UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Ingendahl, M., Woitzel, J., Propheter, N., Wänke, M., & Alves, H. (2023). From Deviant Likes to Reversed Effects: Re-Investigating the Contribution of Unaware Evaluative Conditioning to Attitude Formation. *Collabra: Psychology*, *9*(1). https://doi.org/10.1525/collabra.87462

Social Psychology

From Deviant Likes to Reversed Effects: Re-Investigating the Contribution of Unaware Evaluative Conditioning to Attitude Formation

Moritz Ingendahl[1][a], Johanna Woitzel[1], Nadja Propheter[2], Michaela Wänke[2], Hans Alves[1]

[1] Department of Psychology, Ruhr University Bochum, Germany, [2] Department of Psychology, University of Mannheim, Germany

Keywords: evaluative conditioning, awareness, memory, valence, individual differences

https://doi.org/10.1525/collabra.87462

Collabra: Psychology

Vol. 9, Issue 1, 2023

Evaluative Conditioning (EC) is the change in liking of an object due to its mere pairing with a positive/negative stimulus. A central question in EC research is whether EC effects also emerge without awareness of the stimulus pairings. This is often tested by asking participants after the conditioning whether an object had been paired with positive or negative stimuli. If participants' answers in these memory measures mismatch with the US valence (e.g., "positive" response when an object was paired with a negative stimulus), the pairings are classified as unaware. The last decade of EC research has found mostly no evidence for unaware EC, and sometimes even reversed unaware EC effects when using such memory measures. The present work demonstrates that such valence memory measures underestimate unaware EC effects due to differences between the normed and the subjective US valence. In two simulation studies, a re-analysis of previous studies, and four preregistered experiments (N = 502), we assess when this bias is more or less severe, depending on common procedural variations in EC experiments. We also propose an improved memory measure of aware/unaware EC. Yet, even when the bias was reduced in the corrected measure, no evidence for unaware EC could be found. Overall, our research shows that unaware EC may be difficult to detect with valence memory measures. Also, they support current memory-based and propositional EC accounts.

Attitudes – summary evaluations of objects of thought (Vogel & Wänke, 2016) – are an integral part of people's identity (Smith & Hogg, 2008) and a key predictor of future behavior (Ajzen, 1991). Due to the centrality of attitudes to people's lives (Bohner & Dickel, 2011), it is essential to understand how people come to their attitudes. One intriguing insight from previous decades of attitude research is that people develop attitudes towards people and objects in their environment based on their mere co-occurrence with other positive or negative stimuli. This effect is known as Evaluative Conditioning (EC; De Houwer, 2007; De Houwer et al., 2001; Hofmann et al., 2010; Hütter, 2022), defined as the change in liking toward a neutral conditioned stimulus (CS) due to the pairing with a positive/negative unconditioned stimulus (US). For example, the mere co-occurrence of an unknown stranger (CS) with a cute dog (positive US) might lead to a more positive attitude towards the stranger.

Evaluative Conditioning is relevant to all domains of attitude formation because people encounter stimulus co-occurrences at any time and place (Alves et al., 2020). In the social domain, people's attitudes towards individuals or groups can be influenced by mere co-occurrences in various contexts, such as the workplace, neighborhoods, or by co-occurring with other groups (Baeyens et al., 1990, 1992; Koranyi et al., 2013; Olson & Fazio, 2006; Walther, 2002; Walther et al., 2005). For example, social groups can acquire negative attitudes from stigmatized groups they co-occur with, also known as "sins-of-the-father" effect, or the message bearer may acquire the message's negativity, known as the "kill-the-messenger" effect (Walther et al., 2005). Co-occurrence-based procedures are also used to decrease negative attitudes and bias toward social groups (Forscher et al., 2019; French et al., 2013).

A core question in EC research is the extent to which EC effects require awareness of the pairings of CS and US

a Corresponding author: moritz.ingendahl@ruhr-uni-bochum.de. https://twitter.com/moritzingendahl https://orcid.org/0000-0002-2124-0754

(Hofmann et al., 2010; Stahl et al., 2009, 2016; Sweldens et al., 2014; Walther & Nagengast, 2006). Whereas earlier research argued that EC effects emerge best (or even exclusively) without awareness of the pairings (Baeyens et al., 1990; Olson & Fazio, 2006; Walther & Nagengast, 2006), more recent research has shifted this claim to the opposite (Moran et al., 2023). Current evidence shows that EC effects substantially benefit from or even require awareness (Högden et al., 2018; Stahl et al., 2016).

In the present paper, we identify a methodological problem in a common procedure to quantify aware and unaware EC effects. Specifically, we show that this procedure is biased against unaware EC effects. We identify the conditions that make this bias more or less likely, and further test a simple yet effective way to control it. With this improved methodology, we also examine to which extent EC effects require awareness of stimulus pairings.

Aware and Unaware Evaluative Conditioning

The role of aware and unaware EC in attitude research has been thoroughly discussed in previous reviews (Corneille & Stahl, 2018; Hofmann et al., 2010; Hütter, 2022; Sweldens et al., 2014), and thus, we only provide a short summary here.

One reason why the awareness debate is so prominent in EC is that different process explanations make vastly different predictions regarding the contribution of awareness. According to memory-based accounts (Gast, 2018; Stahl & Aust, 2018), people must consciously retrieve the US valence and apply it to the judgment to show EC effects. These accounts emphasize that EC effects require awareness of the stimulus pairings. Propositional accounts argue that people infer a relation between the CS and the US (e.g., “CS predicts the US”, “CS causes the US”, “CS is the opposite to the US”). Awareness of the pairings is a necessary precondition for inferring the CS-US relation and applying it to the judgment. A more lenient prediction comes from dual-process accounts (Gawronski & Bodenhausen, 2018). Here, EC effects can be due to associative links between CS and US that do not require awareness and propositional reasoning depending on the inferred CS-US relation that require awareness. According to the implicit misattribution account, awareness may actually reduce EC effects (March et al., 2019). Here the assumption is that people misattribute the affect elicited by the US to the CS. Misattribution works best if people are unaware of the affect’s source, and thus, if people are unaware of the pairings.

Because awareness plays a crucial role in all of these theories[1], examining to which extent EC effects require awareness is a straightforward test for these theories. In that regard, current evidence clearly favors memory-based and propositional EC accounts (Corneille & Stahl, 2018; Hofmann et al., 2010; Hütter, 2022; Sweldens et al., 2014). In contrast to earlier findings that EC effects are stronger without awareness (Baeyens et al., 1990; Olson & Fazio, 2006; Walther & Nagengast, 2006), current research suggests that EC effects even require awareness (e.g., Högden et al., 2018; Moran et al., 2021; Stahl et al., 2016). These insights followed several methodological improvements in the last decades, which we will discuss in the following.

Testing the Contribution of (Un)Awareness in EC

Corneille and Stahl (2018) distinguish between experimental and correlational approaches to test the influence of awareness. Experimental approaches usually target participants’ awareness of the CS-US pairings *during the conditioning* through manipulations ranging from parafoveal presentations (Dedonder et al., 2013) and continuous flash suppression (Högden et al., 2018) to subliminal presentation times (Heycke et al., 2018; Heycke & Stahl, 2020; Stahl et al., 2016). These studies have mostly shown no stable evidence for unaware EC.

Correlational approaches, on the other hand, assess awareness of the US valence *after the conditioning* by employing memory measures for the pairings. In such post-conditioning tests, participants recall which CSs were paired with which USs. One primary advantage of such correlative approaches is that they do not disturb the incidental nature of the stimulus pairings during the conditioning. A second advantage is that they are relatively economical. Post-conditioning memory measures do not require specific laboratory equipment and can thus easily be assessed in online studies or field experiments. As for the experimental tests, correlational approaches have often found no evidence of unaware EC, even though some studies indeed found such effects (Hütter et al., 2012; Mierop et al., 2019; Waroquier et al., 2020, p. @242102).

One prominent correlational method to assess the contribution of awareness are valence awareness measures (VAMs), as used by Stahl and colleagues (2009). Here, participants are presented with each CS from the conditioning phase. For each CS, participants are tasked to indicate the valence of the US the CS had been paired with. The logic of this memory test is as follows: If a CS was presented with a positive US and the participant indicates that the CS was paired with a positive US, the respective pairing is classified as aware. If the participant indicates that the CS was paired with a negative US, the pairing is classified as unaware. Likewise, if a CS was paired with a negative US and the participant responds with “positive” (“negative”) in this task, the pairing is classified as unaware (aware).

1 Another reason why studying the role of awareness is important is that aware EC effects are difficult to distinguish from demand effects (for a review, see Corneille & Lush, 2023). According to the propositional account, participants might start liking a CS because they infer that stimulus co-occurrences are a valid source for stimulus evaluation. However, participants might also infer the evaluation the experimenter expects from them (De Houwer, 2006; Walther et al., 2011).

In the most basic procedure, participants are merely given two choice options (i.e., "positive US" and "negative US"). As this method cannot directly assess whether participants guess, some researchers also provide a "don't know" option (Stahl et al., 2009). Yet, responses with "don't know" are also sometimes aggregated with responses classified as unaware (Förderer & Unkelbach, 2013). Other researchers reversed cue and target in the memory test and asked participants to identify the CS that had been paired with positive or negative USs (Bar-Anan et al., 2010; Moran et al., 2021).

Similar to the evidence from other approaches, studies with VAMs often show that EC effects only emerge if participants indicate the "correct" US valence (for a specific CS; see Pleyers et al., 2007). Thus, studies with VAMs offer substantial evidence that EC requires awareness (Alves et al., 2020; Alves & Imhoff, 2023; Förderer & Unkelbach, 2013; Sweldens et al., 2014). In fact, the unaware EC effect even seems to reverse sometimes: For example, in all four experiments of Stahl et al. (2009), the EC effect for unaware pairings was negative, and even significant in one experiment. The same was the case in later research, showing strong reversed EC effects for unaware pairings (Alves & Imhoff, 2023; Förderer & Unkelbach, 2013; Halbeisen et al., 2014; Waroquier et al., 2020). In the following sections, we argue that these reversed unaware EC effects deserve further consideration and point to a severe limitation of VAMs.

(Dis)-Advantages of Valence Awareness Measures

VAMs offer significant advantages but also some disadvantages compared to traditional[2] correlative awareness measures. One advantage compared to recognition memory tests (Walther & Nagengast, 2006) is that they capture only the most relevant information, namely, to what extent people remember the *valence* of the US. According to current EC theories (Gast, 2018; Stahl & Aust, 2018), it is not necessary to remember the *identity* of the US (i.e., which specific US was paired with the CS), but only the valence (i.e., whether it was paired with a positive or negative stimulus). Also, VAMs allow classifications as aware or unaware on the *item level* (instead of participant level), which is indispensable for accurate awareness assessment (Pleyers et al., 2007).

One conceptual problem is the extremely narrow definition of awareness. Essentially, awareness is equated with objective performance in a memory task, which can be considered a strong assumption (Timmermans & Cleeremans, 2015). For instance, people do not necessarily consciously recollect items and contexts (Jacoby, 1991), or they may show above-chance memory performance despite low confidence and a "random guessing" subjective state (Larzabal et al., 2018). Therefore, other researchers have adapted and extended the VAM to also study the subjective feeling of consciousness (Waroquier et al., 2020), which we will come back to in the General Discussion. For the remainder of this article, however, "awareness" in the context of the VAM is treated as an objective performance in a memory task.

On the methodological side, one substantial problem of the VAM is measurement error (Shanks, 2017). The more an awareness measure suffers from measurement error, the worse it differentiates between aware and unaware EC. At extreme levels, an awareness measure that only captures noise reveals equally strong "aware" and "unaware" EC effects.

Another important limitation already discussed by Stahl et al. (2009) is that the correlative assessment also allows reverse causality. One particular type of reverse causality in VAMs is the affect-as-information heuristic (Schwarz, 2011). People may use their attitudes towards the CS as a cue for whether a CS was paired with a positive or a negative US (Bar-Anan & Amzaleg-David, 2014; Hütter et al., 2012). Depending on the actual US valence, these pairings are counted as aware or unaware, leading to an overestimation of aware EC and an underestimation of unaware EC. For example, suppose a CS was paired with a negative US, but the conditioning is ineffective and the CS is evaluated positively. If participants use their CS evaluation as a cue for the answer in the VAM, they would respond with "positive". Because pairings are classified as unaware when participants' responses in the VAM mismatch the US valence, the ineffective pairing is misclassified as unaware. However, if the CS had actually been paired with a positive US, the pairing would be classified as aware EC. Because the CS evaluation is then actually in line with the US valence, the pairing positively contributes to the aware EC effect. Accordingly, affect-as-information would lead to an overestimation of aware EC and an underestimation of unaware EC (see Hütter et al., 2012, for a detailed explanation and simulation).

In summary, previous literature has discussed some limitations of valence awareness measures. Yet, we believe that one primary source of bias has been overlooked so far.

The Role of Subjective US Valence

Valence awareness measures make an implicit but central assumption: Positive USs are positive, and negative USs are negative to all participants. Specifically, VAMs assume that if a participant indicates a US valence opposite of the experimental factor level (e.g., "positive US valence" for a CS conditioned negatively), this is due to the participant being unaware of the US valence (see Table 1) and not because this is the US valence subjectively experienced by the participant.

