Observational evaluative conditioning is sensitive to relational information

Sarah Kasran, Sean Hughes, and Jan De Houwer

Ghent University

Word count (excluding abstract and references): 14348

Author Note

SK, SH, & JDH, Department of Experimental Clinical and Health Psychology, Ghent University. This research was conducted with the support of PhD fellowship FWO18/ASP/119 from the Research Foundation Flanders (FWO) to SK and grant BOF16/MET_V/002 from Ghent University to JDH. Correspondence concerning this article should be sent to Sarah.Kasran@UGent.be.

Abstract

Social learning represents an important avenue via which evaluations can be formed or changed. Rather than learn slowly through trial and error, we can instead observe how another person (a "model") interacts with stimuli and quickly adjust our own behaviour. We report five studies (*n* = 912) that focused on one subtype of social learning, observational evaluative conditioning (OEC), and how it is moderated by relational information (i.e., information indicating how a stimulus and a model's reactions are related). Participants observed a model reacting positively to one stimulus and negatively to another, and were either told that these reactions were genuine, faked, or opposite to the model's actual feelings. Stimulus evaluations were then indexed using ratings and a personalised Implicit Association Test (pIAT). When the model's reactions were said to be genuine, OEC effects emerged in the expected direction. When the model's reactions were said to be faked, the magnitude of self-reported, but not pIAT, effects was reduced. Finally, stating that the model's reactions were opposite to his actual feelings eliminated or reversed self-reported effects and eliminated pIAT effects. We consider how these findings relate to previous work as well as mental-process theories.

Keywords: social learning, observational conditioning, evaluations, relational information

Observational evaluative conditioning is sensitive to relational information

Why do we like some things and dislike others? Although many of our preferences arise from our personal experiences with stimuli, others are acquired and changed via social learning. How we feel about other people, brand products, political ideas, and situations might be heavily influenced by observing the “emotional responses of another person, as conveyed through vocal, facial, and postural manifestations” (Bandura, 1971, p. 13). Many instances of social evaluative learning take place in everyday life. For example, advertisements often seek to persuade viewers of a product’s virtues by displaying how others react as they interact with it. Similarly, we may come to dislike an animal (e.g., a dog) or activity (e.g., flying) after witnessing another person display fear in its presence.

In this paper, we focus on a subtype of social evaluative learning called observational conditioning. This term was originally introduced by Mineka et al. (1984), who found that rhesus monkeys (“observers”) reacted fearfully towards a stimulus after observing another monkey (a “model”) react fearfully to that same stimulus. Baeyens et al. (1996, 2001) extended this research into the domain of attitudes and found that people’s stimulus *evaluations* (i.e., what they like and dislike) also changed when a stimulus was paired with another person’s emotional reactions. In their studies, children (observers) watched videos of a child of the same age (model) reacting neutrally after tasting one novel beverage and negatively after tasting another. Afterwards, the observers rated the beverage that had been followed by the model’s negative reaction as more negative than the other beverage. This work thus demonstrated that changes in liking can occur when people observe a regularity between a stimulus and a model’s evaluative reaction. Whereas the effect studied by Mineka and colleagues is typically referred to as observational fear conditioning, Baeyens and colleagues referred to their effect as observational evaluative conditioning (OEC).

Mental-Process Accounts of OEC

Although past research shows that observing a model's emotional reaction can influence an observer's own stimulus evaluations, it is not yet clear *how* this happens; that is, there is no consensus regarding the mental (i.e., cognitive) processes that are assumed to mediate OEC. Baeyens et al. (2001) forwarded two possible mental-process explanations based on previous theorizing in the observational fear conditioning literature. The first was an intuitively plausible *(social) inferential account* which assumes that an observer makes inferences about the evaluative properties of a stimulus based on how a model reacts to it. For example, a change in liking may occur because of the inference "this beverage is bad". This inference presumably relies on multiple premises, including that the model dislikes the beverage, that the observer is drinking the same beverage as the model, and that the observer and model have similar preferences (Baeyens et al., 2001).

Researchers have also proposed that a model's emotional reaction serves as an unconditioned stimulus (US) that may elicit an unconditioned response (UR) in the observer (e.g., Mineka & Cook, 1993). This proposal (which was referred to as the "direct conditioning hypothesis" by Baeyens et al., 2001) has been quite influential. Although the idea that the model's reaction functions as a US (and observational conditioning therefore constitutes an instance of classical conditioning) is not in itself incompatible with an inferential mental-process account, it has often been accompanied by the assumption that associative mental processes mediate observational conditioning effects (e.g., Askew & Field, 2008; Field, 2006; Heyes, 2012; Olsson & Phelps, 2007; Reynolds et al., 2015, 2018). Specifically, it is assumed that when a stimulus (the conditioned stimulus or CS) and a model's reaction (US) are paired with one another, an association will be formed between the mental representations of the CS and the US or, alternatively, between the representations of the CS and the UR. Because activation can spread via associations from one representation to another, presentation of the CS after CS-US pairings will result not only in the activation of the CS representation but also in the activation of the US representation and/or the UR representation. As such, the CS will elicit a similar emotional response in the observer as the one initially displayed by the model in response to the CS. According to this assumption, only the observer's experience of the CS-US pairings should matter, not the evaluation of the premises mentioned above. This *associative account* thus constitutes a second possible mental-process explanation of OEC effects.

