the targets as quickly as possible as positive or negative by pressing one of two keys. Once a response was made, the next trial started after an intertrial interval of 1000 ms. After 16 practice trials, participants completed a block of 128 test trials in which combinations of the primes (2) and the positive (4) or negative (4) targets were realized equally often, and presented in random order.

For exploratory reasons, after performing the implicit measure, participants were also asked to rate, on a 9-point Likert scale, the pleasantness of cues A, C, X, Y, XF, and YH. The ratings for the cues A and C (presented in a random order) were given first, followed by the ratings for the cues X and Y (presented in a random order), and finally the ratings for the compounds XF and YH (also presented in a random order). Subsequently, participants were asked to judge for cues A and C first (presented in a random order), and then for cues X and Y (also presented in a random order) to what extent they were likely to cause "win" or "loss" in a potential new slot-machine game. This question served as a measure of the causal relations learning between stimuli and outcomes. After the causal ratings, participants were asked to report whether they had discovered the context rule and, if so, to write it down. Finally, but only in Experiments 2 and 3, participants received a list of all stimuli (single and compounds) they experienced in the learning phase. They were asked to

indicate, for each of the presented cues (A, B, C, D) and combinations of cues (AB, CD, XF, YH), the outcome it was paired with. This series of questions served as a measure of contingency awareness. After this last questionnaire was completed, participants were thanked and debriefed.

Results

Experiment 1

IAT. IAT scores were calculated using the D600 scoring algorithm (Greenwald, Nosek and Banaji, 2003). Trials in Blocks 3 and 6 of the IAT were entered as mixed practice blocks whereas trials in Blocks 4 and 7 were treated as mixed test blocks. A positive value for the D600 indicates better performance in the X-positive / Y-negative IAT block than in the X-negative / Y-positive IAT block. In other words, a positive value represents implicit evaluations in line with the target pairings (X paired with win and Y with loss). Although it is well known that caution is required when interpreting the absolute value of IAT scores (e.g., Blanton & Jaccard, 2006), we believe that a positive value of our IAT scores can be interpreted as reflecting a preference for X over Y. Unlike the case in most IAT studies, the target stimuli that we used were nonwords that were unknown to the participants before the experiment. Moreover, a nonword functioned equally often as cue X and cue Y. Hence, the direction of the IAT score (positive or negative) can reflect only the acquired relative preference of X compared to Y.

Most importantly, the difference between the two conditions was significant, $t(54) = 2.21$, $p = .031$, $d = 0.58$. In condition Same, the D600 measure was positive ($M = 0.23$, $SE = 0.11$) and significantly different from zero, $t(27) = 2.09$, $p = .046$, $d = 0.40$. In condition Opposite, the D600 measure was negative ($M = - 0.10$, $SE = 0.10$) but did not significantly differ from zero, $t(27) = -.98$, *ns*.

Explicit valence ratings. Our experiments do not allow for strong conclusions regarding EC of explicit evaluations because explicit evaluations were always assessed after the implicit measure. Nevertheless, for exploratory reasons, we did conduct Cue x Condition ANOVAs on the explicit valence ratings for the target cues X and Y. For this experiment and all subsequent experiments, the

relevant mean explicit valence ratings can be found in Table 1. The analysis revealed a significant main effect of cue, $F(1, 54) = 5.40$, $p = .024$, but not of condition, $F(1, 54) = 1.77$, $p = .19$. Furthermore, the interaction between cue and condition was significant, $F(1, 54) = 10.96$, $p < .01$, *ns,* partial $\eta^2 = .17$. A post-hoc t-test showed a significant effect of the cue in condition Same, $t(27) = 4.247$, $p < .01$, $d = 0.80$, but not in condition Opposite, $t(27) = -0.66$, $p = .51$. This pattern of results reflects that of the IAT data. The Pearson correlation between implicit and explicit ratings (for the cues X and Y) was also significant, $r = .52$, $p < 01$.

Causal ratings. Cue x Condition ANOVAs were also conducted on the causal ratings of the target cues X and Y (see Table 2 for all relevant means for Experiments 1-3). The analysis revealed a significant main effect of the cue, $F(1, 54) = 16.50$, $p < .01$, partial $\eta^2 = .23$, but no effect of the condition, $F(1, 54) < 1$. We also found an interaction, $F(1, 54) = 38.84$, $p < .01$, partial $\eta^2 = .42$. A post-hoc t-test showed a significant effect of the cue in condition Same, $t(27) = 10.80$, $p < .01$, $d = 2.04$, but not in condition Opposite, $t(27) = -1.23$, $p = .23$. Hence, in condition Same, participants indicated that, in a potential new game, the cues X and Y would be followed by the outcome predicted by the context rule, whereas in condition Opposite this effect was not present.