Perfect correspondence between subjective US valence and the experimental valence manipulation is a strong assumption. Many EC studies simply select highly positive/

2 We will discuss more recent and more sophisticated methods than the VAM in the General Discussion.

Table 1. When Valence Awareness Measures Lead to Biased Estimates

Normed US Valence	True Awareness	Subjective US Valence	Response in the VAM	VAM Classification	Influence on EC Effect	Corrected VAM Classification
Positive	Aware	Positive	"Positive"	Aware	--	Aware
Positive	Aware	Negative	"Negative"	Unaware	**Reversed Unaware EC**	**Aware**
Negative	Aware	Positive	"Positive"	Unaware	**Reversed Unaware EC**	**Aware**
Negative	Aware	Negative	"Negative"	Aware	--	**Aware**
Positive	Unaware	Positive	"Positive"/ "Negative"	Aware/ Unaware	Dilution of Aware and Unaware EC	Aware/ Unaware
Positive	Unaware	Negative	"Positive"/ "Negative"	Aware/ Unaware	Dilution of Aware and Unaware EC	Unaware/ Aware
Negative	Unaware	Positive	"Positive"/ "Negative"	Unaware/ Aware	Dilution of Aware and Unaware EC	Aware/ Unaware
Negative	Unaware	Negative	"Positive"/ "Negative"	Unaware/ Aware	Dilution of Aware and Unaware EC	Unaware/ Aware

Note. We take a VAM with no "Don't know" option here. Later, we elaborate on the differences between the two measures in underestimating unaware EC.

negative pictures from standardized databases such as the International Affective Picture System (IAPS; Lang et al., 2008) or the Open Affective Standardized Image Set (OASIS; Kurdi et al., 2017). However, *standardized ratings* do not imply that the valence is consistently perceived across individuals. For example, Stahl et al. (2009) selected positive pictures with a mean evaluation of 7.9 and negative pictures with a mean evaluation of 2.5 on a scale from 1 to 9. The average standard deviation of valence ratings in the IAPS is 1.58. Assuming a normal distribution of the ratings, ~ 4.4% of a random participant sample would evaluate a normatively negative picture from this US set as neutral or even positive (> 4.87 on the 1-9 scale). Likewise, ~ 4.4% would evaluate a normatively positive picture from this US set as neutral or even negative (< 5.58 on the 1-9 scale). Thus, even though a positive (negative) US may be positive (negative) to most participants, it is certainly not positive (negative) to all participants.

There are multiple reasons why people's subjective US evaluations could deviate from the "objective" valence. On the one hand, specific personality traits such as low agreeableness predispose individuals to deviant evaluations of affective stimuli (Ingendahl & Vogel, 2022, 2023). On the other hand, no attitude object is unequivocally liked or disliked by everybody. A cute puppy might be positive to 95% of people but not to those with a severe allergy or a phobia. Likewise, pictures of explosions might be negative to 95% of people but not to pyromaniacs.

For the standard EC effect, it does not matter whether participants experience one out of many USs differently from the normative valence. In the end, if one or two USs of 20 are of opposite valence to a participant, this will only slightly attenuate the EC effect. In a VAM, however, pairings where a normatively positive US is subjectively negative (and vice versa), can have serious consequences. This is depicted in Table 1.

If a CS is paired with a normatively positive US, and the participant is genuinely aware of the pairing, and the US is also subjectively positive to this participant, then the conditioning is classified as aware, and the CS contributes to the aware EC effect. However, if the US is subjectively negative to this participant, the pairing will be falsely classified as unaware, despite the participant being aware of the pairing. Because the US is negative to the participant, it will have a negative effect on the CS evaluations – opposite to the expected direction of an EC effect. Thus, a genuinely aware but reversed EC effect is falsely classified as an unaware EC effect, thereby leading to an underestimation of unaware EC. The corresponding case is if a CS is paired with a normatively negative US, the participant is genuinely aware of the pairing, but the US is subjectively positive to the individual. Again, a genuinely aware reversed EC effect is falsely classified as an unaware EC effect, thereby leading to an underestimation of unaware EC. Note that there is no such systematic bias for aware EC. This is because participants can merely guess the valence of the US for pairings where they are genuinely unaware. Thus, aware EC effects will also be reduced, but only by diluting the experimental manipulation with some pairings that operate against the expected EC effect (see also the following simulation).

To conclude, VAMs implicitly assume that all participants perceive a normatively positive/negative US as such. This assumption is likely violated, even if strongly positive/negative USs are selected from standardized databases. If a participant experiences the US differently from its normed valence, this could lead to an underestimation of unaware EC. Therefore, the reversed unaware EC effects found in previous research (Alves & Imhoff, 2023; Förderer & Unkelbach, 2013; Halbeisen et al., 2014; Stahl et al., 2009; Waroquier et al., 2020) might be caused by pairings where the individual US valence diverges from the normed US valence. This also raises the question of whether unaware EC actually exists but was not found because of this methodological bias. Thus, an improved methodology is necessary to reinvestigate the role of unaware EC in attitude formation.

A Corrected Valence Awareness Measure

If deviations of the subjective US valence from the normed US valence lead to an underestimation of unaware EC, VAMs should be adjusted to account for subjective US valence. We propose several solutions. On an experimental level, researchers could a priori select only those USs that are positive/negative to a specific participant (but see Förderer & Unkelbach, 2013, where reversed effects prevailed despite individual US selection). However, this may not always be possible, for example, because the experiment requires that participants are not exposed to the USs before the conditioning or because the EC effect of a specific US is of interest. Alternatively, one could assess subjective US evaluations post hoc and exclude all pairings where the normed valence does not match the subjective US evaluation (i.e., positive evaluations of negative USs and vice versa). However, excluding pairings from the analysis leads to missing values and reduces statistical power.

As an alternative[3], researchers could correct the VAM based on binary subjective US evaluations (see Table 1). Specifically, participants may be asked to classify the USs themselves as positive or negative. If responses in the VAM match the subjective (and not the normed) US valence, these pairings should be classified as aware. For example, suppose a participant indicates that a CS was paired with a positive US and also evaluates this US as positive. In that case, the pairing should be classified as aware, irrespective of the picture's normed valence.

This correction should have the following effects on the estimated aware and unaware EC effects: The corrected aware EC effect will be smaller than the aware EC effect estimated from the uncorrected VAM. This is because the falsely unaware pairings where the experimental EC effect is reversed are now accurately classified as aware. For the same reason, the corrected unaware EC effect should be larger than in the uncorrected VAM. Also, the corrected aware and unaware EC effects should become smaller the more USs are subjectively opposite to the normed valence. This is simply because the experimental manipulation (normed US valence: positive vs. negative) is less effective if some USs do not have the intended US valence. In this corrected VAM, the estimated aware and the unaware EC effect should be similarly affected by such shifts in US valence, thereby allowing a fairer test to which extent EC effects depend on awareness.

Overview of the Present Research

We examined to what extent valence awareness measures (VAMs) underestimate unaware EC due to differences between normed and subjective US valence and whether unaware EC emerges when correcting for this underestimation. We first assessed how often subjective US evaluations shift from the normed US valence by re-analyzing publicly available data from previous EC experiments. Next, we quantified the degree of underestimation in a simulation study (Simulation A). Here, we took into account the strength of aware and unaware EC, the extent to which participants are genuinely aware of the pairings, and the probability that a subjective US evaluation diverges from the normed US valence. We then conducted three experiments (Experiments 1-3) that tested the simulation's predictions empirically. These experiments also tested the influence of multiple procedural factors on the degree of underestimation of unaware EC, and whether unaware EC emerges when accounting for subjective US valence. Finally, we next extended our framework to the affect-as-information heuristic, which we will return to after Experiment 3.

Re-Analysis of Previous EC Studies

Our reasoning for underestimated unaware EC effects rests on the assumption that the subjective US valence can deviate from the normed US valence. In a first step, we therefore tested whether this requirement is fulfilled in published EC experiments. Because many EC experiments rely on the normed ratings in the IAPS (Lang et al., 2008) or the OASIS (Kurdi et al., 2017), there are few publicly available data sets from experiments where participants also evaluated US stimuli. Still, we identified a convenience sample by searching googlescholar, the OSF, and asking colleagues. We downloaded the data from its OSF directory for each of the studies and identified the variables containing US evaluations. To operationalize US valence shifts, we categorized the US ratings as positive (above scale midpoint), neutral (scale midpoint), or negative (below scale midpoint), depending on the study's rating scale.

We calculated the probabilities of evaluations deviating from the normed valence (i.e., p(subjectively positive|normed negative), etc.). Overall, valence shifts were present in all experiments we considered (see Table 2). The probabilities ranged from 1.5% to 19.1%. The re-analysis therefore shows that shifts in US valence indeed occur in EC experiments. This raises the question of to what extent these shifts in US valence may bias the estimated unaware EC effect, which we tested in a simulation study.

Simulation A: Valence Shifts

We conducted a simulation to quantify the underestimation of unaware EC. For simplicity, valence was a binary construct in this simulation (i.e., positive versus negative). Therefore, normed US valence, subjective US valence, and CS evaluations were treated dichotomously. As a consequence, we relied on the bivariate correlation measure Δp

3 One further alternative is of course switching to more sophisticated awareness measures, such as the process dissociation method (Hütter et al., 2012), the structural knowledge attribution task (Waroquier et al., 2020), or the recent two-button procedure (Stahl et al., 2023). We will explain these methods in the General Discussion. Note that these methods are slightly more complex regarding task instructions and data analysis compared to the VAM.

Table 2. Shifts in US Valence in Previous Studies

Experiment	US Materials	Shifts (Normed Positive)	Shifts (Normed Negative)	Scale
Alves et al. (2020; Exp. 3)	16 IAPS pictures	4.5% (negative)	3.5% (positive)	1-8
Ingendahl & Vogel (2023)	60 OASIS pictures	3.9% (negative) 13.7% (neutral)	4.9% (positive) 12.3% (neutral)	1-7
Ingendahl & Vogel (2022)	60 OASIS pictures	2.6% (negative) 9.5% (neutral)	6.4% (positive) 13.5% (neutral)	1-7
Fan et al. (2021; Exp. 1A)	Pretested images	-- (no positive USs)	4.8% (positive) 4.8% (neutral)	1-9
Mierop et al. (2019; Exp. 1)	200 IAPS pictures	11.5% (negative)	19.1% (positive)	1-2
Luck et al. (2021; Exp. 1)	12 IAPS pictures	2.7% (negative) 5% (neutral)	1.5% (positive) 2.5% (neutral)	1-9

Note. Positive/negative classifications are based on cutoffs at the scale midpoint. In the case of Mierop et al. (2019), data are taken from a speeded classification task, thus leading also to incorrect responses instead of actual valence shifts.

to quantify the EC effect. Δp is defined as p("positive CS evaluation"|positive US) – p("positive CS evaluation"| negative US). It is restricted to values between -1 (perfect reversed EC effect) and 1 (perfect normal EC effect) and can be interpreted like a Pearson correlation (Jenkins & Ward, 1965).

Methods

Parameters

We varied the true strength of aware EC (Δp = 0, 0.25, 0.5) and unaware EC (Δp = 0, 0.25, 0.5). We varied the probability of a valence shift (i.e., a normatively positive US is negative to a participant and vice versa) between 0% and 35% in steps of 2.5%. We also varied the probability of true awareness of a pairing with the values 10%, 25%, 50%, 75%, and 90%. For each combination, we simulated 1,000,000 pairings. 50% of these pairings had a positive and 50% a negative normed US valence. Because we were interested only in the expected value of the EC effect, we did not introduce a nested data structure (i.e., pairings nested in participants).

Estimated Effects

We generated EC effects such that the CS evaluation aligned with the subjective US valence, depending on the true strength of aware/unaware EC. For example, when aware EC was 0, CS evaluations were randomly positive/negative in pairings with true awareness. When aware EC was 0.5, CSs with truly aware pairings had a 50% chance of acquiring the *subjective* US valence. When unaware EC was 0.5, CSs with truly unaware pairings had a 50% chance of acquiring the subjective US valence, etc.

We generated responses in the standard "positive"/"negative" VAM such that participants would respond with the *subjective* US valence when there was true awareness (e.g., responding "positive" if the CS had been shown with a subjectively positive US), and that participants simply guessed when there was no true awareness. Based on the classification in Table 1, we computed four empirical EC effects: the aware and the unaware EC effect estimated via the VAM, and the aware and the unaware EC effect estimated via the corrected VAM.

