Evidence Suggestive of Uncontrollable Attitude Acquisition replicates in an Instructions-based Evaluative Conditioning Paradigm: Implications for Associative Attitude Acquisition

Corneille, Olivier (corresponding author), University of Louvain, Belgium, Place Cardinal Mercier, 10 bte L3.05.01 - 1348, e-mail: olivier.corneille@uclouvain.be, phone number: +32 (10) 47 86 42

Mierop, Adrien, University of Louvain, Belgium, Place Cardinal Mercier, 10 bte L3.05.01 - 1348, e-mail: adrien.mierop@uclouvain.be, phone number: +32 (10) 47 86 40

Stahl, Christoph, Universität zu Köln, Herbert-Lewin-Strasse, 2 – 50931 Köln, Germany, e-mail: christoph.stahl@uni-koeln.de, phone number: +49 (221) 470-3428

Hütter, Mandy, Eberhard Karls Universität Tübingen, Psychology Department, Schleichstr. 4, 72076 Tübingen, Germany, e-mail: mandy.huetter@uni-tuebingen.de, phone number: +49 (7071) 29 72978

This is a preprint version. The published version may be found at

https://doi.org/10.1016/j.jesp.2019.103841

(or upon request to the first author)

Abstract

When presented with neutral stimuli (i.e., CSs) paired with valent ones (i.e., USs), individuals may prove unable to fully reverse the influence of the US on the impression they form about the CS. In a high-powered, pre-registered experiment, we revisited this uncontrollable EC effect in the context of an instruction-based EC procedure. Specifically, standard or reversed learning instructions were given to participants along with two sets of CSs (i.e., unfamiliar consumer products) told to be paired later with positive or negative USs. Evaluative ratings of the CSs were then immediately collected in the absence of direct exposure to CS-US pairings. Fully replicating effects found in the standard (i.e., experienced) EC paradigm, (i) the absolute EC effect was larger in the standard than reversed learning condition, and (ii) outcomes from Multinomial Processing Tree modelling suggested both controllable and uncontrollable EC effects. These findings imply that uncontrollable EC effects cannot be confidently interpreted as evidence for the operation of associative attitude learning.

Keywords: Attitude, Associative Attitude Learning, Dual Process Models, Evaluative Conditioning, Learning, Automaticity, Control.

Evidence Suggestive of Uncontrollable Attitude Acquisition replicates in an Instruction-based Evaluative Conditioning Paradigm: Implications for Associative Attitude Formation.

Introduction

Developed in the broader context of dual-process accounts of cognition (e.g., Evans, 2008; Evans & Stanovich, 2013), dual-learning models of attitudes posit that attitudes can be acquired via two modes. A first mode, coined "propositional", is thought to be based on a non-automatic syllogistic form of reasoning (e.g., Mitchell, De Houwer, & Lovibond, 2009). A second mode, coined "associative", is typically thought to be based on the slow-paced and automatic registration of co-occurrences between stimuli (e.g., Gawronski & Bodenhausen, 2011, 2014; Rydell & McConnell, 2006; Rydell, McConnell, Strain, Claypool, & Hugenberg, 2007; Smith & DeCoster, 2000).

Empirical evidence is scarce, however, that attitudes can be acquired automatically (i.e., unconsciously, efficiently, and independently of processing goals), except perhaps for the controllability feature of automaticity. Therefore, recent evidence for uncontrollable attitude formation represents one of the most convincing pieces of support for dual-learning models of attitude that posit an automatic route to attitude formation (for a detailed discussion, see Corneille & Stahl, 2018). In the present research, we examined whether evidence for uncontrollable attitude formation replicates in an instruction-based paradigm. That is, if uncontrollable EC effects can be elicited based on the brief communication of symbolic information about stimuli. We discuss below the theoretical implications of this question.

Evidence for uncontrollable attitude formation

To date, the most convincing support for uncontrollable attitude formation stems from evaluative conditioning (EC) studies by Hütter and Sweldens (2018). As a procedure, EC consists in the pairing of initially neutral stimuli (i.e., conditioned stimuli, or CSs) with valent stimuli (i.e., the unconditioned stimuli, or USs). As an effect, EC consists in the change in valence of the CSs resulting from their pairing with USs. Because of its simple structure, the EC procedure is the most frequently used paradigm for testing assumptions related to association formation in evaluative learning (e.g., Hahn & Gawronski, 2018).