. **Experiment 2**

Personalized IAT. As in Experiment 1, we used the D600 scoring algorithm (Greenwald et al., 2003). Data of the participants who, when asked at the end of the experiment, did not remember correctly the pairings of the test compounds XF and YH ($n = 11$), were excluded from the analyses. We did so because, according to the assumptions of propositional models, awareness of the contingencies is a crucial factor in the process of proposition formation. Furthermore, people who did not remember the target pairings may have had a distorted memory and might actually have formed a proposition different from what was implied in the learning phase. Within this sample ($n = 35$), the difference between the two conditions was significant, $t(33) = 2.10$, $p = .044$, $d = 0.68$. In condition Same, the D600 measure was positive ($M = 0.23$, $SE = 0.08$) and significantly different from zero, $t(19) = 2.66$, $p = .015$, $d = 0.60$. In condition Opposite, it was negative ($M = - 0.10$, $SE =$

0.14) but not significantly different from zero, $t(14) = -0.71$, *ns*. Note that the difference between conditions was no longer significant when all participants were included in the analysis, $t(44) = 1.39$, *ns*.

Explicit valence ratings. Also for the valence ratings and the causal ratings reported below, only the analyses for the participants who correctly retrieved the target compound co-occurrences will be reported. The analysis for the test cues X and Y did not reveal a significant effect of cue, $F(1, 33) = 2.62$, $p = .12$, partial $\eta^2=.07$, nor of condition, $F(1.33) = 3.09$, $p = .09$, partial $\eta^2=.08$. The interaction Cue x Condition was not significant, $F(1,33) = 2.62$, $p = .12$, partial $\eta^2=.07$. The Pearson correlation between implicit and explicit ratings of the cues X and Y was marginally significant, r = .33, $p = .05$.

Causal ratings. For the test cues X and Y, the ANOVA did reveal a significant main effect of the cue, $F(1, 33) = 4.24$, $p = .047$, partial $\eta^2=.11$, but not of condition, $F(1, 33) < 1$. The Cue x Condition interaction was significant, $F(1, 33) = 30.58$, $p < .01$, partial $\eta^2=.48$. A post-hoc t-test showed a significant effect of the cue in condition Same, $t(19) = 6.53$, $p < .01$, $d = 1.46$, and a marginal effect in condition Opposite, $t(14) = -2.02$, p = .06. Therefore, in condition Same, participants indicated that, in a potential new game, the cues X and Y would be followed by the outcome predicted by the context rule, whereas in condition Opposite this effect was only marginally significant.

Experiment 3

Affective priming. Trials in which an incorrect response was given (7. 90% of all trials) were discarded. All latencies were pre-processed by discarding latencies that were outliers in an individual's reaction time distribution according to Tukey's (1977) extreme outlier criterion (e.g., latencies above the third quartile plus 3 times the individual's interquartile range). The data of three participants who made more than 50% errors and seemingly reacted to the prime rather than to the target were excluded from the analyses. In line with the analyses for Experiment 2, we excluded the

data of eight participants who did not remember the test cue pairings correctly at the end of the experiment.

In this study, we defined trials in which the target had the same valence as the outcome with which the prime co-occurred in the previous learning phase as congruent trials (e.g., X followed by a positive target and Y followed by a negative target). Trials in which the target had the opposite valence as the outcome with which the prime co-occurred in the previous learning phase are incongruent trials (e.g., X followed by a negative target and Y followed by a positive target). We conducted an ANOVA with congruency (congruent vs. incongruent) as a within subjects factor and context condition as a between subjects factor. The two-way interaction between congruency and condition was marginally significant, $F(1,37) = 3.95$, $p = .05$, partial $\eta^2=.10$. Again, in Condition Same, reaction times for congruent trials ($M = 603$ ms, $SE = 24.2$) were marginally faster than reaction times for incongruent trials ($M = 615$ ms, $SE = 26.7$), $t(20) = 1.938$, $p = .07$, $d = .42$. In Condition Opposite, reaction times for congruent ($M = 559$ ms, $SE = 26.2$) and incongruent ($M = 555$ ms, $SE = 27.5$) trials did not differ $t(17) = -0.81$, ns. Note that the Congruence x Condition interaction was not significant when the data of all participants were included in the analyses, $F(1,45) =1.79$, $p = .19$, partial $\eta^2=.04$.