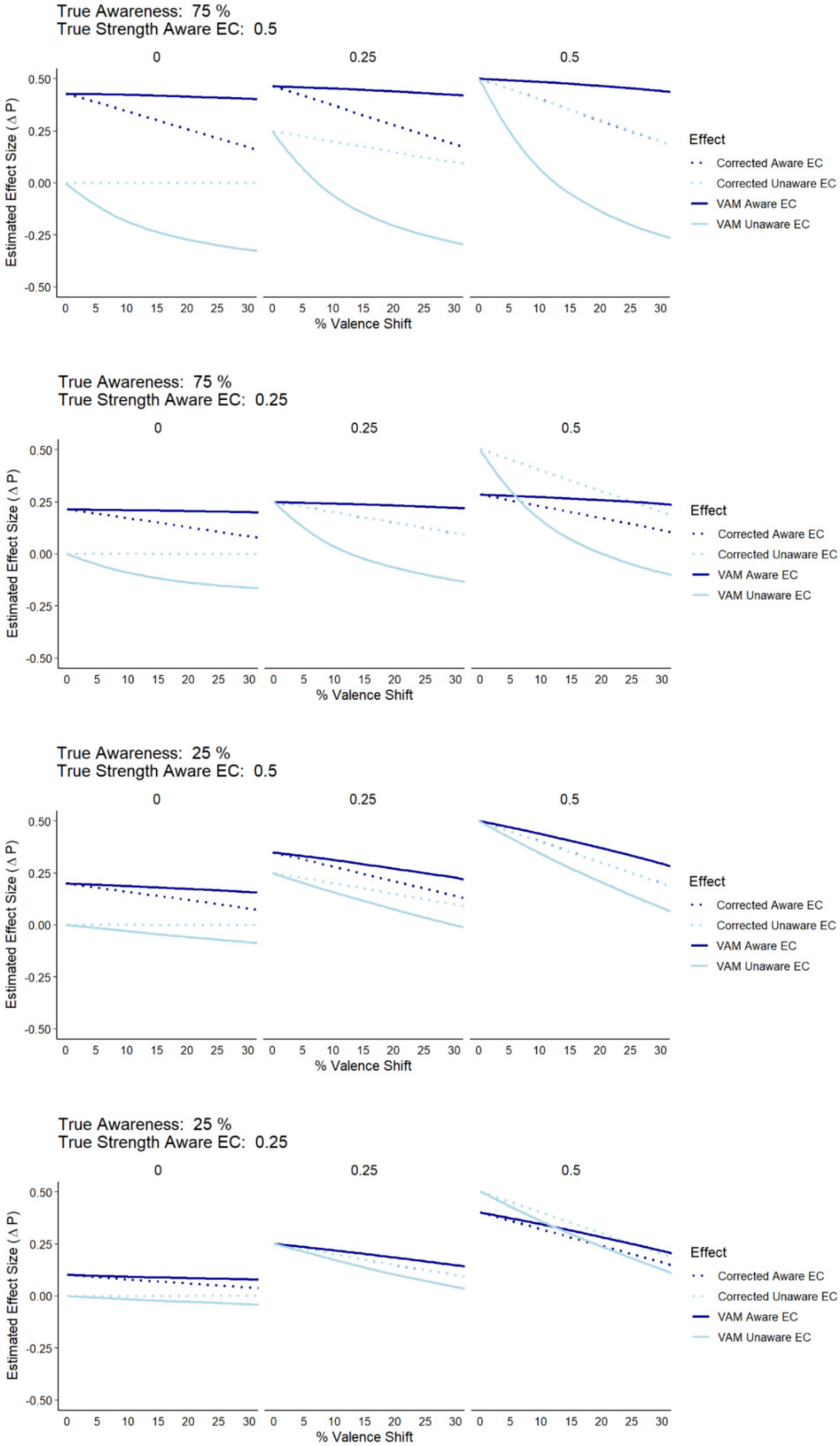
Results

We present the key findings of the simulation in Figure 1. Here, the estimated EC effects are depicted depending on true awareness (25% vs. 75%; separate graphs) and strength of unaware EC (0 vs. 0.25 v. 0.5; separate subplots). We only depict the case where aware EC was medium or strong (Δp = 0.25 or 0.5; separate graphs) because they are arguably the more realistic scenarios based on current evidence on EC (Hofmann et al., 2010). We provide the other plots and the complete code of our simulation on the OSF (https://doi.org/10.17605/OSF.IO/TPA3V).

When awareness was high (75%; Figure 1a/b), valence shifts led to a severe underestimation of the unaware EC effect through VAMs. Even if unaware EC was similarly strong as aware EC (Δp = 0.5; Figure 1a), a small percentage of valence shifts (~ 5-10%) fully eliminated the estimated unaware EC effect, which could even become negative depending on the strength of unaware EC and the frequency of valence shifts. In contrast, VAM estimates of the aware EC effect were only slightly affected by valence shifts and were estimated close to the true effect of Δp = 0.5 and 0.25. Thus, at high awareness, the VAM indeed led to a serious underestimation of unaware EC but not aware EC.

When overall awareness was low (25%, Figure 1c/d), the underestimation of unaware EC was less severe. Estimates still decreased with more valence shifts, but the slope was less steep compared to high awareness. The estimated aware EC effect also decreased with more valence shifts, but to a smaller degree than estimates of the unaware EC effect. The aware EC effect was more strongly underestimated when unaware EC was weak or absent and when awareness was low.

In contrast to these results, the corrected VAM was not systematically biased towards unaware or aware EC effects. Estimates were entirely accurate if there were no valence shifts. With more valence shifts, estimates of aware and unaware EC effects symmetrically decreased, depending on the true strength of aware and unaware EC. When both were of equal strength (both .5 or both .25; Figure 1), va-



a)

b)

c)

d)

Figure 1. Estimated EC Effects as a Function of True Awareness, Strength of Aware EC, Strength of Unaware EC, and Valence Shifts

Note. Different subplots (0 vs. 0.25 vs. 05) show results depending on the true strength of unaware EC. We also conducted a simulation for the VAM with a “don’t know” option, which is provided on the OSF.

lence shifts similarly reduced the estimated aware and unaware EC effects, showing that the corrected VAM was not biased against aware or unaware EC.

Discussion

Our simulation reveals the following core insights: First, the stronger the subjective US valence deviates from the normed US valence, the stronger the underestimation of unaware EC, depending on the strength of aware and unaware EC. For example, if aware and unaware EC are equally strong ($\Delta p = 0.5$), even a few valence shifts (~5-7%) lead to an estimated unaware EC effect of zero. If aware EC is stronger than unaware EC, estimated unaware EC effects can even become negative – as was found in previous EC research (Alves & Imhoff, 2023; Förderer & Unkelbach, 2013; Halbeisen et al., 2014; Stahl et al., 2009; Waroquier et al., 2020).

Second, the degree of bias also depends on the degree of awareness. Underestimation of unaware EC is more severe the more likely participants are aware of a pairing. What seems paradoxical at first sight results from more aware trials with reversed EC that are falsely classified as unaware. Third, correcting the VAM for subjective US valence reduces the bias. Here, both aware and unaware EC estimates are symmetrically reduced the more often subjective US valence deviates from the normed valence.

Because a simulation does not necessarily mirror reality, we conducted three EC experiments to test whether VAMs indeed underestimate unaware EC. In Experiment 1, we first tested whether shifts in US valence lead to an underestimation of unaware EC, whether this bias is reduced if VAMs are corrected, and whether the corrected estimates do actually find evidence for unaware EC.

Experiment 1: Baseline Effect

Experiment 1 used a standard EC paradigm with a subsequent VAM and US evaluation task. Experiment 1 was preregistered on aspredicted (https://aspredicted.org/jc8kd.pdf and https://aspredicted.org/kf386.pdf[4]).

Methods

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in all studies.

Design and Materials

We varied the normed US valence (positive vs. negative) within subjects. There were no additional manipulations. We selected 30 pictures from the Open Affective Standardized Image Set (OASIS; Kurdi et al., 2017) as USs. The 15 positive USs had a mean valence rating above +0.5 SD, and the 15 negative USs had a mean valence rating below -0.5 SD, relative to all 900 OASIS pictures. As the primary goal of this study was to show *what happens if* there are individual shifts in US valence, we chose less extreme USs to make shifts in US valence more likely. Thus, positive USs scored between 5.0 and 5.8 (M = 5.24, SD = 0.22) on the 7-point valence rating scale of OASIS, and negative USs between 2.4 and 3.7 (M = 3.16, SD = 0.40), t = 17.60, p < .001. There was no difference in the mean arousal between pictures of normatively positive (M = 3.88, SD = 0.56) or negative (M = 4.05, SD = 0.87) valence, t = -0.64, p = .527. The computer program randomly selected the pictures for each participant from the set. We used the 32 pseudowords (e.g., TABOMER, SOLEDAN) by Ingendahl and Vogel (2023) as CSs. For each individual participant, a random subset of 20 stimuli was selected to serve as CSs.

Procedure

All experiments were implemented in the software Sosci Survey (Leiner, 2019) and conducted online. The paradigm was adapted from Ingendahl and Vogel (2023). After providing informed consent, participants were first exposed to an EC procedure. Next, they evaluated the CSs, before answering a VAM. Last, participants also evaluated the USs and provided sociodemographic data. In the following, we will discuss each task step by step. Detailed screenshots and material lists are provided in our OSF directory. In line with our local university's ethical guidelines, no specific ethics approval was necessary for this study.

EC Procedure. Participants were told that they would see some unfamiliar brand names presented together with pictures. They should merely look at the sequences and wait for further instructions. In the following conditioning phase, 20 CSs (10 per normed US valence) were presented together with valenced pictures. After a blank screen of 500 ms, a CS-US pair was presented together for 2,500 ms. We counterbalanced between participants whether CSs/USs were shown on the left/right on the screen. Each CS was conditioned five times, leading to a conditioning phase of 20 x 5 = 100 trials that were shown in random order. Each CS was always shown with the same identical US (identity pairing).

CS Evaluation. After the conditioning phase, participants were told that they should now rate the brand names. Participants were presented with the 20 names in random order and asked "How would you evaluate this brand name?" The scale ranged from 1 (*very unpleasant*) to 7 (*very pleasant*). Each CS was shown on a single slide.

VAM. Afterward, we assessed the VAM as by Stahl et al. (Exp. 1, 2009). Participants were tasked to indicate whether

4 During data collection (after ~ 50 participants) we noticed the worst imaginable mistake one could make in a preregistration – pregistering the main hypothesis in the opposite way (i.e., using "smaller" instead of "larger"). We therefore decided to add a second preregistration where the mistake in the initial preregistration was made transparent. By that time, we had not analyzed any data yet.

a CS had been paired with a pleasant or an unpleasant picture[5]. We emphasized in the instructions that participants should only answer whether the CS had been shown with pleasant/unpleasant pictures and not whether they found the CS itself pleasant or unpleasant. In the VAM, a single CS was shown in the middle of the screen, together with the buttons "pleasant" and "unpleasant". The 20 CSs were presented in random order.

US Evaluation. Last, participants were also asked to evaluate the pictures. Again, the 20 USs were presented, each on a single slide and in random order. The USs were also evaluated on a dichotomous scale ("pleasant"/"unpleasant"). To be consistent with the presentation during the conditioning procedure, we first presented the respective US picture for 2,500 ms. Afterward, the picture disappeared, and the two buttons were shown together with the question "What kind of feeling does this picture elicit in you?"

Power Analyses and Sample

We calculated a priori power analyses with G*Power 3 (Faul et al., 2007). As a rough approximation for Experiment 1, we sought to detect a small to medium (d = 0.3) within-subjects effect (main effect/interaction) with 80% power, leading to a necessary sample size of 90 participants. We decided to add a buffer of ten participants. 100 (90 female, 9 male, 1 diverse, M_{Age} = 22.46, SD_{Age} = 3.36; 99 native speakers) participants were recruited via social media, mailing lists, and our university's study portal in exchange for course credits and a lottery for 3 x 15€ online store vouchers.

Results

Valence Shifts

We first tested whether subjective US evaluations diverged from normed US valence. Indeed, 32% of the normed positive USs were evaluated negatively by participants, and 31% of the normed negative USs were evaluated positively. As preregistered, we tested whether the probability of a valence shift differed between positive and negative USs by defining a multilevel binomial regression in lme4 (Bates et al., 2019) with the highest converging random effect structure (Barr et al., 2013). Normed US valence was the only predictor, coded 0.5 (positive) and -0.5 (negative), which did not show a significant difference, b = 0.05, CI_{95} = [-0.14, 0.24], z = 0.53, p = .595.

CS Evaluations

We next analyzed CS evaluations in three multilevel regressions. In our first model, which was not preregistered, normed US valence was the only predictor, coded 0.5 (positive) and -0.5 (negative)[6]. CS Evaluations were z-standardized (grand mean). Due to this coding, the valence main effect is the standardized mean difference (~Cohen's d) in evaluations between positively and negatively conditioned CSs – thus, representing the EC effect. This model found a standard positive EC effect, b = 0.20, CI_{95} = [0.11, 0.29], t = 4.54, p < .001.

We next conducted two preregistered models, which included either the uncorrected VAM or the corrected VAM as a predictor in addition to normed valence and the interaction. In each model, awareness was coded with 0 (unaware) and 1 (aware). Due to this coding, the main effect of normed valence is the EC effect for unaware pairings, and the interaction indicates the change in the EC effect for aware pairings. To compare the difference in the unaware EC effects based on the uncorrected and corrected VAM, we provide 95% confidence intervals around each effect estimate.

The mean CS evaluations are visualized in Figure 2. As expected, the model with the uncorrected VAM as an awareness measure revealed a negative EC effect for unaware CSs, b = -0.53, CI_{95} = [-0.66, -0.40], t = -8.21, p < .001. Thus, CSs paired with normed positive USs were evaluated *less* positively than CSs paired with normed negative USs for pairings classified as unaware (see Figure 2a). This effect changed the direction for aware CSs, as shown by the normed US valence x awareness interaction, b = 1.23, CI_{95} = [1.11, 1.44], t = 15.13, p < .001. Hence, the EC effect was positive for aware CSs, b = 0.75, CI_{95} = [0.64, 0.86], t = 13.29, p < .001. The main effect of awareness was not significant, b = 0.02, CI_{95} = [-0.06, 0.10], t = 0.45, p = .651.

The pattern changed when using the corrected VAM as an awareness measure (see Figure 2b). Here, the unaware EC effect was descriptively negative but not significant anymore, b = -0.12, CI_{95} = [-0.27, 0.03], t = -1.63, p = .104. The interaction was significant, b = 0.48, CI_{95} = [0.30, 0.67], t = 5.24, p < .001, leading to a positive aware EC effect, b = 0.36, CI_{95} = [0.26, 0.47], t = 6.83, p < .001. There was also an unexpected main effect of the corrected measure with more positive evaluations for aware pairings, b = 0.14, CI_{95} = [0.05, 0.23], t = 2.99, p = .003.

Awareness

In an exploratory manner, we also investigated to what extent awareness (either by the VAM or the corrected VAM) depended on normed valence, using two further logistic multilevel regressions. When using the uncorrected VAM, awareness was slightly above chance (57%, p < .001). Awareness was considerably higher when using the corrected VAM (67%). Neither of the measures showed an effect of normed valence, all zs < 1.26, all ps > .209.