Hütter and Sweldens (2018) tested the uncontrollability assumption by providing their participants with instructions prior to the CS-US pairing procedure. Participants were either requested to apply (i.e., standard condition) or to reverse (i.e., reversal condition) the influence of the USs to form an accurate impression of the CSs. An absolute EC effect of smaller magnitude was observed in the reversal as compared to the standard condition (see also Gawronski, Balas & Creighton, 2014). In other words, participants were able to reverse the EC effect, but not entirely so. Hütter and Sweldens (2018) suggested that evaluative influences of opposite directions (i.e., a controllable reversed evaluative effect, and an uncontrollable assimilative evaluative effect) could have interfered with each other in the reversal condition.

Hütter and Sweldens (2018) isolated and weighed these controllable and uncontrollable processes by relying on multinomial processing tree (MPT) modelling (e.g., Riefer & Batchelder, 1988).[1] They modelled expected response frequencies as a function of US valence and control instructions, as if both a controllable and an uncontrollable process would contribute to participants' evaluations. They estimated three parameters reflecting a

[1] We refer to uncontrollability as a feature of automaticity of a *process* (i.e., in the absence of full control, CS evaluations are assimilated to US valence instead of reflecting random judgments). None of the authors assumes, however, that EC *effects* cannot be fully controlled. See also the General Discussion for a nuanced discussion of how MPT parameters should be interpreted and validated.

controllable process (i.e., the *c*-parameter), an uncontrollable process (i.e., the *u*-parameter) and a response bias (i.e., the *r*-parameter). The *c*-parameter reflects the probability to correctly apply (standard condition) versus reverse (reversal condition) US valence in forming an impression of the CS. By contrast, the *u*-parameter reflects the probability of assimilative evaluations irrespective of instructions in both conditions. Estimates of both of these parameters were significantly larger than zero, suggesting that, in addition to a controllable process, an uncontrollable process contributed to the formation of attitudes in their experiments.

Hütter and Sweldens (2018) further found that the *c*-parameter, but not the *u*-parameter was sensitive to experimental manipulations implemented at learning (i.e., financial incentives or attentional load). This finding attested to the functional independence of these parameters. They were careful not to interpret their findings in terms of associative attitude formation. As just mentioned, however, dual-process models of attitude typically see associative attitude formation as an automatic process, and the possibility of uncontrollable attitude formation supports this view.

Instruction-based EC

In instruction-based EC procedures, participants are provided with instructions about upcoming events and evaluative measures are collected directly after this instruction stage. That is, they are collected in the absence of direct exposure to the CS-US pairings. Effects of instruction-based procedures have been investigated in a diversity of evaluative learning paradigms, such as mere exposure (Van Dessel, Mertens & Smith, 2017), approach-avoidance training (Van Dessel, De Houwer, Gast, & Smith, C., 2015), and evaluative conditioning (De Houwer, 2006; Gast & De Houwer, 2013; Hütter & De Houwer, 2017) paradigms.

Evaluative learning effects found in instruction-based procedures are consistent with a propositional account of evaluative learning. According to the propositional view, participants "derive symbolic meaning from a particular proximal event (i.e., an event that occurs here and now in space and time), being the words and sentences of which the instructions are composed" (De Houwer & Hughes, 2016, p. 482-483). Propositional learning predicts that evaluative learning effects can be quickly elicited based on formal reasoning applied to symbolic information.

Applied to our current uncontrollable EC question, a propositional account may suggest that, in order to reverse an impression, participants have to form a veridical impression in the first place. As a result, two propositions bearing opposite evaluative implications may be established about the CS-US relations under reversal instructions. It is also possible that one of these propositions (occasionally, the veridical one) shows a retrieval advantage at judgment time, resulting in veridical evaluations under reversal instructions. More generally, it should be noted that propositional models allow for the automatic formation of propositions about stimuli relations (De Houwer, 2018). Therefore, it is conceivable that participants automatically form propositions implying assimilative evaluations under reversal instructions conditions.

This propositional account stands in contrast with associative learning views, which typically assume a mere, slow-paced, registration of stimulus co-occurrences. Some associative models acknowledge the possibility of a fast formation of associations based on declarative information (e.g., Gawronski & Bodenhausen, 2006, 2011). However, the replication of evaluative learning effects in an instruction-based procedure lends credence to a reinterpretation of these effects in terms of propositional learning.