Baeyens et al. (2001) conducted an initial empirical test of these different accounts. Yet their results did not provide clear support for one account over the other. On the one hand, when the observers were told that the model did not drink the same beverages as them, OEC effects failed to emerge. This suggests that if one undermines one of the crucial premises mentioned above, the observer will not make the final inference about the stimulus' evaluative properties. Hence, this finding seems consistent with an inferential account of OEC. On the other hand, there was evidence to suggest that telling observers that the model was not drinking the same beverages as they were reduced the observers' attention to the video of the model. If so, then the absence of an OEC effect could also have been explained by a reduced activation of the US or UR representations and thus a reduced opportunity for association formation (i.e., consistent with an associative account of OEC).

The authors also reported a dissociation between the emergence of OEC effects and the observers' memory for the spatiotemporal relations between the CSs and the model's negative reactions (i.e., contingency memory): OEC effects emerged even though very few observers could afterwards indicate which type of beverage had been followed by negative model reactions. Although this argues against the idea that OEC is mediated by conscious inferences about stimulus properties, it is worth noting that the contingency awareness measure that Baeyens et al. used was not optimal (for recent reviews of this issue see

Corneille & Stahl, 2019; Sweldens et al., 2014). Aside from the fact that this type of post-experimental measure assesses memory at the end rather than awareness during the learning phase, the measure may not have been sensitive enough. First, it required active effort on the part of the observers: rather than being asked to separately taste each beverage and report if it had been followed by a negative reaction, they were free to look at and taste all beverages and then asked to indicate which one had been followed by negative reactions, leaving open the possibility that not all beverages were considered. Second, the measure presented the beverages in a different context than during acquisition: while the beverages also contained an irrelevant feature (colour) during acquisition, they were colourless during the contingency memory test, which may have created confusion. In sum, we cannot conclude with certainty that the observers in these studies were actually unaware of the contingencies.

Taken together, the question of whether OEC effects are due to inferential or associative processes was not resolved by the original studies of Baeyens and colleagues (1996, 2001). Since then, however, findings have emerged in other areas of psychological science that may also inform this debate. In what follows we discuss some of these findings.

Findings from Other Social Learning Research

Even though different terminology is used in different areas of research (e.g., social transmission; Jones et al., 2007; Weisbuch et al., 2009; social referencing; Moses et al., 2001; Mumme & Fernald, 2003), social learning studies often involve pairing a stimulus with an emotional reaction of a model. Hence, the effects obtained in these studies could be considered as instances of OEC and provide information about its underlying processes.

Some of these findings appear to be in line with an inferential account. For example, a social referencing study found that infants' emotional responses towards ambiguous objects (novel toys) were influenced by how another person reacted in the presence of those objects, but only for objects that the other person looked at while showing the emotional reaction (Moses et al., 2001). In a different study wherein adult participants viewed pairings of neutral stimuli and pictures of emotional facial expressions, their evaluations of those stimuli also only changed in line with the facial expression if the gaze of the face was directed toward the stimuli (Bayliss et al., 2007). Such findings support an inferential account because they suggest that beliefs about the *relation* between a stimulus and a model's reaction matter. That is, if they simply co-occur but do not seem to be causally related, the observer might not infer evaluative properties of the stimulus from its co-occurrence with an emotional reaction.

In contrast, work elsewhere seems to argue against an inferential account. Within the domain of (racial) bias and prejudice, considerable research indicates that our evaluations of another person are sensitive to how other people behave nonverbally toward that individual, or even toward others from that person's social group (Castelli et al., 2008, 2012; Skinner et al., 2017, 2020; Skinner & Perry, 2020). Some researchers have reported evidence of this kind of social transmission of bias even when the pattern of nonverbal bias in the observed target-model interactions was very subtle and observers were not considered to be consciously aware of this pattern (Weisbuch et al., 2009; Weisbuch & Ambady, 2009). While this again suggests that contingency awareness may not be necessary for the effect to emerge, it is worth noting that the authors based this conclusion on the fact that a separate sample of participants in a small pilot study was unable to detect the pattern of nonverbal biased behaviour in the videos. Therefore, it remains to be seen if this conclusion would hold if contingency awareness and changes in liking were assessed in the same participants.

Taken together, research in other areas of social learning paints a mixed picture with regard to the role that inferences are assumed to play. In addition, while we consider these studies to be informative for the current debate, we should note that they often relied on different paradigms, which might limit generalization to more typical observational conditioning research. For example, participants in the study by Bayliss et al. (2007) did not see an actual person interacting with the target object, but simply viewed a picture of a face in the middle of the screen, which "changed" its emotional expression and gaze direction shortly before a picture of the target object was presented on the side of the screen.[1]

Findings from Evaluative Conditioning Research

Another literature that seems closely related to OEC research is that of evaluative conditioning (EC), which focuses on the impact of pairing stimuli on evaluative responses (e.g., the finding that pairing a CS, such as the name of a brand, with a valenced US, such as a picture of puppies, leads the CS to be liked more; for a review see Hofmann et al., 2010). A similar debate is taking place within EC research as in OEC research with regard to the role that propositions and associations play in EC effects. Although the EC literature usually makes reference to a propositional rather than an inferential account, the core idea is the same: unlike associations, propositions have a truth value, can encode information about the specific way in which stimuli are related, and can be used as premises in inferential reasoning (De Houwer, 2009, 2018; Mitchell et al., 2009).