5 We chose the labels pleasant/unpleasant instead of positive/negative in all measures to be consistent with Stahl et al. (2009).

6 Note that we preregistered effect codings of 1 and -1, but switched to 0.5 and -0.5 so that the regression weight represents the standardized mean differences (~ Cohen's d). The different coding does not change the t-/p-values.

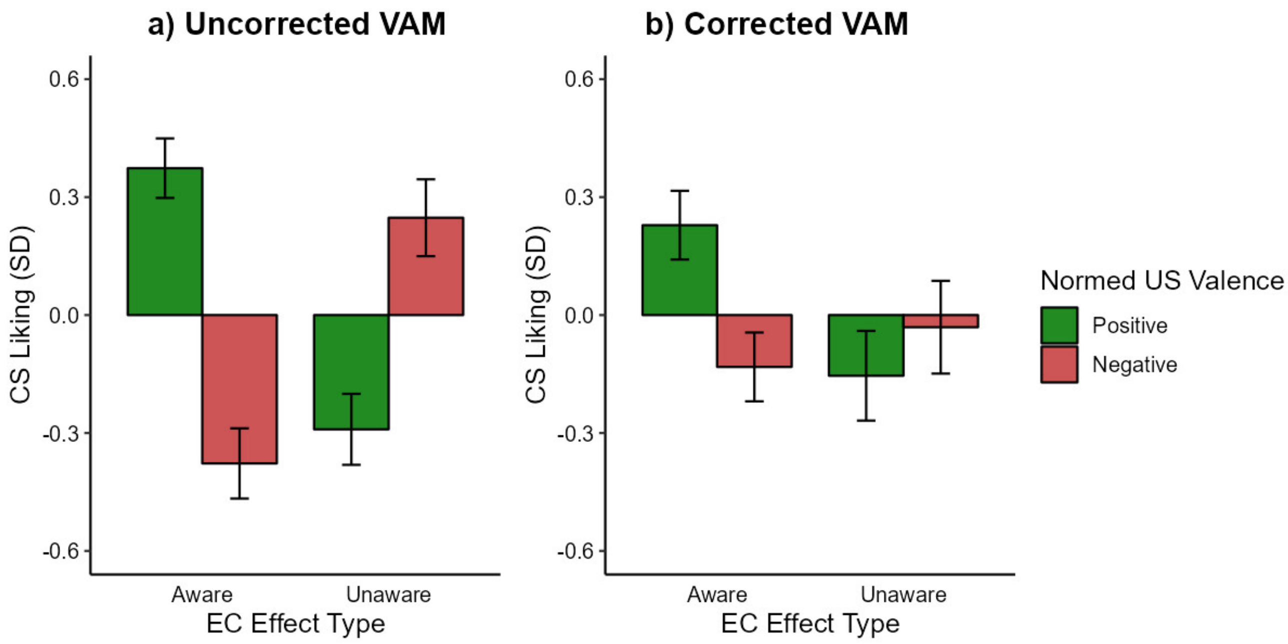


Figure 2. Figure 2 Mean CS Evaluation in Experiment 1 as a Function of Normed US Valence, Awareness Measure, and EC Effect Type

Note. Error bars represent 95% confidence intervals. Descriptive statistics with means and standard deviations are also provided on the OSF.

Discussion

Experiment 1 supports our reasoning and the results from Simulation A. Even though the subjective US valence diverged from the normed valence in one-third of the pairings, the overall EC effect was positive and unsuspicious. However, dissecting this overall EC effect into an unaware and an aware effect revealed vastly different results. With the uncorrected VAM, the unaware EC effect was reversed, whereas the aware EC effect was positive. When using the corrected VAM, the unaware EC effect was not different from zero, whereas the aware EC effect was reduced but still robust and significant.

These findings show that VAMs substantially underestimate unaware EC if the subjective US valence diverges from the normed valence. This can go as far as the unaware EC effect is reversed. A corrected VAM that controls for subjective US valence eliminates this problem, even though the unaware EC effect was descriptively still negative in this experiment. The aware EC effect was smaller in the corrected VAM, most likely because aware pairings with a reversed EC effect were now accurately classified as aware.

Based on these findings, the question arises when the underestimation of unaware EC is more or less extreme in an EC experiment and, as a consequence, when the chances are better or worse of detecting unaware EC. In that regard, Simulation A already shows three key moderators for the degree of underestimation. First, underestimation depends on the true strength of unaware and aware EC. These moderators are, however, difficult to influence directly without making specific theoretical assumptions beforehand. The other moderator is the proportion of truly aware pairings in an EC experiment. Specifically, if people are mainly aware of the pairings, the VAM should severely underestimate unaware EC (Figure 1), because more reversed aware pairings occur that are falsely classified as unaware. If awareness is overall low, the bias should be less severe.

In the following experiments, we therefore varied the degree of awareness in the EC paradigm. One straightforward method is to alter the number of repetitions for each pairing. If a CS is paired only once with the US instead of five times, awareness of the pairing should be lower.

Experiment 2: Number of Repetitions

Experiment 2 tested the influence of the number of repetitions on awareness and, consequently, on the degree of underestimation of unaware EC. Experiment 2 was preregistered on aspredicted.org (https://aspredicted.org/jb4gg.pdf).

Methods

The experiment followed the same procedure as Experiment 1, except that we manipulated the number of presentations per CS-US pairing between participants. In an experimental condition, each pairing was shown only once. The control condition was an exact replication of Experiment 1, thus with five repetitions per CS-US pairing. Note that the number of CSs was still identical between the conditions. Experiment 2 therefore featured a mixed design with the between-subjects[7] factor number of presentations (one vs. five presentation) and the within-subjects factor normed US valence (positive vs. negative).

We conducted a-priori power analyses in G*Power 3 (Faul et al., 2007). We sought to detect a small to medium (*f* = 0.15) between-within interaction with 80% power, leading to a necessary sample size of 90 participants. However, we decided to recruit 150 participants to have sufficient power within each between-subjects condition. Our final sample size was N = 151 (111 female, 39 male, 1 diverse, M_{Age} = 22.62, SD_{Age} = 3.72; 138 native speakers). Again, they were recruited via social media, mailing lists, and our university's study portal in exchange for course credits and a lottery for 3 x 15€ online store vouchers.

Results

Valence Shifts

We tested our hypotheses with similar multilevel regressions as in Experiment 1. Subjective US evaluation diverged from the normed US valence for 30% of the USs. Valence shifts were slightly less frequent for positive USs, b = -0.17, CI_{95} = [-0.37, 0.43], z = -2.16, p = .031. There was no significant main effect of number of presentations (1 = one, -1 = five), b = 0.01, CI_{95} = [-0.08, 0.09], z = 0.22, p = .828, and no significant interaction with normed valence, b = 0.03, CI_{95} = [-0.13, 0.18], z = 0.32, p = .751.

CS Evaluations

The first model predicted CS evaluations from normed US valence (again coded with 0.5 and -0.5) and found a strong main effect, b = 0.22, CI_{95} = [0.14, 0.29], t = 5.55, p < .001, indicative of a standard EC effect. Number of presentations (1 = one, -1 = five) had no significant main effect, b = 0.002, CI_{95} = [-0.05, 0.05], t = 0.09, p = .925. However, the EC effect was overall weaker in the one-presentation condition, b = -0.11, CI_{95} = [-0.18, -0.03], t = -2.69, p = .008.

As in Experiment 1, we conducted two separate models that also included the different awareness measures as predictors. Each model contained the predictors normed US valence (0.5 vs. -0.5), awareness (1 vs. 0), and presentation number (1 vs. -1), and all higher-order interactions. Mean CS evaluations are visualized in Figure 3.

Again, the model with the uncorrected VAM as a predictor revealed a negative EC effect for unaware CSs, b = -0.49, CI_{95} = [-0.60, -0.38], t = -9.09, p < .001. This effect changed direction for aware CSs, b = 1.25, CI_{95} = [1.12, 1.39], t = 18.11, p < .001, resulting in a positive EC effect for aware pairings, b = 0.77, CI_{95} = [0.67, 0.86], t = 15.91, p < .001. In contrast to our expectations, the unaware EC effect was only descriptively less negative in the one-presentation condition, as shown by the non-significant Normed US Valence x Number of Presentations interaction, b = 0.02, CI_{95} = [-0.08, 0.13], t = 0.40, p = .686. However, there was a significant three-way interaction, b = -0.19, CI_{95} = [-0.33, -0.06], t = -2.81, p = .005. As shown in Figure 3, the aware EC effect was smaller in the one-presentation condition. All other effects in this model were not significant; all *t*s < 1.04, all *p*s > .297.

As in Experiment 1, this pattern changed in a regression model that specified the corrected VAM as a predictor (see Figure 3b). Here, the overall unaware EC effect, expressed by the normed US valence main effect, was not significant, b = -0.07, CI_{95} = [-0.18, 0.05], t = -1.16, p = .247. The interaction with awareness was significant, b = 0.48, CI_{95} = [0.33, 0.62], t = 6.52, p < .001, leading to an aware EC effect, b = 0.41, CI_{95} = [0.31, 0.50], t = 8.50, p < .001. There was also an unexpected main effect of the corrected measure, b = 0.19, CI_{95} = [0.12, 0.26], t = 5.08, p < .001. The three-way interaction was not significant, b = 0.02, CI_{95} = [-0.13, 0.16], t = 0.24, p = .814. Except for an insignificant trend that the unaware EC effect was slightly lower (and negative) in the few-presentations condition, b = -0.10, CI_{95} = [-0.22, 0.01], t = 1.72, p = .085, there were no other notable effects, all *t*s < 0.65, all *p*s > .519.

Awareness (Exploratory)

Overall, 56% of the pairings were classified as aware when using the VAM and 60% when using the corrected VAM. In an exploratory manner, we also investigated the effects of number of presentations and normed US valence on awareness. For that purpose, we predicted each awareness measure (uncorrected and corrected VAM) with separate multilevel logistic regressions that were not preregistered. Each model contained the predictors normed US valence (0.5 vs. -0.5) and presentation number (1 vs. -1) and the interaction. For the VAM, there was no main effect or interaction of normed US valence or number of presentations, all *z*s < 1.62, all *p*s > .105. For the corrected VAM, however, there was a main effect of number of presentations, b = -0.21, CI_{95} = [-0.30, -0.11], z = 4.17, p < .001, such that awareness was lower in the one-presentation (56%) than in the five-presentations condition (65%). Unexpectedly, there was also a significant interaction, b = 0.21, CI_{95} = [0.06, 0.36], z = 2.69, p = .007, such that the difference between the two conditions was stronger for normed negative USs. The main effect of valence was not significant, b = -0.08, CI_{95} = [-0.23, 0.07], z = -1.08, p = .282.

Discussion

Experiment 2 replicated the findings from Experiment 1. Due to shifts in subjective US valence from normed US valence, the unaware EC effect estimated from the VAM was negative. Correcting the VAM for subjective US valence substantially reduced this negative EC effect. Yet, even after correction, no evidence for a positive unaware EC effect was found.

7 We manipulated the number of presentations between participants in order to have a control condition identical to Experiment 1.

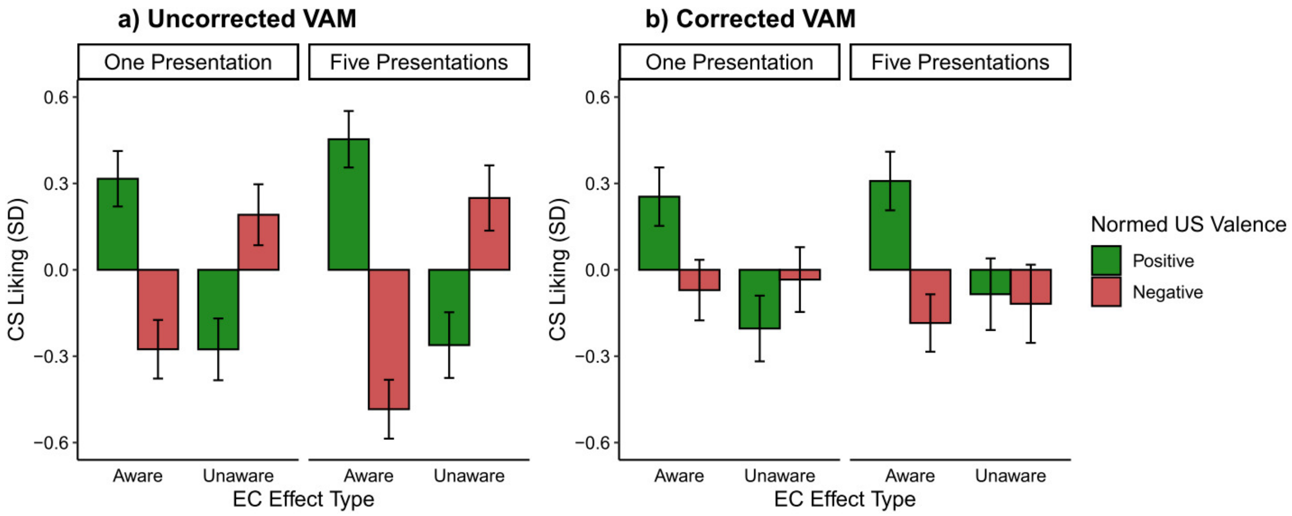


Figure 3. Mean CS Evaluation in Experiment 2 as a Function of the Number of Presentations, Normed US Valence, Awareness Measure, and EC Effect Type

Note. Error bars represent 95% confidence intervals.

In line with the results from our simulation, the estimated aware EC effect was significantly reduced when there was no awareness. In contrast to the simulation results, the uncorrected VAM's unaware EC was only slightly less negative when there was less awareness (i.e., in the one-presentation condition). Also, the correction seemed to benefit mostly the unaware EC effect in the five-presentations condition but not in the one-presentation condition, where the unaware EC effect estimated from the corrected VAM was still negative.