We examined here whether evidence for uncontrollable attitude formation (i.e., residual assimilative EC effect inferred from an incomplete reversal of evaluations under reversal instructions, and evidence for a significant *u*-parameter in the MPT model) is replicated in an instruction-based EC procedure. As just discussed, such effects would imply that uncontrollable EC effects cannot be confidently interpreted anymore as an indicator of associative attitude learning, as either a propositional account or a sophisticated associative account may explain these effects.

Method

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. The pre-registration, material, program script, raw data and analytic scripts data are publicly available on Open Science Framework (osf.io/eubs2).

Participants. To determine sample size, we drew on the effect size of the control instructions × US valence interaction on CS ratings obtained in a previous EC study that relied on a standard EC paradigm ($\eta^2_G = .07$). Using a mixed ANOVA with $\alpha = \beta = .05$, a sample of 169 participants is required to achieve high statistical power ($1 - \beta = .95$). We requested 180 participants on the online platform 'Testable Minds' to accommodate potential data loss. Eventually, we received data from 188 participants ($M_{age} = 35.22$, $SD_{age} = 11.84$; 87 females, 101 males).

Procedure. We adapted the procedure used by Hütter and Sweldens (2018, Exp. 4). The main changes are that (i) we slightly adapted the instructions for designing the instruction-based version of their procedure and (ii) we used different CSs (i.e., consumer products instead of brand logos). Specifically, participants were randomly assigned to either a standard or reversal condition with the following instructions (for the sake of comparison and brevity, identical paragraphs are referred to by using […] for the reversal condition):

Standard Condition

New products are introduced to the market on a regular basis. Previous research on advertising shows that we quickly form impressions of unknown products and that these impressions can be strongly influenced by the context in which we encounter a product *for the first time. Specifically, this means that we may learn to like products that appear with positive images and learn to dislike products that appear with negative images.*

In advertising, such images may very well provide information about a product. For instance, marketers might invest more money in advertising for products that have proved successful in the past or are expected to sell well. Conversely, low quality marketing might be reflective of a low-quality product.

In this research, we investigate whether people are able to APPLY the positive or negative quality of the images to the product. In the next phase, you will therefore be presented with different products appearing repeatedly with positive or negative images. Hence, you should start LIKING the product when paired with a POSITIVE image. Conversely, you should start DISLIKING the product when paired with a NEGATIVE image. Afterwards, we will ask you how you feel about the different products.

In the next phase of the experiment, a given product will be presented six times together with either a pleasant or an unpleasant picture. Pleasant pictures will depict, for example, babies, pets, or flowers. Unpleasant pictures will depict, for example, garbage, injuries, or obviously dangerous animals.

You will now see two consecutive screens. On one screen, you will see the products that will be presented with pleasant pictures during the perception phase. On the other screen, you will see the products that will be presented with unpleasant pictures. Each screen will be presented for 2.5 min. The transition between screens will proceed automatically.

It is very important that you remember which products will be presented with which type (i.e., positive or negative) of pictures. You will definitely need this information to finish the task successfully. After this memorization phase, you will go through the perception phase. Remember that, in that second - perception phase – it will be important that:

You LIKE pictures presented with POSITIVE pictures

You DISLIKE pictures presented with NEGATIVE pictures

Once you have paid careful attention to the above instructions, please press the spacebar to see the products on which you will be later asked to form an impression.

Reversal Condition

[…]
While such images may well provide information about a product in advertising, many cases are conceivable in which the direct application of the positive or negative quality of an image would lead to erroneous impressions. For example, products may be of low quality despite positive advertising. Conversely, high-quality products can be put in a negative light by competitive products.
In this research, we investigate whether people are able to REVERSE the influence of the positive or negative quality of the images to the product. In the next phase, you will therefore be presented with different products repeatedly appearing together with positive or negative images. Hence, you should start DISLIKING the product when paired with a POSITIVE image. Conversely, you should start LIKING the product when paired with a NEGATIVE image. Afterwards, we will ask you how you feel about the different products.
[…]
You DISLIKE pictures presented with POSITIVE pictures
You LIKE pictures presented with NEGATIVE pictures
[…]

Participants were then presented with two picture sets, each consisting of 12 unfamiliar consumer products (i.e., the CSs), presumably to be paired with positive or negative pictures (i.e., the US) in the perception phase of the experiment. The order of the presentation of the sets was randomized. The two picture sets were headed, respectively:

"These products will be presented with PLEASANT pictures during the perception phase:"

"These products will be presented with UNPLEASANT pictures during the perception phase:"

Following these instructions, each picture set was presented for 151 seconds, which amounted to the total CS-US exposure time in Hütter and Sweldens (2018; Exp. 4).