Explicit ratings. Also for the valence ratings and the causal ratings reported below, only the analyses for the participants who showed a correct retrieval of the target co-occurrences will be reported. The analysis for the test cues X and Y revealed a significant effect of cue, $F(1, 37) = 5.80$, $p = .021$, but not of condition, $F < 1$. We did not find an interaction $F(1, 37) = 2.81$, $p =. 10$, partial $\eta^2=.071$. The Pearson correlation between implicit and explicit ratings of the cues X and Y was not significantly different from zero, $r = .15$, *ns.*

Causal ratings. For the test cues X and Y the ANOVA did reveal a main effect of the cue, $F(1, 37) = 17.49$, $p < .01$, partial $\eta^2=.32$, and a marginal effect of condition, $F(1, 37) = 3.50$, $p = .07$, partial $\eta^2=.09$. We found a significant Cue x Condition interaction, $F(1, 37) = 13.67$, $p < .01$, partial $\eta^2=.27$. A post-hoc t-test showed a significant effect of the cue in condition Same, $t(20) = 5.78$, p

$< .01$, $d = 1.26$, but not in condition Opposite, $t(17) < 1$. Therefore, in condition Same, participants indicated that, in a potential new game, the cues X and Y would be followed by the outcome predicted by the context rule, whereas in condition Opposite this effect was not present.

General Discussion

The main aim of our research was to test the influence of context variables on EC of implicit evaluations. In three experiments, we compared the implicit evaluations of two target stimuli (X and Y) in two conditions that differed only with regard to the context in which the stimuli were presented. In both conditions, the test stimuli X and Y were consistently presented in compound with another neutral cue (F or H) and consistently followed by the same, positive or negative outcome (XF+, YH-). The context was manipulated by presenting the target cues intermixed with context cues that were paired with outcomes according to a certain rule. In condition Same, each of the context cues was consistently paired with the same outcome, regardless of whether it was presented alone or in compound with another cue (A+, B+, AB+; C-, D-, CD-). In condition Opposite, each of a series of other context cues was paired with a certain outcome when presented alone and with the opposite outcome when presented in a compound with another cue (A+, B+, AB-; C-, D-, CD+). Given that in both conditions the target stimulus X always co-occurred with a positive outcome and the target stimulus Y always co-occurred with a negative outcome, purely associative models of implicit attitudes that take into account only the pairings would predict no difference between the conditions. Results showed, however, that the two conditions differed significantly in terms of implicit evaluations towards the target cues. Across all three experiments, various implicit measures always reflected a preference for the positively paired cue X in condition Same. This preference was always absent in condition Opposite. The current results thus demonstrate that EC of implicit evaluations can be moderated by rules that are implied by the context.

We did not, however, find a reversed EC effect in the condition Opposite. The latter finding stands in contrast to recent results by Peters and Gawronski (2011, Experiments 1 & 2) who found reversed EC of implicit evaluation when participants were verbally instructed after CS-US pairings to mentally reverse the valence of the US. More specifically, the authors provided participants with information about the personality traits of four fictitious people. Persons A and B were paired in 75 % of the cases with positive USs whereas persons C and D were paired in 75 % of the cases with negative USs. The participants' task was, for each pairing, to guess whether the information provided about each of the four individuals was true or false. Information provided by pairings involving persons A and C was said to be true for the frequent valence and false for the infrequent valence, whereas the information provided by pairings involving persons B and D was said to be false for the frequent valence and true for the infrequent valence. Whereas A was preferred over C (i.e., standard EC effect), B was liked less than D (i.e., reversed EC effect). A possible explanation for why we did not find such a reversed EC effect might be that we did not directly instruct participants about how to interpret the CS-US pairings. Instead, our manipulation was more indirect and could have succeeded only if (a) participants noticed the context rule that governed the pairings involving the context cues and (b) they actually applied those rules to make inferences about the USs that were paired with the target cues X and Y when these cues were presented on their own. Although we took measures to ensure that participants did encode the context rule, it might be that some participants did not actually use the rule to make inferences about the target cues X and Y. From this perspective, it is interesting to note that in a related study conducted at our lab, but with a simplified rule and the explicit task to state conclusions about the target stimuli (please see Footnote 3 for details), we did find significant reversed EC. This study was more similar to the studies of Peters and Gawronski in that it was much clearer to participants how the CS-US pairings had to be interpreted.