Because propositions can encode the specific relation between events, one of the main approaches for testing the involvement of propositional processes in EC has been to study the impact of information about the precise nature of the CS-US relation. The rationale here is that if EC depends on propositions about the CS-US relation, then EC effects should be

---

[1] The study of Bayliss et al. (2007) can be situated in a wider literature on mere gaze effects which shows that participants evaluate objects that are looked at by others more positively than objects that are looked away from, even in the absence of emotional expressions (e.g., Bayliss et al., 2006; Corneille et al., 2009). Many studies in this literature suggest that inferences play a role: gaze effects were eliminated when observers believed the model could not see the stimulus (Manera et al., 2014) or were unaware of the contingencies (Bry et al., 2011); effects were reduced or even reversed when the model was considered untrustworthy (King et al., 2011; Treinen et al., 2012); and effects were amplified when multiple models were used (Capozzi et al., 2015). Although we consider this research to be closely related to OEC, proponents of an associative account of observational conditioning might argue that these gaze effects fall beyond the scope of such an account, because there would not seem to be a clear US (i.e., a model's gaze is not inherently valenced but can be construed as either positive or negative depending on its relation to the location of the stimulus; see also Bry et al., 2011). If so, these findings would not be considered relevant to the debate between inferential and associative accounts of OEC.

moderated by such relational information. Many recent studies have shown that this is the case: when relational qualifiers or the broader context signal that the CS and US are opposite to one another, EC effects are – under certain circumstances – reversed (i.e., the CS acquires a valence opposite to the valence of the US; e.g., Fiedler & Unkelbach, 2011; Förderer & Unkelbach, 2012; Moran et al., 2017). EC effects have also been shown to vary depending on whether CSs are thought to cause, predict, or be unrelated to USs (Hughes, Ye, Van Dessel, et al., 2019). However, the impact of relational information is not always straightforward. This is especially true for automatic evaluations (i.e., evaluations measured under conditions that are assumed to be suboptimal for cognitive processing, such as when there is little time or people are engaged in multiple tasks; see Moors & De Houwer, 2006). Such evaluations are often merely attenuated rather than reversed by oppositional information (i.e., the impact of CS-US pairings is reduced; e.g., Moran & Bar-Anan, 2018; Peters & Gawronski, 2011; Zanon et al., 2014). Although this complicates the conclusions that can be drawn from this body of research, the finding that relational information moderates EC has induced some researchers to assign a large (or even exclusive) role to propositional processes (for a review, see De Houwer et al., 2020).

Similar to what we discussed for OEC, a common finding that was initially viewed as evidence against a propositional account of EC was the demonstration of EC effects in the apparent absence of contingency awareness (for a review on the role of contingency awareness in EC see Sweldens et al., 2014). However, much of this evidence has been heavily criticized on multiple grounds, with a recent review concluding that there seems to be little evidence for EC in the absence of contingency awareness (Corneille & Stahl, 2019).

Based on these and other types of evidence, most contemporary accounts of EC assign an important role to propositional processes (for a recent overview of theoretical accounts, see Corneille & Stahl, 2019). Although it is of course possible that OEC is mediated by different processes than EC, the evidence for the involvement of propositional and inferential processes in EC strengthens the case for an inferential account of OEC.

The Current Research

As we discussed above, the debate about whether OEC is due to associative or inferential processes was not settled. In the literature on observational (fear) conditioning, many researchers still assume that observational conditioning effects are mediated by association formation (e.g., Askew & Field, 2008; Heyes, 2012; Olsson & Phelps, 2007), despite the fact that in both social learning and EC research evidence has since been obtained that has strengthened the case for an inferential account of observational conditioning.

In the current research we set out to provide a new, direct test of an inferential account of OEC. Inspired by EC research, we manipulated relational information as a way to test whether inferential processes play a role in OEC. An inferential account of OEC would assume that the observer's inference about the evaluative properties of the CS is based on several premises, including a proposition about the relation between the CS and the model's evaluation of that CS. Unlike an associative account, this account predicts that information which affects this proposition (i.e., relational information) would also influence the resulting inference and thus the OEC effect. Therefore, we examined whether additional information about the exact nature of the relationship between the CS and the model's reaction (US) influenced the strength and direction of OEC effects. This was expected to (a) provide information about a potentially important moderator of OEC effects and (b) inform theorizing about the mental processes driving observational (evaluative) conditioning.[2]

---

[2] One might ask why it is necessary to test whether OEC is sensitive to relational information, given that we already know that EC is (and OEC can be considered a subtype of EC if the model's reaction serves as a US). We believe this is relevant for two reasons. First, although it might seem reasonable to assume that OEC is moderated by the same factors as EC, such assumptions need to be empirically investigated before we can hold them with certainty. Second, while accounts of EC have increasingly incorporated the idea that propositions and inferences play a role, associative accounts remain influential in the observational conditioning literature. Hence, a direct test of this prediction using a typical observational conditioning paradigm has theoretical value.

In our studies participants watched videos wherein an individual (the model) tasted two different types of cookies, reacting positively to one cookie (CS_{pos}) and negatively to the other (CS_{neg}). Critically, we manipulated the relationship between the cookies (CSs) and the model's reactions (USs) by providing relational information prior to the observation phase. In Experiments 1-2, half of the participants were told that the model would display his honest opinion of the cookies, whereas the other half were told that he would *fake* his reactions to the cookies. In Experiments 3-4b, a third group was told that the model would show the *opposite* reaction to what he actually felt.