One reason might be that the reduced number of pairings also weakened unaware EC. Thus, there might be less bias but also weaker unaware EC, which may cancel each other out. Therefore, other manipulations might be necessary to boost unaware EC. In that regard, the implicit misattribution account (Jones et al., 2009; March et al., 2019) predicts EC in conditions where people are unaware of the pairings. According to this account, people misattribute the affect from the US to the CS. Earlier findings indicate that misattribution may work best if people confuse the source of the affective experience – in particular, if the same CS is paired with different USs (Hütter & Sweldens, 2013; Sweldens et al., 2010). We therefore decided to conduct another experiment using such a procedure. Another advantage of using different USs is that it may be more difficult to trace which CS was paired with which US, leading to lower awareness. According to our simulation, this should also lead to less underestimation of unaware EC, therefore providing optimal conditions to detect unaware EC. However, correcting VAMs for shifts in US valence is more complex if a CS was paired with multiple USs because multiple US evaluations need to be accounted for. One possibility is correcting based on the mean valence of all USs shown with a CS.

Experiment 3: Identity vs. Valence Pairing

Experiment 3 tested the influence of a valence pairing procedure on the underestimation (and correction) of the unaware EC effect. Experiment 3 was preregistered on [aspredicted.org](https://aspredicted.org/zg97s.pdf) (https://aspredicted.org/zg97s.pdf) .

Methods

The experiment followed the same procedure as Experiment 1, except that we manipulated the type of pairing between participants. Therefore, the design of this study was a mixed design with the between-subjects factor of pairing procedure (identity vs. valence) and the within-subjects factor of normed US valence (positive vs. negative). In an identity-pairing condition, each CS was consistently paired with the same US, replicating Experiment 1. In a valence-pairing condition, each CS was paired with five different USs (of the same normed valence). Specifically, CS1 was shown with US1, US2, US3, US4, and US5; CS2 was shown with US2, US3, US4, US5, US6, etc., such that all USs were shown equally often. The rest of the tasks were identical to the identity-pairing condition and the previous experiments. However, because now five US evaluations belonged to one response in the VAM, we computed an average US evaluation for all the five USs shown with a particular CS. The corrected VAM in the valence-pairing condition was calculated based on this average valence (i.e., a "positive" response in the VAM and a [mis-]match with the average US valence across all USs paired with that CS).

We conducted the same power analysis as in Experiment 2, leading to a desired sample size of N = 150 (129 female, 20 male, 1 diverse, M_{Age} = 22.77, SD_{Age} = 5.00; 145 native speakers). Again, participants were recruited via social media, mailing lists, and our university's study portal in exchange for course credits and a lottery for 3 x 15€ online store vouchers.

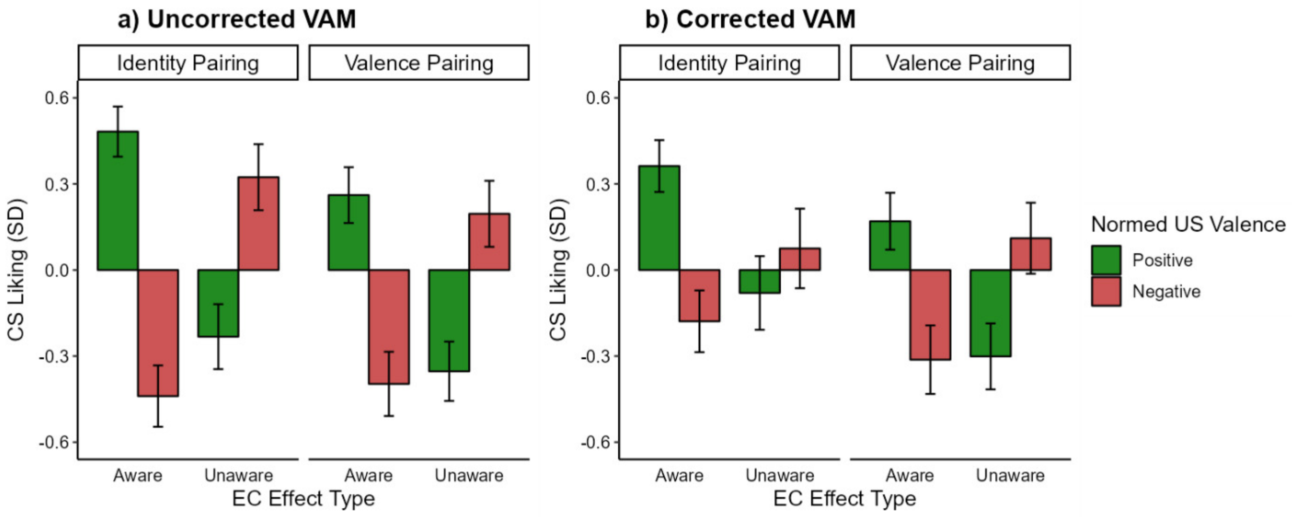


Figure 4. Mean CS Evaluation in Experiment 3 as a Function of Pairing Procedure, Normed US Valence, Awareness Measure, and EC Effect Type

Note. Error bars represent 95% confidence intervals.

Results

Valence Shifts

Subjective US evaluations diverged from normed US valence in 29% of pairings. Unexpectedly, valence shifts were less frequent for normed positive USs (26%) than normed negative USs (33%), b = -0.31, CI_{95} = [-0.49, -0.14], z = -3.45, p < .001. This effect was only present in the identity-pairing condition, as shown by a significant pairing procedure x normed US valence interaction, b = 0.22, CI_{95} = [0.05, 0.40], z = 2.48, p = .013. The main effect of pairing procedure was not significant, b = -0.002, CI_{95} = [-0.08, 0.08], z = -0.05, p = .961.

CS Evaluations

We tested our hypotheses by specifying the same multilevel regressions as in Experiment 2, using the same coding scheme (here, with valence-pairing coded 1 and identity pairing coded -1). The first model found a significant main effect of normed US valence on CS evaluations, b = 0.21, CI_{95} = [0.12, 0.29], t = 4.92, p < .001, indicative of a standard EC effect. Evaluations of CSs were also overall lower in the valence-pairing condition, b = -0.07, CI_{95} = [-0.12, -0.02], t = -2.64, p = .009. The overall EC effect was weaker in the valence-pairing condition, b = -0.12, CI_{95} = [-0.20, -0.04], t = -2.82, p = .005.

As before, we conducted two separate models with each awareness measure as a predictor. The mean CS evaluations are visualized in Figure 4.

Again, the model with the uncorrected VAM as an awareness measure showed a negative EC effect for unaware pairings, b = -0.55, CI_{95} = [-0.66, -0.45], t = -10.13, p < .001. This effect changed direction for aware pairings, b = 1.34, CI_{95} = [1.21, 1.45], t = 19.52, p < .001, resulting in a positive aware EC effect, b = 0.79, CI_{95} = [0.69, 0.89], t = 16.16, p < .001. In contrast to our expectations, the unaware EC effect did not differ between the conditions, as shown by the non-significant pairing procedure x normed US valence interaction, b = 0.003, CI_{95} = [-0.10, 0.11], t = 0.06, p = .949. In addition, there was a small three-way interaction, b = -0.14, CI_{95} = [-0.27, -0.001], t = -1.97, p = .049. As can be seen in Figure 4, the aware EC effect was smaller in the valence-pairing condition. Additionally, CS evaluations were overall more negative in the valence-pairing condition, b = -0.06, CI_{95} = [-0.12, -0.004], t = -2.11, p = .036. All other effects in this model were not significant, all ts < 0.51, all ps > .611.

As before, this pattern changed when using the corrected VAM as a predictor (see Figure 4b)[8]. The negative unaware EC effect was reduced but still negative, b = -0.28, CI_{95} = [-0.40, -0.16], t = -4.67, p < .001. The interaction with the corrected VAM was significant, b = 0.80, CI_{95} = [0.65, 0.94], t = 10.87, p < .001, leading to a positive aware EC effect, b = 0.51, CI_{95} = [0.42, 0.61], t = 10.49, p < .001. In addition, the corrected unaware EC effect was more negative in the valence-pairing condition (see Figure 4b), b = -0.13, CI_{95} = [-0.25, -0.01], t = -2.11, p = .035. There were no other significant effects, all ts < 1.61, all ps > .108.

8 In this analysis, we corrected the VAM responses in the valence-pairing condition based on the mean evaluation of the five USs the CS was shown with (i.e., < 50% positive US evaluations). In an exploratory manner, we also investigated whether the correction was more effective if the US valence was consistently positive/negative for a certain CS (i.e., 100% positive/negative evaluations). This was not the case, the analysis is provided in the R Markdown of our OSF directory.

Awareness

Last, we also investigated the effects of the pairing procedure on awareness using the same analysis as in Experiment 2. This time, we preregistered this analysis. For the uncorrected VAM, estimated awareness levels were 51% for the valence-pairing and 61% for the identity-pairing conditions, b = -0.15, CI_{95} = [-0.23, -0.07], z = -3.82, p < .001. In addition, awareness estimated via VAM was higher for normed positive USs, b = 0.23, CI_{95} = [0.03, 0.44], z = 2.19, p = .028. For the corrected VAM, estimated awareness levels were 54% in the valence-pairing and 70% in the identity-pairing conditions, b = -0.29, CI_{95} = [-0.39, -0.19], z = -5.47, p < .001. In addition, awareness estimated via the corrected VAM was higher for normed positive USs, b = 0.17, CI_{95} = [0.01, 0.32], z = 2.12, p = .034. All other effects were not significant, all *z*s < 1.58, all *p*s > .114.

Discussion

Experiment 3 replicated the pattern of our simulation and previous experiments for an identity-pairing procedure. Here, the uncorrected VAM underestimated the influence of unaware EC. Correcting for shifts in US valence reduced this bias but still provided no evidence for an unaware EC effect.

The results from the valence pairing procedure deserve further consideration. Here, the same degree of underestimation of unaware EC was found for the standard VAM as in the identity-pairing condition – despite overall less awareness. Also, correcting for shifts in US valence did not eliminate the bias here. The corrected unaware EC effect was still significantly negative and even more negative than in the identity-pairing condition. These results are inconsistent with the idea that reversed unaware EC effects are exclusively due to subjective valence changes and imply that other processes must also contribute.

One explanation might be the affect-as-information heuristic (Bar-Anan & Amzaleg-David, 2014; Hütter et al., 2012; Stahl et al., 2009). Participants might infer the US valence from their attitudes towards the CS. For example, if a CS was paired with a positive US, they might respond in the VAM based on their attitude towards the CS. If the CS is positive to the participant, then the answer in the VAM is also "positive", and the pairing is classified as aware. If the CS is negative to the participant (i.e., the conditioning was ineffective), then the answer in the VAM is also "negative", and the pairing is classified as unaware. Thus, affect-as-information might lead to similar reversed unaware EC effects as shifts in US valence do.

This additional process might explain why we found results inconsistent with our first simulation. If participants also rely on affect-as-information, a consistent reversed unaware EC effect emerges that cannot be corrected for. The less participants are aware of the pairings, the more are "unaware" EC effects biased by affect-as-information and not by valence shifts. This might also lead to less impact of the correction for US valence shifts when awareness is low. To test this explanation, we first extended our simulation to incorporate affect-as-information as a process.

Simulation B: Valence Shifts and Affect-As-Infomation

Methods

This simulation relied on the exact same methodology as Simulation A, but with an additional parameter to account for affect-as-information. Under affect-as-information, people rely on their CS evaluations as a cue for whether the CS was paired with a positive or negative US. For illustration, we varied the likelihood of affect-as-information in the simulation between 0% (leading to the same results as in Simulation A) and 25%, results for other parameter levels are provided on the OSF. In the case of 25%, participants would respond to 25% of the pairings with "positive" if they also liked the CS, and with "negative" if they disliked the CS. The rest of the simulation was identical to Simulation A.

Results

The core results from this simulation are displayed in Figure 5. More fine-grained variations with different levels of aware EC and affect-as-information are provided on the OSF.

When allowing for affect-as-information, the unaware EC effect was reversed even without any subjective valence shifts. The biasing influence of valence shifts was overall reduced. Even though a higher chance of valence shifts still reduced the estimated unaware EC effect, the slope was less steep than when there was no affect-as-information.

Furthermore, the corrected measure was also biased by affect-as-information and underestimated the unaware EC effect but overestimated the aware EC effect. The less participants were aware of the pairings, the less correcting for US shifts improved the accuracy in estimating the true unaware EC effect.

Discussion

Extending the simulation to affect-as-information explains why our previous experiments did not reveal the expected influence of overall awareness on the underestimation of unaware EC. If people rely on affect-as-information, reversed unaware EC can occur, independent of valence shifts. Affect-as-information has a stronger influence on the estimated unaware EC effect the *less* participants are aware of CS-US pairings. Valence shifts have a stronger influence the *more* participants are aware of CS-US pairings. Thus, in combination, the overall proportion of awareness will not influence the size of the unaware EC effect. Because the corrected VAM is also biased by affect-as-information, the corrected VAM will also show reversed unaware EC effects.