It is also interesting to note that Peters and Gawronski did not find a full reversal when participants were only informed about the correct interpretation of the CS-US after all the pairings

had been presented, instead of after each trial. It is therefore possible that the presence of the relational information at the time of encoding represents a crucial factor in obtaining a reversed EC effect. In our studies, particularly in condition Opposite, different participants might have discovered the context rule at different moments during the learning phase. Consequently, the relational information embedded in the context may have been available at the time of the encoding of the CS-US pairings for only some of the participants. This could explain the lack of a reverse EC effect in our studies and its presence in studies in which the encoding of both valenced and relational information is ensured to happen at the same time. Again, note that we did find a reversed EC effect in a study that resembled the Peters and Gawronski studies (see Footnote 3). Because a simpler rule was used in that experiment than in Experiments 1-3, participants might have become aware of the rule sooner and have already used this information during the first presentation of the target cues X and Y.

Regardless of the exact reasons for why we did not observe a reversal of the EC effect in Experiments 1-3, the fact that we did find a modulation of EC of implicit evaluations has important theoretical implications. First, as we explained in the introduction, purely associative models postulate that the experience of pairings should unequivocally lead to the formation of associations between the mental representations of those stimuli. Because implicit evaluation is assumed to be determined only by associations in memory, one should observe an EC effect that is determined by the experienced pairings in a way that is independent of what the context implies about the meaning of those pairings. The fact that EC was significantly moderated by the properties of the context in which these pairings were experienced does not fit well with these ideas.

According to propositional models, on the other hand, the effect of CS-US pairings on liking is mediated by the formation of a proposition about the CS-US relation. As the content of this proposition depends not only on the CS-US pairings but also on the context in which these pairings occur, aspects of the context should be able to moderate the effect of stimulus pairings on liking. Hence, we predicted EC effects on implicit measures that are moderated by the context rule that is

implemented in the pairings of other stimuli. The most straightforward prediction from these

models would have been that participants inferred a reversed relation between CS and US in

condition Opposite, which according to a purely propositional model, should have led to a

significantly reversed EC effect. The fact that we did not find significant reversal speaks against this

interpretation. However, propositional models can account for the lack of a reversed EC effect in

condition Opposite if certain additional assumptions are made. As we explained in the introduction,

one could assume that participants formed two propositions that are relevant for the evaluation of

the target stimuli: one determined by the context and one determined by the co-occurrences. Both

propositions could cancel each other out, resulting in a null implicit evaluation. In sum, the current

results do not refute a strict propositional model but additional assumptions are needed before such

a model can explain the exact pattern of results.

Importantly, the current results can also be explained by a hybrid associative-propositional

account of EC of implicit evaluations that assumes that both associative and propositional processes

contribute to such EC effects. As we explained in the introduction, associative and propositional

processes might cancel each other out in condition Opposite, which would explain the null-finding

in this condition. On the other hand, there might be circumstances in which propositional processes

dominate implicit evaluations, for instance, when unambiguous information is presented about how

CS-US pairings should be interpreted. This would explain why Peters and Gawronski (2011) did

find a reversed EC effect when participants were instructed to mentally reverse the meaning of CS-

US pairings (also see Footnote 3). A hybrid account also fits well with the observation of Peters and

Gawronski that verbal information has a stronger effect when presented immediately after a CS-US

pairing. In such cases, propositional inferences could overrule the formation of associations based

on CS-US pairings and thus dominate implicit evaluations (see Peters & Gawronski, 2011, for

details). Note that these ideas are in line with recent dual process models such as the Associative

and Propositional Evaluation (APE) model (Gawronski & Bodenhausen, 2006, 2011). Regardless of

whether one favors a strictly propositional account or a hybrid account of our results, it is important

to note that our results confirm the idea that propositional processes do (co-)determine EC of implicit evaluations (also see Peters & Gawronski, 2011).

To conclude, our results show that EC of implicit evaluations depends not only on the objective CS-US pairings but also on elements in the context that specify the implications of these CS-US pairings. Although we are not the first to observe such context effects (i.e., Peters & Gawronski, 2011; see Förderer & Unkelbach, 2011, for a study on EC of explicit evaluations), we are the first to show that these context effects are not restricted to explicit verbal instructions about how CS-US pairings should be interpreted. Our results suggest that context effects in EC of implicit evaluations might be widespread and confirm the idea that propositional processes do contribute to EC of implicit evaluations.

References

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. doi: 10.1016/0005-7967(92)90136-5.

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.