Following the observation phase, evaluative responses to the cookies were measured. In line with EC research, we included not only self-reported ratings but also a measure of automatic evaluations, in order to try to obtain convergent evidence for our predictions across multiple measures (i.e., to avoid basing our conclusions only on self-reports). We used a variant of the Implicit Association Test (IAT; Greenwald et al., 1998) to measure automatic evaluations. In a typical IAT, participants are asked to categorize positive and negative stimuli based on their valence on some trials, whereas they have to categorize the target stimuli based on a different feature (e.g., which brand they belong to) on other trials. Since participants have to use the same set of response keys on both trial types, the speed with which they can categorize a target stimulus with the same response key as positive (vs. negative) stimuli is taken as an index of how positively they evaluate the target stimulus. Given the nature of the task, evaluations measured within the IAT are usually considered to be more automatic (i.e., measured under conditions that are suboptimal for cognitive processing) than self-reported stimulus evaluations. In the current research, we opted for the personalised version of the IAT (pIAT), which requires participants to sort stimuli with the same keys as liked and disliked stimuli (Olson & Fazio, 2004), because responses to a standard IAT (which requires participants to sort stimuli with the same keys as normatively positive or negative stimuli) might simply have reflected knowledge about the *model's* preferences, whereas we were interested in the observer's *own* preferences.

All experiments were conducted in accordance with the General Ethical Protocol of the Ethical Committee of the Faculty of Psychology and Educational Sciences at Ghent University. Stimulus materials, scripts, raw and processed data, and all R code used for analyses are available on the Open Science Framework (https://osf.io/9rta3/). Designs and analysis plans were pre-registered for Experiments 1, 3, 4a, and 4b (https://osf.io/s4n69, https://osf.io/26u3v, https://osf.io/863rq, https://osf.io/k2w94). Experiment 2 was not pre-registered due to an oversight; however, all relevant documents were uploaded prior to data collection (https://osf.io/y5g7x/). Any deviations from these pre-registrations are listed in the "Deviations from pre-registration" document on the OSF page (https://osf.io/tpdwf/).

Experiment 1

In Experiment 1 we examined if our observational conditioning procedure would lead to OEC effects and whether these effects would be influenced by relational information. Our first hypothesis was that after observing a model react positively to one CS and negatively to another, observers would evaluate the former more positively than the latter, both on self-report and pIAT measures. Our second hypothesis was that relational information would moderate OEC effects, such that participants who were told that the model expressed his honest opinion would show the above effects whereas those told that the model had faked his reactions would not.

Method

Participants and Design

Participants were recruited via the online platform Prolific Academic (https://www.prolific.co/) and completed the experiment in exchange for €1.40. Participants who had incomplete data or who encountered technical issues (*n* = 39) were excluded and replaced during data collection, resulting in a sample of 165 participants (94 women, M_{age} = 31.7, SD_{age} = 7.8, age range: 18-50 years). A 2 (*Stimulus*: CS paired with positive vs. negative reaction) x 2 (*Relational Information*: genuine vs. faked reaction) design was employed, with the first factor manipulated within and the second manipulated between participants. Stimulus identity (CS1 vs. CS2 paired with the positive reaction), evaluative measure order (self-reports vs. pIAT first), and pIAT block order (learning-consistent vs. learning-inconsistent block first) were counterbalanced across participants.[3]

Stimuli

CSs and USs. Two differently shaped cookies (a circle and a triangle) with fictional names ("Vekte" and "Empeya") served as CS1 and CS2. We filmed multiple videos of three different models who were instructed to eat a cookie and display positive or negative nonverbal reactions (USs). Each video (10 seconds long) showed the model taking a cookie from a plate, taking a bite, and displaying a reaction for approximately five seconds. The cookie shapes were clearly visible and the corresponding name label was placed next to the plate. For each model, we selected two videos per category (CS1-positive, CS1-negative, CS2-positive, CS2-negative) and asked an independent sample of participants to rate the model's reactions in terms of believability and valence. We selected one video per category from the model with the highest believability ratings (a 23-year-old male model) and ensured that the reactions in the CS1 vs. CS2 videos did not differ significantly in terms of valence.[4]

[3] As the data were collected online, server issues made it difficult to achieve perfect counterbalancing (e.g., arriving at exactly 10 participants in each of the 16 cells of Experiment 1). However, in all of our experiments counterbalancing was close to complete, with numbers per cell never deviating more than one participant from the planned cell size (in Experiment 1, for example, numbers ranged from 9 to 11 participants per cell).

[4] Pretest materials, data, and analyses are available at https://osf.io/4vbxz/. One half of the videos was rated by one sample of participants and the other half by a second so that they never saw one model reacting in different ways to the same cookie. Believability ratings differed from 0 (*neutrality*) for all four videos (all *p*s < .001). The two positive videos did not differ significantly in terms of reaction valence, $t(95.81) = -1.38$, $p = .17$, and neither did the two negative videos, $t(89.34) = 1.35$, $p = .18$.

pIAT. The CS names served as target labels, while the categories “I like” and “I dislike” served as attribute labels. Target stimuli consisted of four edited pictures of each CS (upright vs. vertically flipped, coloured vs. grayscale) and the CS names in two different fonts, resulting in six target stimuli per CS category. Attribute stimuli consisted of twelve positive and negative words (*pleasure, holidays, rainbows, gifts, peace, friends, sickness, accidents, abuse, death, fear,* and *pain*) presented in a regular font (Arial, 5.5%).