Participants then rated the CSs on a scale ranging from 1 (very negative) to 9 (very positive (in order to probe the asymmetry in absolute EC effect between the standard and reversal conditions), and on a dichotomous evaluation task ("positive" vs. "negative"; for MPT modelling).

Results

Data were analyzed using R. The evaluative ratings were analyzed using the 'ezANOVA' and 'anovaBF' functions (from the 'ez' and the 'BayesFactor' packages, Lawrence, 2016; Morey & Rouder, 2015). MPT modelling was conducted using the *multiTree* software (Moshagen, 2010).

Evaluative ratings. CS ratings were submitted to a 2 (US valence: positive or negative) × 2 (Instructions: standard or reversal) mixed-design ANOVA, with the second factor varying between participants (see Table 1 for the full pattern of means). We observed no main effect of Instructions, $F(1, 186) = 1.42$, $\eta^2_g = .00$, $p = .23$, BF01 = 5.93 ± 7%. We obtained a main effect of US valence, $F(1, 186) = 33.84$, $\eta^2_g = .12$, $p < .001$, BF10 > 1000: On average, participants rated CSs paired with positive USs more positively ($M = 5.94$, $SD = 2.02$) than CSs paired with negative USs ($M = 4.63$, $SD = 2.19$). This effect was qualified by an Instructions × US valence interaction, $F(1, 186) = 89.35$, $\eta^2_g = .25$, $p < .001$, BF10 > 1000: In absolute value, the instruction-based EC effect was stronger in the standard condition ($F(1,107) = 141.19$, $\eta^2_g = .47$, $p < .001$, BF10 > 1000) than in the reversal condition ($F(1,79) = 8.86$, $\eta^2_g = .08$, $p = .004$, BF10 = 93.83 ± 2.28%), $F(1,186) = 18.88$, $\eta^2_g = .09$, $p < .001$, BF10 = 78.16 ± 0% (see Figure 1).

MPT modelling. The initial model with three free parameters (*c*, *u*, *r*) fitted the data well, $G^2(1) = 0.66$, $p = .42$. The estimate of the *c*-parameter, $c = .36$; 95% CI [.33, .38], was significantly larger than 0, $\Delta G^2(1) = 609.36$, $p < .001$, suggesting the contribution of a controllable process (i.e., evaluations in line with reversal instructions). The estimate of the *u*-parameter, $u = .30$; 95% *CI* [.26, .34], was also significantly larger than 0, $\Delta G^2(1) = 201.68$, $p < .001$, suggesting the contribution of an uncontrollable assimilative process (i.e., evaluations reflecting US valence regardless of standard or reversal instructions). The *r*-parameter, $r = .62$; 95% *CI* [.59, .65, was significantly different from 0.5, $\Delta G^2(1) = 65.47$, $p < .001$, indicating a response bias toward the 'positive' response.

General Discussion

The present research replicates uncontrollable EC effects in an instruction-based EC procedure: a smaller absolute EC effect was found in the instructed reversal than in the instructed standard condition. In addition, MPT modelling suggests the contribution of both controllable and uncontrollable processes. The presence of a *u*-parameter in an instruction-based EC procedure is particularly informative. The advantage of a process dissociation approach is that it allows modelling independent parameters while keeping the task constant (Hütter & Klauer, 2016). Therefore, conclusions made about mental processes are not contingent on comparisons made across (e.g., direct versus indirect) tasks. Task dissociation procedures likely involve a host of hardly interpretable structural and functional differences (for a discussion in the context of the controllability of EC, see Hütter & Sweldens, 2018).

Uncontrollable EC is interpreted as one of the most compelling pieces of evidence for associative attitude formation (see Corneille & Stahl, 2018). The replication of a *u*-parameter in an instruction-based EC procedure implies that uncontrollable EC effects cannot be confidently interpreted anymore as an indicator of associative learning. We explain why this is the case in the remainder of this discussion.