De Houwer, J. (2006). Using the Implicit Association Test does not rule out an impact of conscious

propositional knowledge on evaluative conditioning. *Learning and Motivation, 37*, 176-187.

doi: 10.1016/j.lmot.2005.12.002.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The*

Spanish Journal of Psychology, 10, 230-241. http://www.ucm.es/info/psi/docs/journal/

De Houwer, J. (2009a). How do people evaluate objects? A brief review. *Social and Personality*

Psychology Compass, 3, 36-48. doi: 10.1111/j.1751-9004.2008.00162.x.

De Houwer, J. (2009b). Conditioning as a source of liking: There is nothing simple about it. In

Wänke, M. (Ed.). *Frontiers of Social Psychology: The Social Psychology of Consumer*

Behavior. New York: Psychology Press. http://www.psypress.com/

De Houwer, J. (2009c). The propositional approach to associative learning as an alternative for

association formation models. *Learning & Behavior*, 37, 1-20. doi: 10.3758/LB.37.1.1.

De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional

approach and vice versa. *Perspectives on Psychological Science, 6*, 202-209.

doi:10.1177/1745691611400238.

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.

De Houwer, J. & Vandorpe, S. (2010). Using the Implicit Association Test as a measure of causal

    learning does not eliminate effects of rule learning. *Experimental Psychology*, *57*, 2010, 61-

    67. doi: 10.1027/1618-3169/a000008.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition: Their meaning and use.

    Annual Review of Psychology, 54, 297-327. doi: 10.1146/annurev.psych.54.101601.145225.

Fazio, R. H., Jackson, J. R., & Dunton, B. C., Williams, C. J. (1995). Variability in automatic

    activation as an unobstrusive measure of racial attitudes: A bona fide pipeline? *Journal of

    Personality and Social Psychology, 69*, 1013-1027. doi: 10.1037//0022-3514.69.6.1013.

Fiedler, K., & Unkelbach, C. (2010). Evaluative conditioning depends on higher-order encoding

    processes. *Cognition and Emotion*. doi: 10.1080/02699931.2010.513497.

Fiedler, K., & Unkelbach, C. (in press). Creating contrastive CS-US relations reverses evaluative

    conditioning effects. *Journal of Experimental Social Psychology.

    http://www.journals.elsevier.com/journal-of-experimental-social-psychology/*

Förderer, S. & Unkelbach, C. (2011). Hating the cute kitten or loving the aggressive pit-bull: EC

    effects depend on CS – US relations. *Cognition and Emotion.* doi:

    10.1177/1948550611398413.

Gast, A. & De Houwer, J. (in press). Evaluative conditioning in the absence of directly experienced

    CS-US pairings. *Quarterly Journal of Experimental Psychology.

    http://www.tandf.co.uk/journals/pp/02724987.html*

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. doi: 10.1016/B978-0-12-385522-0.00002-0.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464-1480. doi: 10.1037/0022-3514.74.6.1464.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216. doi: 10.1037/0022-3514.85.2.197.

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.

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorising in implicit attitude research: Propositional and behavioral alternatives. *Psychological Record, 61*, 465–496.* http://thepsychologicalrecord.siuc.edu/

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

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.

Olson, M., & Fazio, R. H. (2004). Reducing the Influence of Extrapersonal Associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality & Social Psychology, 86(5),* 653-667, doi: 10.1037/0022-3514.86.5.653.

Payne, B. K., Cheng, C. M., Govorun, O., Stewart, B. D. (2005). An Inkblot for Attitudes: Affect Misattribution as Implicit Measurement. *Journal of Personality and Social Psychology*, 89, 277-293. doi: 10.1037/0022-3514.89.3.277.

Peters, K. R., Gavronski, B. (2011). Are We Puppets on a String? Comparing the Impact of Contingency and Validity on Implicit and Explicit Evaluations. *Personality and Social Psychology Bullettin*, 37 (4), 557-569. doi: 10.1177/0146167211400423.

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

Roefs, A., Huijding, J., Smulders, F. T. Y., MacLeod, C. M., De Jong, P. J., Wiers, R. W., & Jansen, A. T. M. (2011). Implicit measures of association in psychopathology research. *Psychological Bulletin, 137*, 149-193. doi: 10.1037/a0021729.

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. doi: 10.1037/0022-3514.91.6.995.

Stahl, C., Unkelbach, C. & Corneille, O. (2009). On the respective contributions of awareness of US valence and US identity in attitude formation through evaluative conditioning. *Journal of Personality and Social Psychology, 97*, 404-420. doi: 10.1037/a0016196.