Procedure

The experiment was programmed in Inquisit 4.0 and hosted via Inquisit Web (Millisecond Software, Seattle, WA). After providing informed consent and demographic information, participants read a cover story stating that two companies were each developing a new type of cookie. They were shown pictures of the CSs and their corresponding names and asked to remember these throughout the experiment. Thereafter they read the relational information, watched the videos, and completed the evaluative measures, followed by exploratory questions.

Relational Information. Prior to the OEC phase, participants were told that they would watch videos of a participant eating the two cookies. Those in the ‘genuine reaction’ condition were then informed that “before eating the cookies this participant was told to (visually) show whether he genuinely liked or disliked the cookies”. Those in the ‘faked reaction’ condition were informed that “before eating the cookies this participant was told to (visually) fake that he liked one cookie and disliked the other cookie”. Subsequently, participants completed a check to see if they remembered the cookie names and the relational information. Incorrect responses led to re-exposure to the names and the information, followed again by the manipulation check until it was successfully completed.

OEC Procedure. Participants watched two different videos. In one video the model tasted one cookie (CS_{pos}) and reacted positively by showing a facial expression of enjoyment and then eating the entire cookie. In the other video the model tasted a second cookie (CS_{neg}) and showed disgust via his facial expression and body language. Both videos were presented three times each in a random order, with an inter-trial-interval (ITI) of three seconds.

Evaluative Ratings. Participants were asked to provide ratings of each CS using a scale from -10 to +10 with 0 as a neutral point. Four different questions were asked for each CS using the following anchors: *very bad – very good, very negative – very positive, I would dislike it very much – I would like it very much,* and *very unpleasant – very pleasant*. The eight questions were presented in a random order.

pIAT. Prior to the pIAT, participants were again asked to report the CS names and reminded of the names if necessary. They were then told that they had to categorise stimuli as quickly and accurately as possible.

On each pIAT trial, a stimulus was presented in the middle of the screen and had to be classified according to two labels presented on the top left and right of the screen using the D and K keys. Error feedback was provided in the form of a red ‘X’ presented for 200 ms before the trial ended (ITI: 400 ms).

The pIAT consisted of seven blocks. In Block 1 (practice block; 16 trials) the category labels were the two CS names, and participants had to sort pictures and names of the CSs. In Block 2 (practice block; 16 trials) they had to classify valenced words in terms of whether they belonged to the category of things they liked or to the category of things they disliked (note that no error feedback was presented for this trial type in any of the blocks). In Blocks 3-4 (test blocks; 32 trials each) the two trial types were combined, requiring participants to sort CSs into the two CS categories as well as valenced words in terms of whether they belonged to the “I like” or the “I dislike” category. In Block 5 (practice block; 16 trials) participants again had to categorise only the CSs, but the response mapping was reversed relative to the previous blocks (i.e., the CS categories switched location). Finally, in Blocks 6-7 (test blocks; 32 trials each) participants once again encountered both trial types, with the same response mapping for like-dislike trials but the switched response mapping for CS trials. Trial order within each block was random and the relevant labels remained on screen throughout each block. Because pIAT block order was counterbalanced, for half of the participants the initial response mappings were consistent with the OEC phase (i.e., sorting the CS_{pos} with the same key as things they liked and the CS_{neg} with the same key as things they disliked), whereas for the other half the initial response mappings were inconsistent with the OEC phase (i.e., sorting the CS_{pos} with the same key as things they disliked and the CS_{neg} with the same key as things they liked).

Exploratory Questions. Finally, memory for the pairings, believability of the videos and information, hypothesis awareness, demand compliance, and reactance were assessed. These questions were included for exploratory purposes and are not discussed further unless otherwise stated (see Supplementary Materials).

Results

Data Preparation

The evaluative ratings were averaged to create two mean scores (one for the CS_{pos} and another for the CS_{neg}). A difference score was then created by subtracting the CS_{neg} rating from the CS_{pos} rating, so that positive scores indicated a preference for the CS_{pos} over the CS_{neg} whereas negative scores indicated the opposite pattern. Reaction times on the pIAT were used to calculate participant-level scores according to the D1-algorithm (Greenwald et al., 2003). Positive pIAT scores reflected a more positive evaluation of the CS_{pos} relative to the CS_{neg}, negative scores reflected the opposite.

Data for participants who had error rates above 30% across the pIAT (n = 3) or above 40% for any test block (n = 6), or who responded faster than 400 ms on more than 10% of trials (n = 1) were removed. This resulted in a final sample of 155 participants (89 women, M_{age} = 31.6, SD_{age} = 7.7).

Analytic Strategy

All hypothesis tests were conducted at the $\alpha = .05$ significance level. For both dependent variables (ratings and pIAT scores), one-sample *t*-tests were used to investigate whether the scores differed from zero (i.e., if one CS was evaluated more positively than the other). Two-sample *t*-tests were then used to determine if scores differed as a function of the relational information received by participants (genuine vs. faked reaction). We supplemented these significance tests with Bayesian analyses. All reported Bayes Factors (BF) indicate the probability of the alternative hypothesis compared to the null hypothesis given the observed data (Rouder et al., 2009). We also checked whether any of the counterbalanced method factors improved model fit using the Akaike information criterion (AIC). If they did, we conducted an analysis of variance (ANOVA) to test for the effect of relational information in the presence of those method factors as well as for the effects of the factors themselves (the results are reported in the Supplementary Materials and will only be discussed here if relevant to the main findings).