To test empirically whether affect-as-information also biases unaware EC estimates (estimated by uncorrected and corrected VAMs), we conducted a final experiment where reversed unaware EC could be produced primarily by affect-as-information and not by valence shifts. Specifically, we altered the paradigm of our previous experiments such that

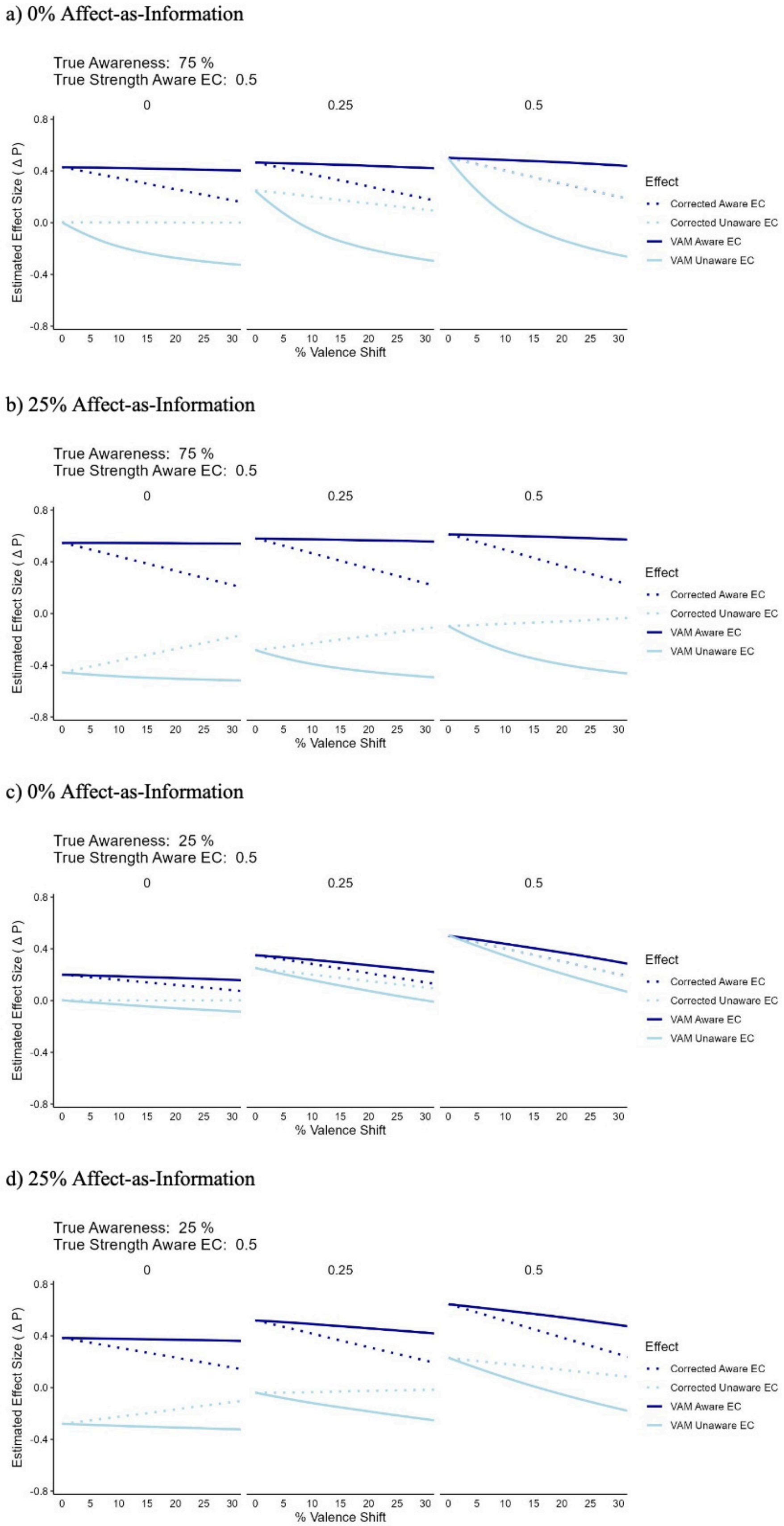


Figure 5. Results From a Simulation with 0 vs. 25% Probability of Affect-As-Information

Note. Visualizations for other parameter levels are provided on the OSF.

awareness of the pairings was extremely low and shifts in US valence were unlikely to occur.

Experiment 4: Affect-As-Information

Experiment 4 tested whether reversed EC effects still emerge in our paradigm when valence shifts are unlikely to bias the unaware EC effect. Specifically, we used US materials where US shifts were unlikely and awareness was overall low. Experiment 4 was preregistered on aspredicted.org (https://aspredicted.org/u9kd6.pdf).

Methods

The experiment followed the same procedure as Experiment 1 with the following differences. First, we used 48 (instead of 20) CSs, from which 24 were conditioned positively and 24 negatively. Second, we used USs from previous EC research where shifts in US valence were less likely. Specifically, we used the US materials by Ingendahl and Vogel (2022, 2023). Thirty positive pictures from the OASIS (Kurdi et al., 2017) with normed valence ratings higher than +1 SD (OASIS valence rating > 5.56 on a scale of 1-7) served as normatively positive USs, and 30 pictures with normed valence ratings below -1 SD (2.86) as normatively negative USs. As in Experiments 1-3, the USs were rated at the beginning of the experiment. Third, each CS-US pairing was shown only once, making awareness of the pairing rather unlikely. Last, we counterbalanced the order of the VAM and the CS evaluation between participants. Half of the participants first filled out the VAM and then evaluated the CSs, the rest of the participants had the standard order as in our previous experiments.

We used the same power analysis as in Experiment 1, leading to a final sample size of $N = 100$ (44 female, 56 male, $M_{Age} = 34.70$, $SD_{Age} = 12.43$; 99 native speakers). This time, participants were recruited via Prolific Academic.

Results

Valence Shifts

Subjective US evaluations diverged from normed US valence for 5% of the normed positive and 5% of the normed negative pictures. The probabilities of valence shifts did not depend on normed US valence, order, or the interaction, all *zs* < 1.28, *ps* > .202.

CS Evaluations

We conducted the same multilevel regressions as in the previous studies. In our first model, normed US valence (again coded with 0.5 and -0.5) had a main effect on CS evaluations, $b = 0.09$, $CI_{95} = [0.03, 0.15]$, $t = 2.94$, $p = .004$, thereby showing an overall EC effect. Order (1 = memory first, -1 = evaluation first) had no main effect, $b = 0.03$, $CI_{95} = [-0.03, 0.09]$, $t = 0.91$, $p = .365$. However, the EC effect was slightly but not significantly weaker in the memory-first condition, $b = -0.05$, $CI_{95} = [-0.11, 0.01]$, $t = -1.78$, $p = .078$.

As before, we conducted two models with each awareness measure as a predictor. Each model contained the predictors normed US valence (0.5 vs. -0.5), awareness measure (1 vs. 0), and order (1 vs. -1), and all higher-order interactions. The mean CS evaluations are visualized in Figure 6.

Again, the model with the uncorrected VAM as an awareness measure revealed a negative EC effect for unaware CSs, $b = -0.34$, $CI_{95} = [-0.42, -0.26]$, $t = -8.46$, $p < .001$. This effect changed direction for aware pairings, $b = 0.81$, $CI_{95} = [0.70, 0.92]$, $t = 14.55$, $p < .001$, leading to a positive aware EC effect, $b = 0.47$, $CI_{95} = [0.40, 0.55]$, $t = 12.51$, $p < .001$. Unexpectedly, the aware EC effect was smaller when memory was assessed first, as shown by the significant three-way interaction, $b = -0.16$, $CI_{95} = [-0.27, -0.05]$, $t = -2.83$, $p = .005$. All other effects in this model were not significant, all *ts* < 1.02, all *ps* > .307.

This pattern did not substantially change when using corrected VAMs (see Figure 6b). Here, the unaware EC effect was slightly less negative but still strong and significant, $b = -0.31$, $CI_{95} = [-0.39, -0.24]$, $t = -7.81$, $p < .001$. The interaction was significant, $b = 0.76$, $CI_{95} = [0.65, 0.87]$, $t = 13.62$, $p < .001$, leading to a positive aware EC effect, $b = 0.44$, $CI_{95} = [0.37, 0.52]$, $t = 11.79$, $p < .001$. As for the uncorrected VAM, the aware EC effect was smaller when memory was assessed first, as shown by the significant three-way interaction, $b = -0.14$, $CI_{95} = [-0.25, -0.03]$, $t = -2.52$, $p = .012$. All other effects in this model were non-significant; all *ts* < 1.27, all *ps* > .203.

Awareness

For both the uncorrected VAM (53%, $p = .003$) and the corrected VAM (53%, $p < .001$), awareness levels were slightly above chance. For the uncorrected VAM, awareness was higher for pairings with normed positive USs, $b = 0.32$, $CI_{95} = [0.14, 0.51]$, $t = 3.36$, $p < .001$. This was also the case for the corrected VAM, $b = 0.36$, $CI_{95} = [0.15, 0.58]$, $t = 3.31$, $p < .001$. All other effects were not significant, all *zs* < 0.79, all *ps* > .430.

Discussion

Our last experiment is fully in line with an affect-as-information explanation. Even though shifts in subjective US valence were rare (5%), the unaware EC effect estimated via the uncorrected VAM was reversed. Furthermore, even after correcting for shifts in US valence, the unaware EC effect was still reversed. This implies that participants also relied on an affect-as-information heuristic. The effect was likely stronger than in Experiment 3 because awareness levels were very low but still slightly above chance in this study, leading to more responses based on affect-as-information.

General Discussion

To what extent evaluative conditioning (EC) requires awareness of stimulus pairings has been a major debate in the last decades. Recent findings question whether EC without awareness exists. These findings are often based

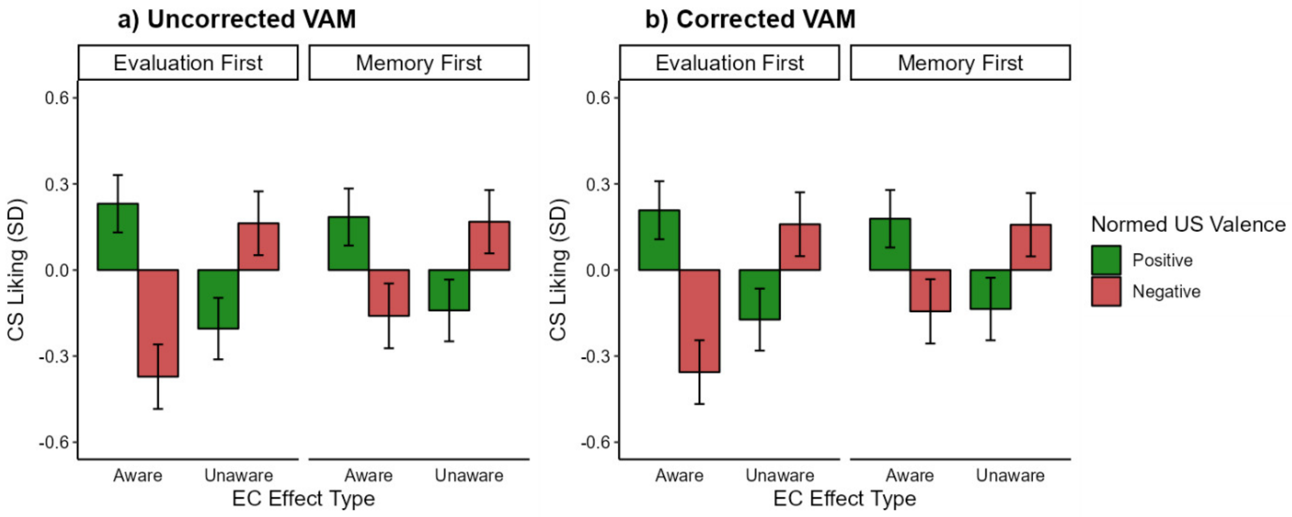


Figure 6. Mean CS Evaluation in Experiment 4 as a Function of Order, Normed US Valence, Awareness Measure, and EC Effect Type

Note. Error bars represent 95% confidence intervals.

on post-conditioning memory tests where participants indicate whether an object was paired with positive or negative stimuli. We proposed that these valence awareness measures (VAMs) systematically underestimate unaware EC effects because the subjective US valence may diverge from the normed US valence, leading to reversed aware EC effects being misclassified as unaware. This might explain the absence of unaware EC and the sometimes reversed unaware EC effects found in previous EC studies. We also proposed a corrected VAM that merely requires binary US classifications and effectively controls for subjective US valence, eliminating this bias.

We tested this with a re-analysis of previous EC studies, two simulations, and four preregistered experiments. Our re-analysis of previous EC studies showed that subjective US evaluations indeed diverge from normed US valence, creating sufficient conditions for the bias. Our first simulation showed that even the small probabilities of valence shifts found in the re-analysis might suffice to eliminate all chances of detecting even medium to strong unaware EC effects. The simulation also showed that the bias should be more pronounced the more people are aware of stimulus pairings. It also showed that the corrected VAM should eliminate the bias. We next tested these predictions in three preregistered experiments. Here, we found reversed unaware EC effects when using the standard VAM, which vanished when using the corrected VAM. Yet, even the corrected measure showed no evidence of (positive) unaware EC effects.

Our experiments did not confirm the prediction that the bias in the VAM is less pronounced under conditions with low awareness (e.g., reduced number of pairings or pairings with different USs). Therefore, we also tested whether a second bias might be responsible for these results – the affect-as-information heuristic. A second simulation confirmed this, showing that affect-as-information also leads to an underestimation of unaware EC, which is more pronounced the less participants are aware of the stimulus pairings. A final experiment showed a reversed unaware EC effect even under conditions where US valence shifts were unlikely to bias unaware EC. This supports the idea that affect-as-information also biases estimates of unaware EC effects.

Our research offers important implications for EC research – both on a theoretical and a methodological level – but also implications beyond EC.