Tukey, J. W. (1977) *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.

Verguts, T., & Fias, W. (2009). Similarity and rules united: Similarity- and rule-based processing in a single neural network. *Cognitive Science*, *33*, 243-259. doi: 10.1111/j.1551-6709.2009.01011.x

Wiers, R. W., & Stacey, A. W. (2006). Implicit cognition and addiction. *Current Directions in Psychological Science, 15*, 292-296. doi: 10.1111/j.1467-8721.2006.00455.x.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist,*

35, 151-175. doi: 10.1037/0003-066X.35.2.151.

Footnotes

[1] Note that we use the term implicit evaluation to refer to a behavioural phenomenon (i.e., the fact that a stimulus evokes an evaluative response automatically) rather than to a type of mental representation (e.g., implicit attitudes; see De Houwer, 2009a, 2011). Although the study of implicit evaluations imposes limits on theories about the mental representations that mediate implicit evaluation, it should not rely on a priori assumptions about the nature of those representations (e.g., the assumption that implicit and explicit attitudes are separate entities).

[2] Associative learning is defined as changes in behaviour that are due to relations between events (e.g. De Houwer, 2009c).

[3] In a further experiment, we employed a procedure similar to that used in condition Opposite of Experiments 1, 2, and 3. The differences with Experiments 1-3 are as follows. First, because we were interested mainly in obtaining a reversed EC effect in condition Opposite, only this condition was realized. The second important change was made in order to simplify the implementation of the context rule. All stimulus compounds consisted of the same cue R in combination with one other stimulus. Participants were told that cue R reverses the outcome of the cue with which it is paired, and were then presented with A+, AR-, B-, BR+, XR+ and YR-. This rule is less complex than the one used in Experiments 1-3 and could be demonstrated with a smaller number of different stimuli. A third important change was that participants were asked to make causal judgements about X and Y before administrating an IAT that was identical to the IAT of Experiment 1. We implemented this change to give participants the opportunity to relate the context rule to the target cues X and Y and thus to form the propositions about the target cues that were implied by this information. The D600 measure was coded and scored as in Experiment 1. As predicted, the D600 measure for the single cues X and Y was negative ($M = -0.15$, $SE = 0.49$) and significantly different from zero, $t(80) = -$

2.98, $p < .01$, $d = 0.33$. In terms of explicit evaluation, the preference for the negatively paired cue was significantly different from zero, ($M = -1.38$, $SE = 0.49$) $t(80) = -2.85$, $p < .01$, $d = 2.85$. This result confirms the findings of Peters and Gawronski (2011) and shows that reverse EC effects of implicit evaluation are not specific to the manipulation (verbal information on how to encode the pairings vs. rules implied by context pairings) and stimuli (adjectives describing a person vs. nonwords that result in winning or losing) that they used. The causal ratings indicated that the cue that was in the losing (winning) compound was deemed, when on its own, more likely to cause win (loss) in a potential new game, $t(80) = -8.51$, $p < .01$, $d = 0.95$. Finally, we observed a positive correlation between implicit and explicit ratings of the target cues, $r = .53$, $p < .01$.

Table 1. Means (and standard errors) of the explicit valence ratings for the target cues in Experiments 1 – 3. X was always paired with a positive outcome whereas Y was always paired with a negative outcome.

Experiments	Cue	Condition Same	Condition Opposite
Experiment 1	X	6.68 (0.48)	5.07 (0.48)
	Y	3.21 (0.50)	5.68 (0.50)
Experiment 2	X	6.10 (0.51)	5.60 (0.59)
	Y	3.90 (0.49)	5.60 (0.56)
Experiment 3	X	5.86 (0.40)	4.95 (0.43)
	Y	4.00 (0.40)	4.61 (0.43)

Table 2. Means (and standard errors) of the causal ratings for the target cues in Experiment 1 – 3

    A high score indicates that the cue is deemed to cause a win in a potential new game. X was

    always paired with a positive outcome whereas Y was always paired with a negative

    outcome.

Experiments	Cue	Condition Same	Condition Opposite
Experiment 1	X	7.82 (0.54)	4.64 (0.54)
	Y	1.89 (0.50)	5.89 (0.50)
Experiment 2	X	7.30 (0.60)	3.00 (0.69)
	Y	2.05 (0.54)	5.40 (0.63)
Experiment 3	X	6.86 (0.59)	3.67 (0.64)
	Y	2.33 (0.50)	3.39 (0.54)