Hypothesis Testing

Self-Reported Evaluations. Table 1 shows the means and standard deviations for both dependent variables in each condition. A self-reported OEC effect emerged: the mean difference between CS_{pos} and CS_{neg} ratings (M = 6.12, SD = 7.24) was significantly larger than zero, $t(154) = 10.52$, $p < .001$, Cohen's $d = 0.85$, 95% CI [0.66, 1.03], $BF_{10} > 10\ 000$. This OEC effect was moderated by the relational information: the effect was larger in the genuine reaction condition ($M = 8.82$, $SD = 7.36$) than in the faked reaction condition ($M = 3.59$, $SD = 6.17$), $t(144.8) = 4.77$, $p < .001$, $d = 0.77$, [0.43, 1.11], $BF_{10} = 8384.62$. Interestingly, OEC effects emerged for those who were told that the model's reactions were genuine, $t(74) = 10.38$, $p < .001$, $d = 1.20$, [0.90, 1.49], $BF_{10} > 10\ 000$, but also for those who were told that the model's reactions were faked, $t(79) = 5.20$, $p < .001$, $d = 0.58$, [0.34, 0.82], $BF_{10} > 10\ 000$.

Automatic Evaluations (pIAT). An overall automatic OEC effect emerged in the sense that the mean pIAT score ($M = 0.20$, $SD = 0.40$) was positive, indicating a relative preference for the CS_{pos} over the CS_{neg}, $t(154) = 6.05$, $p < .001$, $d = 0.49$, [0.32, 0.65], $BF_{10} > 10\ 000$. This effect was also moderated by the type of relational information: pIAT scores were larger when the model's reactions were said to be genuine ($M = 0.26$, $SD = 0.36$) compared to when they were said to be faked ($M = 0.13$, $SD = 0.43$), $t(150.64) = 2.02$, $p = .02$, $d = 0.32$, [0.002, 0.640], $BF_{10} = 2.11$. Once again, OEC effects were evident in both the genuine reaction condition, $t(74) = 6.34$, $p < .001$, $d = 0.73$, [0.47, 0.98], $BF_{10} > 10\ 000$, and the faked reaction condition, $t(79) = 2.75$, $p = .004$, $d = 0.31$, [0.08, 0.53], $BF_{10} = 8.20$.

Table 1

Mean Differential Ratings and pIAT Scores per Condition (Experiment 1)

Variable	Relational information			
	Genuine reaction		Faked reaction	
Ratings (difference between CSs)	8.82	(7.36)	3.59	(6.17)
pIAT scores	0.26	(0.36)	0.13	(0.43)

Note. Values between parentheses indicate standard deviations.

Discussion

Experiment 1 indicated that our OEC procedure resulted in significant effects: after watching a model react positively when eating one cookie and negatively when eating another, participants preferred the former over the latter on both self-report and pIAT measures. These effects were moderated by relational information: larger effects emerged when the model was said to have expressed a genuine, relative to a faked, reaction.

Nevertheless, two points are worth noting. First, and contrary to predictions, participants in the faked reaction group still showed an OEC effect. Second, the impact of relational information on pIAT scores was only significant when a specific scoring algorithm (D1) was used (and not when a D4 score was used; see Supplementary Materials for more information), and even then, the Bayes Factor indicated only weak evidence. One possibility is that this weak impact of relational information on pIAT scores was due to the brief nature of the information provided (i.e., a single sentence before the observation phase). We therefore decided to conduct a second experiment to replicate and strengthen our initial findings while also using more elaborate relational information.

Experiment 2

Method

Participants and Design

After replacing participants with incomplete data (n = 15), our sample consisted of 162 participants (86 women, M_{age} = 31.7, SD_{age} = 8.9, range: 15-53 years) recruited via Prolific Academic in exchange for €1.40. This sample size provided sufficient power (.93) to observe a medium-sized difference in pIAT scores between conditions (d = 0.50). The design was identical to Experiment 1.

Stimuli

The same stimuli were used as in Experiment 1.

Procedure

The procedure was similar to Experiment 1, with several exceptions. First, we revised the relational information to make it more elaborate and salient. Those in the genuine reaction condition were now told that the videos were taped during a consumer test in which the model had to show his honest reactions, and that he had been asked to clearly display whether he liked or disliked the cookies in order to capture his first impressions. Those in the faked reaction condition were told that the videos were taped during the casting for an advertisement, that the person was paid by one of the companies to participate, and that in order to judge his acting skills the company had asked him to fake that he liked their cookie and disliked the other.

Second, in order to make the relational information more salient we first provided the information about the cookie names and checked whether participants could remember them. Only then did we provide the relational information and ask participants to complete a manipulation check about this information.

Third, we checked if participants could still remember the relational information at the end of the study and whether they took this information into account when forming their CS evaluations. Additionally, participants in the faked reaction group were asked if they liked one cookie more than the other. If they replied “Yes”, they were asked (in an open-ended format) to report why, given that the model faked his reactions. If they replied “No”, we asked them to report why not.