Theoretical Implications for EC

In line with current EC research (Corneille & Stahl, 2018; Hütter, 2022; Moran et al., 2023), we do not find any evidence for unaware EC. Despite a potential bias in previous findings introduced by US valence shifts, correcting for this bias in our studies nevertheless led to estimated unaware EC effects close to zero. However, our research also shows that unaware EC estimates might be convoluted by affect-as-information, and therefore, it is not entirely certain that unaware EC does not exist. Still, our results show solid aware EC effects, which fully supports current memory-based and propositional EC accounts (De Houwer, 2018; Gast, 2018; Stahl & Aust, 2018). These accounts postulate that people need to consciously encode and retrieve the US valence in order to show EC effects, or that they need to deliberately infer a relation between CS and US. Note, however, that affect-as-information can also lead to an overestimation of the aware EC effects found in our studies.

Our findings are also inconsistent with the implicit misattribution account (March et al., 2019). In that regard, our results question the generalizability of one particular result in support of the misattribution account. Our studies used USs of low evocativeness that were presented simultaneously with the CSs. In one of the studies, we even had a condition where USs varied between trials. According to the misattribution account, these should be the optimal conditions to detect unaware EC (Jones et al., 2009; March et al., 2019) – which we did not find. Therefore, our results align with current research questioning the implicit misattribution account (Moran et al., 2021).

From a dual-process perspective, our results do not offer evidence for associative attitude change (Gawronski & Bodenhausen, 2018). Associative attitude formation is supposed to depend solely on the mere co-occurrence of CS and US by an automatic formation of links in memory. Because these links should establish even without conscious awareness of (or explicit memory of) the pairings, associative EC should also emerge when participants are unaware of the pairings. This was not the case in our experiments. Overall, this aligns with recent findings, suggesting that the evidence for associative attitude change is rather mixed (for reviews, see Corneille & Stahl, 2018; Hütter, 2022).

Beyond these insights for current EC theories, our findings also show when and why EC effects may reverse. EC is defined as the *change* in the liking of a CS due to its pairing with a positive/negative stimulus (De Houwer, 2007), thereby leaving formally open whether the pairings lead to assimilation (positive US positive CS) or contrast (positive US negative CS). The common finding in EC is assimilation (Alves & Imhoff, 2023; Hofmann et al., 2010); however, under certain circumstances, EC effects can reverse. For example, research supporting the propositional account shows that a contrastive CS-US relation (e.g., CS is opposed to the US) leads to contrast effects (De Houwer et al., 2020; Kurdi et al., 2023; Unkelbach & Fiedler, 2016). We show that the contrastive unaware EC effect found in previous research is likely not due to a specific cognitive process that is theoretically meaningful for EC. Instead, the reversed unaware effect seems to be based partly on interindividual differences in the affective reaction toward the US. These differences, combined with a biased awareness measure, lead to the pattern that unaware EC effects empirically reverse. This point brings us to the methodological implications of our findings.

Methodological Implications for EC

Our research identifies a new pitfall in VAMs – individual shifts in US valence leading to an underestimation of unaware EC. This might also explain some of the reversed unaware EC effects found in previous research (Alves & Imhoff, 2023; Halbeisen et al., 2014; Stahl et al., 2009; Waroquier et al., 2020). In a few cases, subjective US valence might have diverged from normed US valence, leading to a negative estimate of unaware EC. Our simulations show that even these few cases may be sufficient to cause such a bias. Yet, we also provide a simple solution. Merely assessing binary US evaluations suffices to control for the bias and leads to more accurate estimates of unaware EC.

However, even though the bias from valence shifts can be controlled for, another type of bias cannot be controlled for – affect-as-information (Schwarz, 2011). Previous research has already discussed this problem (Bar-Anan & Amzaleg-David, 2014; Corneille & Stahl, 2018; Hütter et al., 2012; Stahl et al., 2009; Waroquier et al., 2020). Our findings suggest that affect-as-information can lead to an underestimation of unaware EC, independent of US valence shifts, even in the corrected VAM. This bias has a stronger impact on the estimated unaware EC effect the less participants are aware of the pairings.

In combination, both biases severely decrease the chances of finding unaware EC with VAMs. Also, they give a too optimistic estimate of aware EC. Even though VAMs are very economic and can be implemented even in less controlled study designs, we recommend not using VAMs if the goal is to accurately estimate aware and unaware EC effects. Therefore, the question arises which alternative methods should be used, and to what extent shifts in individual US valence may also bias other methods.

First, we have not yet discussed VAMs with an additional "don't know" response option, which is supposed to capture mere guessing when participants are truly unaware. We also conducted a simulation with such a measure, which is provided on the OSF. Under the assumption that the button is used exclusively if people are genuinely unaware, the major change to our first simulation is that responses where participants indicate the opposite normed valence (e.g., "positive" responses when the US was normatively negative) now entirely capture reversed aware EC, thus leading to a bias that is entirely contingent on the strength of aware EC, even when the probability of a valence shift is very low and independent of the overall levels of awareness.

Second, Waroquier et al. (2020) recently developed a new VAM task, which is the standard VAM followed with a second question assessing participants' subjective attribution of memory – actual memory, a feeling/intuition, the CS valence, or random guessing. Interestingly, Waroquier et al. (2020) found evidence of "random guessing" EC (see also Jurchiș et al., 2020), although Stahl et al. (2023) did not replicate this finding. Notably, Waroquier et al. (2020) also found a strong reversed EC effect when participants answer in the VAM were classified as "unaware". According to our reasoning, this reversed unaware EC effect should be strong for responses where participants indicated actual memory or using the CS evaluation in the second question, and weakest for random guessing. A re-analysis of Waroquier et al. (2020) suggest that this is indeed the case (see the OSF for detailed results), but there are only few data points to test this, and therefore a replication of Waroquier et al. (2020) with an additional US evaluation is necessary.

Third, an alternative and more sophisticated method is the process dissociation technique (Hütter et al., 2012; Hütter & Sweldens, 2013; Mierop et al., 2017). This task requires participants in an inclusion condition to respond with "pleasant" or "unpleasant" depending on the US valence the CS had been shown with. If participants in the inclusion condition do not remember, they are instructed to use the same response keys to report their attitudes toward the CS. Participants in an exclusion condition are instructed to reverse their responses in the latter case, that is, responding "pleasant" for a negative attitude and "unpleasant" for a positive attitude towards the CS. This allows separating explicit memory from memory-independent EC using a multinomial processing tree (Erdfelder et al., 2009). The first task in this procedure (i.e., responding based on valence memory) is similar to the VAM; therefore, one may suspect that US valence shifts also impact this technique. However, in contrast to VAMs, US valence shifts introduce noise but no bias to the dissociation procedure. This might result in a bad model fit but not in a systematic bias of unaware EC.

Last, a very recent technique is the two-button procedure developed by Stahl and colleagues (2023). Here, participants receive two button pairs with “pleasant” and “unpleasant” and are required to use the first button pair if they remember the US valence and the second if they do not remember it. Because of these separate button pairs, shifts in US valence should not bias unaware EC. Even if participants found a normed positive US negative and were aware of the pairing, they will nevertheless use the first button pair, which does not influence the estimate for unaware EC.

In summary, only VAMs, but not the process dissociation procedure (Hütter et al., 2012) or the two-button test (Stahl et al., 2023) seem to be biased by individual US valence shifts. Thus, even though VAMs are arguably the more parsimonious measure, they are not an adequate method to estimate unaware EC.

Implications beyond EC

Beyond EC, our research offers valuable implications for dissociating between aware and unaware (or memory-dependent and memory-independent) processes in general. Any memory test that requires a response based on a psychological construct (e.g., valence) will inevitably misclassify responses as long as the manipulation has imperfect construct validity. Such memory tests are common for instance in the source memory literature (e.g., Bell et al., 2012; Symeonidou & Kuhlmann, 2022; Ventura-Bort et al., 2020). Depending on the specific effect of interest, this may lead to an under- or overestimation of unaware processes.

For example, let us presume that we study the effect of trustworthy behavior (e.g., lending other people money, cheating on a test) on the likeability of individuals and that we want to investigate whether the effect of the trustworthiness manipulation prevails when participants forget the actual information. Suppose our memory test asks whether a certain individual had shown trustworthy or untrustworthy behavior. In that case, we get the same problem as with the VAM and aware/unaware EC: Even if a pilot study has shown that the behaviors we use as stimuli are mostly trustworthy or untrustworthy to participants, some participants may have a different perception. Therefore some “unaware” responses may actually be aware but with the opposite effect than intended. Due to this bias, trustworthiness information may appear to have no or even a negative influence if participants forget the information.

Limitations

Despite several strengths, such as theoretical reasoning supported by mathematical simulations, or sufficiently powered preregistered experiments, there are several limitations of our research.

First, we argue that reversed unaware EC effects reported in previous research may be due to subjective valence shifts. Our findings suggest that this may indeed be the case. However, an exact replication of these studies would be necessary to state this confidently. We do not consider the increase in knowledge sufficient to take this effort, also, because the US pictures used in these studies are outdated and thus might lead to different subjective valence shifts nowadays (Schwarz & Strack, 2014). Related to this point, our studies were conducted online, which is a less controlled environment than a laboratory. However, direct comparisons often show no inferior data quality of online experiments, even for reaction-time-based tasks (Hilbig, 2016; Houben & Wiers, 2008).

Second, we treated valence conceptually as a binary construct in this research. However, CSs and USs may not only be positive and negative but also neutral. Also, all valence judgments are influenced by measurement error, and therefore, deviations from the normed valence can, in part, also reflect random fluctuation. Future research might investigate whether a more fine-grained assessment of valence memory or US valence improves the correction for valence shifts.

Lastly, we do not find any evidence for unaware EC using an objective memory test. However, this does not prove that unaware EC does not exist and that it could not be measured in the future by developing further refined procedures. In that regard, our research identifies a problem in one frequently used method, but as discussed above, this problem does not apply to more recent and arguably more sophisticated methods to test the contribution of awareness. Accordingly, our findings should further motivate EC researchers to embrace these new methods.

Contributions

Contributed to conception and design: MI, JW, NP, MW, HA

Contributed to acquisition of data: MI, NP, JW, HA

Contributed to analysis and interpretation of data: MI

Drafted and/or revised the article: MI, JW, HA

Approved the submitted version for publication: MI, JW, NP, MW, HA

Acknowledgements

We acknowledge support by the Open Access Publication Funds of the Ruhr-Universität Bochum.

Funding Information

This research was supported by the Open Access Publication Funds of the Ruhr University Bochum.

Competing Interests

The authors have no competing interests to declare.

Supplemental Material

All supplemental material can be found on this paper's project page on the OSF: https://doi.org/10.17605/OSF.IO/TPA3V

Data Accessibility Statement

All the stimuli, presentation materials, participant data, and analysis scripts can be found on this paper's project page on the OSF: https://doi.org/10.17605/OSF.IO/TPA3V

Submitted: June 04, 2023 PDT, Accepted: July 16, 2023 PDT

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-t

Alves, H., Högden, F., Gast, A., Aust, F., & Unkelbach, C. (2020). Attitudes from mere co-occurrences are guided by differentiation. *Journal of Personality and Social Psychology*, *119*(3), 560–581. https://doi.org/10.1037/pspa0000193

Alves, H., & Imhoff, R. (2023). Evaluative context and conditioning effects among same and different objects. *Journal of Personality and Social Psychology*, *124*(4), 735–753. https://doi.org/10.1037/pspa0000323

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

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1992). The content of learning in human evaluative conditioning: Acquired valence is sensitive to US-revaluation. *Learning and Motivation*, *23*(2), 200–224. https://doi.org/10.1016/0023-9690(92)90018-h

Bar-Anan, Y., & Amzaleg-David, E. (2014). The effect of evaluation on co-occurrence memory judgement. *Cognition and Emotion*, *28*(6), 1030–1046. https://doi.org/10.1080/02699931.2013.867835

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*(12), 2313–2335. https://doi.org/10.1080/17470211003802442

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., & Fox, J. (2019). *Package „lme4"*. https://cran.r-project.org/web/packages/lme4/lme4.pdf

Bell, R., Giang, T., & Buchner, A. (2012). Partial and specific source memory for faces associated to other- and self-relevant negative contexts. *Cognition & Emotion*, *26*(6), 1036–1055. https://doi.org/10.1080/02699931.2011.633988

Bohner, G., & Dickel, N. (2011). Attitudes and Attitude Change. *Annual Review of Psychology*, *62*(1), 391–417. https://doi.org/10.1146/annurev.psych.121208.131609

Corneille, O., & Lush, P. (2023). Sixty Years After Orne's *American Psychologist* Article: A Conceptual Framework for Subjective Experiences Elicited by Demand Characteristics. *Personality and Social Psychology Review*, *27*(1), 83–101. https://doi.org/10.1177/10888683221104368

Corneille, O., & Stahl, C. (2018). Associative Attitude Learning: A Closer Look at Evidence and How It Relates to Attitude Models. *Personality and Social Psychology Review*, *23*(2), 161–189. https://doi.org/10.1177/1088868318763261

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*(2), 176–187. https://doi.org/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*(2), 230–241. https://doi.org/10.1017/s1138741600006491

De Houwer, J. (2018). Propositional Models of Evaluative Conditioning. *Social Psychological Bulletin*, *13*(3), 1–21. https://doi.org/10.5964/spb.v13i3.28046

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*(6), 853–869. https://doi.org/10.1037/0033-2909.127.6.853

De Houwer, J., Van Dessel, P., & Moran, T. (2020). Chapter Three—Attitudes beyond associations: On the role of propositional representations in stimulus evaluation. In B. Gawronski (Ed.), *Advances in Experimental Social Psychology* (Vol. 61, pp. 127–183). Academic Press. https://doi.org/10.1016/bs.aesp.2019.09.004

Dedonder, J., Corneille, O., Bertinchamps, D., & Yzerbyt, V. (2013). Overcoming Correlational Pitfalls: Experimental Evidence Suggests That Evaluative Conditioning Occurs for Explicit But Not Implicit Encoding of CS–US Pairings. *Social Psychological and Personality Science*, *5*(2), 250–257. https://doi.org/10.1177/1948550613490969