Associative learning typically assumes that associations are formed based on the mere registration of stimulus co-occurrences in the environment, and that "codes used by the associative system are not verbalizable, not easily used for attributions, not easily converted into propositions, and not easily used for syllogistic reasoning" (McConnell & Rydell, 2014, p. 214). Even though symbolic information contained in instructions also involves spatio-temporal co-occurrences (i.e., words in a sentence), it would be bold to assume that participants *merely register* these co-occurrences (De Houwer & Hughes, 2016). Rather, participants likely generate information and inferences relevant to the co-occurrences that are symbolically referred to in the instructions. Our understanding is that this active self-generation process does not qualify as a mere registration of stimulus inputs in current theories of associative attitude learning. However, these theories may be adapted so as to cover the possibility of a more active role of the observer.

Alternatively, one may conclude from our findings that associative attitude learning allows for the fast formation of associations based on declarative information (e.g., Gawronski & Bodenhausen, 2006, 2011). In this case, however, a single-process propositional learning account would still provide a straightforward and parsimonious alternative to dual-process views that assume this more sophisticated form of associative learning.

Finally, a third possible conclusion is that the evidence in favor of uncontrollable EC reported here and in past research is actually misinterpreted as indicative of an uncontrollable attitude formation process. If this was the case, the present evidence would be irrelevant to uncontrollable attitude formation. We believe it is worth briefly elaborating on this third possibility. MPT models represent a set of assumptions about how the data are generated, and these assumptions are reflected in the parameters' labels (e.g., u = uncontrollable). But there is no guarantee of a one-to-one relationship between model parameters and latent psychological processes. Finding a non-zero u-parameter does not in itself imply that instruction-based EC is partly uncontrollable. Learning-independent mechanisms may contribute to the size of the u-parameter in both the experiential and instructed paradigms. Therefore, psychological interpretation of the parameters needs to be validated in experimental studies (cf. Hütter & Klauer, 2016). Hütter and Sweldens (2018) conducted experiments to support the interpretation of the u-parameter as an uncontrollable learning process. Nevertheless, all authors of the current article believe that more research is needed to further illuminate the interpretation of the *u*-parameter and its relation to theories of attitude formation.

Conclusion:

The three interpretational possibilities discussed in the General Discussion all converge towards the same conclusion: the replication of uncontrollable EC effects in an instruction-based EC procedure is consistent with either a propositional account of these effects or a sophisticated associative one. Of these two possible accounts, however, a propositional learning account currently provides the most parsimonious explanation of uncontrollable evaluative learning effects.

References

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*, 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,* 176–187. doi: 10.1016/j.lmot.2005.12.002

De Houwer, J. (2018). Propositional models of evaluative conditioning. *Social Psychological Bulletin*, *13*, e28046.

De Houwer, J., & Hughes, S. (2016). Evaluative conditioning as a symbolic phenomenon: On the relation between evaluative conditioning, evaluative conditioning via instructions, and persuasion. *Social Cognition*, *34*(5). Special Issue: Conceptual, Theoretical, and Methodological Challenges in Evaluative Conditioning Research, 480-494. Doi: 10.1521/soco.2016.34.5.480.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255-278.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223-241.

Gast, A., & De Houwer, J. (2012). Evaluative conditioning without directly experienced pairings of the conditioned and the unconditioned stimuli. *The Quarterly Journal of Experimental Psychology, 65*, 1657-1674.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*(5), 692-731.

Gawronski, B., & Bodenhausen, G. V. (2011). The Associative-Propositional Evaluation Model: Theory, Evidence, and Open Questions. *Advances in Experimental Social Psychology*, *44*, 59-127.

Gawronski, B., & Bodenhausen, G. V. (2014). The associative-propositional evaluation model: Operating principles and operating conditions of evaluation. *Dual-process theories of the social mind*, 188-203.

Gawronski, B., Balas, R., & Creighton, L. A. (2014). Can the formation of conditioned attitudes be intentionally controlled? *Personality and Social Psychology Bulletin*, *40*(4), 419-432.

Hahn, A., & Gawronski, B. (2018). Implicit social cognition. In J. T. Wixted (Ed.), *The Stevens' handbook of experimental psychology and cognitive neuroscience, Volume 4* (4th ed., pp. 395-427). New York, NY: Wiley.

Hütter, M., & De Houwer, J. (2017). Examining the contributions of memory-dependent and memory-independent components to evaluative conditioning via instructions. *Journal of Experimental Social Psychology*, *71*, 49-58.