Results

Data Preparation and Analytic Strategy

Data were prepared and analysed as in Experiment 1. Participants were excluded if they produced error rates above 30% across the pIAT (*n* = 7) or above 40% for any test block (*n* = 5), or if they responded faster than 400 ms on more than 10% of trials (*n* = 2). The final sample consisted of 148 participants (81 women, M_{age} = 31.5, SD_{age} = 8.9).

Hypothesis Testing

Self-Reported Evaluations. Table 2 shows the means and standard deviations in each condition. Overall a self-reported OEC effect emerged: participants preferred the CS_{pos} over the CS_{neg}, (*M* = 7.52, *SD* = 6.95), *t*(147) = 13.18, *p* < .001, *d* = 1.08, [0.88, 1.29], BF_{10} > 10 000. This effect was moderated by relational information, such that it was larger when the model's reactions were said to be genuine (*M* = 9.56, *SD* = 6.39) relative to when they were said to be faked (*M* = 5.54, *SD* = 6.93), *t*(145.57) = 3.67, *p* < .001, *d* = 0.60, [0.26, 0.94], BF_{10} = 142.13. Once again, OEC effects were significant in both the genuine, *t*(72) = 12.79, *p* < .001, *d* = 1.50, [1.16, 1.83], BF_{10} > 10 000, and the faked reaction conditions, *t*(74) = 6.92, *p* < .001, *d* = 0.80, [0.54, 1.06], BF_{10} > 10 000. Finally, there was an interaction between relational information and task order: relational information moderated the OEC effect when participants provided their ratings after completing the pIAT but not when they provided their ratings before completing the pIAT (see Supplementary Materials).

Automatic Evaluations (pIAT). Although an overall OEC effect again emerged, such that the CS_{pos} was preferred over the CS_{neg} (*M* = 0.23, *SD* = 0.41), *t*(147) = 6.84, *p* < .001, *d* = 0.56, [0.39, 0.73], BF_{10} > 10 000, pIAT scores were not found to differ as a function of relational information, *t*(145.42) = 0.71, *p* = 0.24, *d* = 0.12, [-0.21, 0.44], BF_{10} = 0.34. OEC effects emerged regardless of whether the model's reactions were described as genuine, *M* =

0.26, *SD* = 0.42, *t*(72) = 5.21, *p* < .001, *d* = 0.61, [0.36, 0.86], BF_{10} > 10 000, or as faked, *M* = 0.21, *SD* = 0.40, *t*(74) = 4.44, *p* < .001, *d* = 0.51, [0.27, 0.75], BF_{10} = 1188.17.

Table 2

Mean Differential Ratings and pIAT Scores per Condition (Experiment 2)

Variable	Relational information			
	Genuine reaction		Faked reaction	
Ratings (difference between CSs)	9.56	(6.39)	5.54	(6.93)
pIAT scores	0.26	(0.42)	0.21	(0.40)

Note. Values between parentheses indicate standard deviations.

Discussion

Experiment 2 sought to replicate and strengthen the findings of Experiment 1. On the one hand, self-reports were once again moderated by relational information, with larger OEC effects in the genuine relative to the faked reaction condition. On the other hand, and unlike in Experiment 1, automatic OEC effects were not moderated by relational information: similar pIAT effects were found in the genuine and faked reaction conditions.

Reflecting on these findings, one may ask: why did relational information moderate automatic evaluative responding in Experiment 1 but not in Experiment 2? Although it is possible that – contrary to our intentions – the relational information was less salient to participants in Experiment 2, it should be noted that the evidence for an impact of relational information on pIAT scores was also rather unconvincing in Experiment 1. So far, the overall trend of evidence supporting the idea that relational information moderates automatic OEC effects is therefore weak.

One possible explanation for this outcome is that the stimuli used in Experiments 1-2 may have been suboptimal. Reading through participants' responses to the exploratory questions (see Supplementary Materials) revealed two issues. First, many participants referred to the shapes of the cookies when asked why they preferred one CS over the other, with most considering the round shape to be more familiar than the triangular shape. Second, many reported difficulties remembering which cookie names and shapes belonged together. Because the pIAT contained pictures of the cookies (without their names printed underneath) participants were required to mentally retrieve the corresponding cookie name on some pIAT trials. This is an extra step that was irrelevant to our research question and may have introduced noise to reaction times. Experiment 3 sought to eliminate both methodological issues by keeping the shape of the cookies constant and only varying their names.

Another question is why the faked reaction information merely reduced OEC effects and did not eliminate them. Once again, exploring our data in greater detail proved informative. The distribution of effects in Experiments 1-2 suggested that the impact of relational information on self-reported evaluations was not due to an overall shift in participant-level OEC effects (see Figures 1-2 for the distributions of the variables in the different experiments). Instead, it seemed to be mainly due to the complete absence of an OEC effect in a small subgroup (~ 15 participants) of the faked reaction group. This suggests that only a small number of participants were strongly influenced by the information that the model faked his reactions. For this small group, the information might have implied that the model's reactions were not a valid source for inferring the valence of the CSs, leading them to evaluate both CSs in the same way. For others, however, this same information may have created an informational 'vacuum': it implies that the model's reactions may not be a valid source for inferring the valence of the CSs, but it does *not* imply anything about the model's genuine evaluations of the CSs. Thus, in the absence of any other information as to the properties of the CSs, many participants may decide to rely on the model's reactions anyway.