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial Processing Tree Models. *Zeitschrift für Psychologie / Journal of Psychology*, *217*(3), 108–124. https://doi.org/10.1027/0044-3409.217.3.108

Fan, X., Bodenhausen, G. V., & Lee, A. Y. (2021). Acquiring favorable attitudes based on aversive affective cues: Examining the spontaneity and efficiency of propositional evaluative conditioning. *Journal of Experimental Social Psychology*, *95*, 104139. https://doi.org/10.1016/j.jesp.2021.104139

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Förderer, S., & Unkelbach, C. (2013). On the stability of evaluative conditioning effects: The role of identity memory, valence memory, and evaluative consolidation. *Social Psychology*, *44*(6), 380–389. https://doi.org/10.1027/1864-9335/a000150

Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019). A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology*, *117*(3), 522–559. https://doi.org/10.1037/pspa0000160

French, A. R., Franz, T. M., Phelan, L. L., & Blaine, B. E. (2013). Reducing Muslim/Arab Stereotypes Through Evaluative Conditioning. *The Journal of Social Psychology*, *153*(1), 6–9. https://doi.org/10.1080/00224545.2012.706242

Gast, A. (2018). A Declarative Memory Model of Evaluative Conditioning. *Social Psychological Bulletin*, *13*(3), 1–23. https://doi.org/10.5964/spb.v13i3.28590

Gawronski, Bertram, & Bodenhausen, G. V. (2018). Evaluative Conditioning From the Perspective of the Associative-Propositional Evaluation Model. *Social Psychological Bulletin*, *13*(3 SE-Review Article), 1–33. https://doi.org/10.5964/spb.v13i3.28024

Halbeisen, G., Blask, K., Weil, R., & Walther, E. (2014). The role of recollection in evaluative conditioning. *Journal of Experimental Social Psychology*, *55*, 162–168. https://doi.org/10.1016/j.jesp.2014.07.005

Heycke, T., Gehrmann, S., Haaf, J. M., & Stahl, C. (2018). Of two minds or one? A registered replication of Rydell et al. (2006). *Cognition and Emotion*, *32*(8), 1708–1727. https://doi.org/10.1080/02699931.2018.1429389

Heycke, T., & Stahl, C. (2020). No evaluative conditioning effects with briefly presented stimuli. *Psychological Research*, *84*(4), 1020–1027. https://doi.org/10.1007/s00426-018-1109-1

Hilbig, B. E. (2016). Reaction time effects in lab- versus Web-based research: Experimental evidence. *Behavior Research Methods*, *48*(4), 1718–1724. https://doi.org/10.3758/s13428-015-0678-9

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*(3), 390–421. https://doi.org/10.1037/a0018916

Högden, F., Hütter, M., & Unkelbach, C. (2018). Does evaluative conditioning depend on awareness? Evidence from a continuous flash suppression paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(10), 1641–1657. https://doi.org/10.1037/xlm0000533

Houben, K., & Wiers, R. W. (2008). Measuring implicit alcohol associations via the Internet: Validation of Web-based implicit association tests. *Behavior Research Methods*, *40*(4), 1134–1143. https://doi.org/10.3758/brm.40.4.1134

Hütter, M. (2022). An integrative review of dual- and single-process accounts of evaluative conditioning. *Nature Reviews Psychology*, *1*(11), 640–653. https://doi.org/10.1038/s44159-022-00102-7

Hütter, M., & Sweldens, S. (2013). Implicit misattribution of evaluative responses: Contingency-unaware evaluative conditioning requires simultaneous stimulus presentations. *Journal of Experimental Psychology: General*, *142*(3), 638–643. https://doi.org/10.1037/a0029989

Hütter, M., Sweldens, S., Stahl, C., Unkelbach, C., & Klauer, K. C. (2012). Dissociating contingency awareness and conditioned attitudes: Evidence of contingency-unaware evaluative conditioning. *Journal of Experimental Psychology: General*, *141*(3), 539–557. https://doi.org/10.1037/a0026477

Ingendahl, M., & Vogel, T. (2022). Stimulus Evaluation in the Eye of the Beholder: Big Five Personality Traits Explain Variance in Normed Picture Sets. *Personality Science*, *3*. https://doi.org/10.5964/ps.7951

Ingendahl, M., & Vogel, T. (2023). (Why) Do Big Five Personality Traits moderate Evaluative Conditioning? The role of US extremity and pairing memory. *Collabra: Psychology*, *9*(1), 74812. https://doi.org/10.17605/OSF.IO/3UJ7K

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. https://doi.org/10.1016/0749-596x(91)90025-f

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied*, *79*(1), 1–17. https://doi.org/10.1037/h0093874

Jones, C. R., Fazio, R. H., & Olson, M. A. (2009). Implicit misattribution as a mechanism underlying evaluative conditioning. *Journal of Personality and Social Psychology*, *96*(5), 933–948. https://doi.org/10.1037/a0014747

Jurchiş, R., Costea, A., Dienes, Z., Miclea, M., & Opre, A. (2020). Evaluative conditioning of artificial grammars: Evidence that subjectively-unconscious structures bias affective evaluations of novel stimuli. *Journal of Experimental Psychology: General*, *149*(9), 1800–1809. https://doi.org/10.1037/xge0000734

Koranyi, N., Gast, A., & Rothermund, K. (2013). „Although quite nice, I was somehow not attracted by that person“: Attitudes toward romantically committed opposite-sex others are immune to positive evaluative conditioning. *Social Psychological and Personality Science*, *4*(4), 403–410. https://doi.org/10.1177/1948550612467037

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods*, *49*(2), 457–470. https://doi.org/10.3758/s13428-016-0715-3

Kurdi, B., Morehouse, K. N., & Dunham, Y. (2023). How do explicit and implicit evaluations shift? A preregistered meta-analysis of the effects of co-occurrence and relational information. *Journal of Personality and Social Psychology*, *124*(6), 1174–1202. https://doi.org/10.1037/pspa0000329

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

Larzabal, C., Tramoni, E., Muratot, S., Thorpe, S. J., & Barbeau, E. J. (2018). Extremely long-term memory and familiarity after 12 years. *Cognition*, *170*, 254–262. https://doi.org/10.1016/j.cognition.2017.10.009

Leiner, D. (2019). *Sosci Survey*. https://www.soscisurvey.de

Luck, C. C., Patterson, R. R., & Lipp, O. V. (2021). Be careful what you say! – Evaluative change based on instructional learning generalizes to other similar stimuli and to the wider category. *Cognition and Emotion*, *35*(1), 169–184. https://doi.org/10.1080/02699931.2020.1816912

March, D. S., Olson, M. A., & Fazio, R. H. (2019). The Implicit Misattribution Model of Evaluative Conditioning. *Social Psychological Bulletin*, *13*(3 SE-Review Article), 1–25. https://doi.org/10.5964/spb.v13i3.27574

Mierop, A., Hütter, M., & Corneille, O. (2017). Resource Availability and Explicit Memory Largely Determine Evaluative Conditioning Effects in a Paradigm Claimed to be Conducive to Implicit Attitude Acquisition. *Social Psychological and Personality Science*, *8*(7), 758–767. https://doi.org/10.1177/1948550616687093

Mierop, A., Hütter, M., Stahl, C., & Corneille, O. (2019). Does attitude acquisition in evaluative conditioning without explicit CS-US memory reflect implicit misattribution of affect? *Cognition and Emotion*, *33*(2), 173–184. https://doi.org/10.1080/02699931.2018.1435505

Moran, T., Hughes, S., Hussey, I., Vadillo, M. A., Olson, M. A., Aust, F., Bading, K., Balas, R., Benedict, T., Corneille, O., Douglas, S. B., Ferguson, M. J., Fritzlen, K. A., Gast, A., Gawronski, B., Giménez-Fernández, T., Hanusz, K., Heycke, T., Högden, F., ... De Houwer, J. (2021). Incidental Attitude Formation via the Surveillance Task: A Preregistered Replication of the Olson and Fazio (2001) Study. *Psychological Science*, *32*(1), 120–131. https://doi.org/10.1177/0956797620968526

Moran, T., Nudler, Y., & Anan, Y. B. (2023). Evaluative Conditioning: Past, Present, and Future. *Annual Review of Psychology*, *74*(1), 245–269. https://doi.org/10.1146/annurev-psych-032420-031815

Olson, M. A., & Fazio, R. H. (2006). Reducing Automatically Activated Racial Prejudice Through Implicit Evaluative Conditioning. *Personality and Social Psychology Bulletin*, *32*(4), 421–433. https://doi.org/10.1177/0146167205284004

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*(1), 130–144. https://doi.org/10.1037/0278-7393.33.1.130

Schwarz, N. (2011). Feelings-as-information theory. *Handbook of Theories of Social Psychology*, *1*, 289–308.

Schwarz, N., & Strack, F. (2014). Does merely going through the same moves make for a "direct" replication? Concepts, contexts, and operationalizations. *Social Psychology*, *45*(4), 305–306.

Shanks, D. R. (2017). Regressive research: The pitfalls of post hoc data selection in the study of unconscious mental processes. *Psychonomic Bulletin & Review*, *24*(3), 752–775. https://doi.org/10.3758/s13423-016-1170-y

Smith, J. R., & Hogg, M. A. (2008). Social identity and attitudes. In *Attitudes and attitude change* (pp. 337–360). Psychology Press.

Stahl, C., & Aust, F. (2018). Evaluative Conditioning as Memory-Based Judgment. *Social Psychological Bulletin*, *13*(3 SE-Review Article), 1–30. https://doi.org/10.5964/spb.v13i3.28589

Stahl, C., Aust, F., Bena, J., Mierop, A., & Corneille, O. (2023). A conditional judgment procedure for probing evaluative conditioning effects in the absence of feelings of remembering. *Behavior Research Methods*. https://www.researchgate.net/publication/367545584_A_conditional_judgment_procedure_for_probing_evaluative_conditioning_effects_in_the_absence_of_feelings_of_remembering

Stahl, C., Haaf, J., & Corneille, O. (2016). Subliminal evaluative conditioning? Above-chance CS identification may be necessary and insufficient for attitude learning. *Journal of Experimental Psychology: General*, *145*(9), 1107–1131. https://doi.org/10.1037/xge0000191

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*(3), 404–420. https://doi.org/10.1037/a0016196

Sweldens, S., Corneille, O., & Yzerbyt, V. (2014). The Role of Awareness in Attitude Formation Through Evaluative Conditioning. *Personality and Social Psychology Review*, *18*(2), 187–209. https://doi.org/10.1177/1088868314527832

Sweldens, S., Van Osselaer, S. M. J., & Janiszewski, C. (2010). Evaluative Conditioning Procedures and the Resilience of Conditioned Brand Attitudes. *Journal of Consumer Research*, *37*(3), 473–489. https://doi.org/10.1086/653656

Symeonidou, N., & Kuhlmann, B. G. (2022). Better memory for emotional sources? A systematic evaluation of source valence and arousal in source memory. *Cognition and Emotion*, *36*(2), 300–316. https://doi.org/10.1080/02699931.2021.2008323

Timmermans, B., & Cleeremans, A. (2015). How can we measure awareness? An overview of current methods. In M. Overgaard (Ed.), *Behavioral Methods in Consciousness Research* (pp. 21–46). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199688890.003.0003

Unkelbach, C., & Fiedler, K. (2016). Contrastive CS-US Relations Reverse Evaluative Conditioning Effects. *Social Cognition*, *34*(5), 413–434. https://doi.org/10.1521/soco.2016.34.5.413

Ventura-Bort, C., Dolcos, F., Wendt, J., Wirkner, J., Hamm, A. O., & Weymar, M. (2020). Item and source memory for emotional associates is mediated by different retrieval processes. *Neuropsychologia*, *145*, 106606. https://doi.org/10.1016/j.neuropsychologia.2017.12.015

Vogel, T., & Wänke, M. (2016). *Attitudes and Attitude Change*. Taylor & Francis. https://doi.org/10.4324/9781315754185

Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology*, *82*(6), 919–934. https://doi.org/10.1037/0022-3514.82.6.919

Walther, E., & Nagengast, B. (2006). Evaluative conditioning and the awareness issue: Assessing contingency awareness with the Four-Picture Recognition Test. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(4), 454–459. https://doi.org/10.1037/0097-7403.32.4.454

Walther, E., Nagengast, B., & Trasselli, C. (2005). Evaluative conditioning in social psychology: Facts and speculations. *Cognition and Emotion*, *19*(2), 175–196. https://doi.org/10.1080/02699930441000274

Walther, E., Weil, R., & Düsing, J. (2011). The Role of Evaluative Conditioning in Attitude Formation. *Current Directions in Psychological Science*, *20*(3), 192–196. https://doi.org/10.1177/0963721411408771

Waroquier, L., Abadie, M., & Dienes, Z. (2020). Distinguishing the role of conscious and unconscious knowledge in evaluative conditioning. *Cognition*, *205*, 104460. https://doi.org/10.1016/j.cognition.2020.104460

Supplementary Materials

Peer Review History

Download: https://collabra.scholasticahq.com/article/87462-from-deviant-likes-to-reversed-effects-re-investigating-the-contribution-of-unaware-evaluative-conditioning-to-attitude-formation/attachment/178769.docx?auth_token=X4KRO-leRpQYxsdiX51H

Figures

Download: https://collabra.scholasticahq.com/article/87462-from-deviant-likes-to-reversed-effects-re-investigating-the-contribution-of-unaware-evaluative-conditioning-to-attitude-formation/attachment/178770.zip?auth_token=X4KRO-leRpQYxsdiX51H