Hütter, M., & Klauer, K. C. (2016). Applying processing trees in social psychology. *European Review of Social Psychology*, *27*(1), 116-159.

Hütter, M., & Sweldens, S. (2018). Dissociating controllable and uncontrollable effects of affective stimuli on attitudes and consumption. *Journal of Consumer Research, 45,* 320-349.

McConnell, A. R., & Rydell, R. J. (2014). The Systems of Evaluation Model: A dual-systems approach to attitudes. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 204-217). New York: Guilford.

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

Riefer, D. M. & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95,* 318-339.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: a systems of reasoning analysis. *Journal of personality and social psychology*, *91*(6), 995.

Rydell, R. J., McConnell, A. R., Strain, L. M., Claypool, H. M., & Hugenberg, K. (2007). Implicit and explicit attitudes respond differently to increasing amounts of counterattitudinal information. *European Journal of Social Psychology*, *37*(5), 867-878.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*(2), 108-131.

Van Dessel, P., De Houwer, J., Gast, A., & Smith, C. (2015). Instruction-based approach-avoidance effects: changing stimulus evaluation via the mere instruction to approach or avoid stimuli. *Experimental Psychology*, *62*(3), 161–169.

Van Dessel, P., Mertens, G., & Smith, C. (2017). The mere exposure instruction effect: mere exposure instructions influence liking. *Experimental Psychology*, *64*(5), 299–314.

Table 1. Response frequencies observed in the dichotomous evaluative measure as a function of US valence and instructions.

	Standard		Reversal	
	CS+	CS-	CS+	CS-
“positive” responses	1075	354	462	615
“negative” responses	221	942	498	345

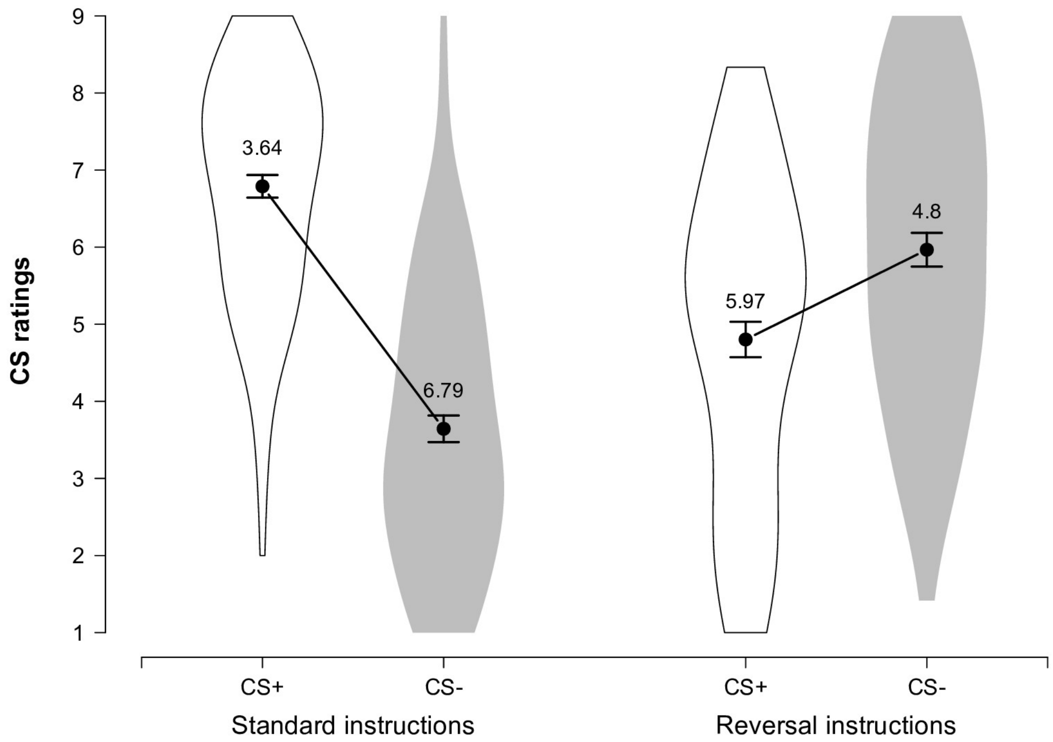

Figure 1. Evaluative ratings of the CSs as a function of US valence and instruction condition. Filled circles represent observed means. Error bars represent standard errors of means. CS- = negatively conditioned stimuli. CS+ = positively conditioned stimuli.