In that case, we would expect to find a clearer impact on OEC effects if relational information has unambiguous implications for how the valence of the CSs should be inferred from the model's reactions. Experiment 3 therefore included a third, "opposite reaction" manipulation, which involved telling participants that the model showed the opposite of what he felt. We predicted that in this group CSs would be evaluated in a way opposite to the valence of the reactions they were paired with (i.e., a reversed OEC effect).

Experiment 3

Method

Participants and Design

After replacing participants with incomplete data (*n* = 22), our sample consisted of 213 participants (90 women, M_{age} = 28.68, SD_{age} = 7.7, range: 18-50 years) recruited via Prolific Academic in exchange for €1.80. The sample size was based on a power calculation indicating we needed a minimum sample of *n* = 206 in order to have .90 power to detect a medium-sized main effect of relational information (η^2_p = 0.059).

A 2 (*Stimulus*: CS paired with positive vs. negative reaction) x 3 (*Relational Information*: genuine vs. faked vs. opposite reaction) design was used, with the first factor manipulated within and the second manipulated between subjects. We counterbalanced the same method factors as in Experiments 1-2.

Stimuli

Only circle-shaped cookies were used, so that the CSs differed only in terms of their names ("Empeya" vs. "Plogo") and not their shapes.[5] This meant that only the two videos containing circle-shaped cookies were used. To counterbalance stimulus identity, we used

[5] We also replaced the name "Vekte" by "Plogo" based on other studies from our lab suggesting that overall tendencies to prefer one nonword over the other emerge less frequently when "Plogo" is compared to "Empeya".

photo- and video-editing software to edit the name labels in these two videos and create matching sets. Therefore, unlike in Experiments 1-2, all participants observed the same positive and negative reactions, and only the cookie names on the labels in these respective videos were varied as a function of stimulus identity. The pIAT stimuli were changed accordingly: target stimuli now consisted of the cookie names (rather than both names and pictures). Each name was presented in multiple combinations of rotations and fonts (six stimuli per CS) in order to prevent participants from categorizing the CSs based on purely perceptual features (see also De Houwer & Vandorpe, 2010; Zanon et al., 2014).

Procedure

A similar procedure was used as in Experiments 1-2, with two notable changes. First, to account for the fact that the cookies looked identical, we told participants that they were produced by the same company but based on different recipes. Second, while the genuine and faked reaction groups were given information similar to the information given in Experiment 2, a third group received opposite reaction information. Specifically, they were told that the videos were taped during the casting for a cookie advertisement and that in order to judge an actor's skills, the company had asked him to show the *opposite* reaction to what he felt about each cookie.

Results

Data Preparation and Analytic Strategy

Data were prepared and excluded as in Experiments 1-2. We excluded data from participants who had error rates above 30% across the pIAT (n = 1), above 40% for any test block (n = 17), or who responded faster than 400 ms on more than 10% of trials (n = 23). The final sample consisted of 173 participants (74 women, M_{age} = 29.3, SD_{age} = 7.7).

The analytic strategy was updated because we now had three conditions. We first conducted a one-way ANOVA to investigate whether ratings and pIAT scores differed as a function of relational information (genuine vs. faked vs. opposite reaction). We also used pairwise *t*-tests (with Holm-Bonferroni correction of the *p*-values for multiple comparisons) to investigate which conditions (if any) differed from each other.

Hypothesis Testing

Self-Reported Evaluations. Table 3 shows the means and standard deviations in the three conditions. The self-reported OEC effect was moderated by relational information type, $F(2,170) = 50.10$, $p < .001$, $\eta^2_p = 0.37$, 90% CI [0.27, 0.45], $BF_{10} > 10\,000$. All conditions differed from each other (genuine-faked: $p = .008$, genuine-opposite: $p < .001$, faked-opposite: $p < .001$). As expected, scores in the genuine reaction group ($M = 10.67$, $SD = 6.68$) indicated a strong preference for the CS_{pos} over the CS_{neg}, $t(59) = 12.38$, $p < .001$, $d = 1.60$, [1.21, 1.98], $BF_{10} > 10\,000$. Although smaller, scores in the faked reaction group ($M = 6.54$, $SD = 7.58$) also indicated a preference for the CS_{pos} over the CS_{neg}, $t(54) = 6.39$, $p < .001$, $d = 0.86$, [0.55, 1.17], $BF_{10} > 10\,000$. Critically, a reversed pattern emerged in the opposite reaction group ($M = -4.02$, $SD = 9.97$), with participants evaluating the CS_{neg} more positively than the CS_{pos}, $t(57) = -3.07$, $p = .002$, $d = 0.40$, [0.13, 0.67], $BF_{10} = 18.73$. In absolute terms, this reversed OEC effect was smaller than the standard effect in the genuine reaction group, $t(99.1) = 4.25$, $p < .001$, $d = 0.77$, [0.40, 1.17], $BF_{10} = 992.61$, but not smaller than the effect in the faked reaction group, $t(106.1) = 1.52$, $p = .13$, $d = 0.28$, [-0.09, 0.65], $BF_{10} = 1.02$.

Automatic Evaluations (pIAT). Relational information type also moderated pIAT scores, $F(2,170) = 6.72$, $p = .002$, $\eta^2_p = 0.07$, [0.02, 0.14], $BF_{10} = 18.09$. In addition, there was a significant interaction between relational information type and evaluative measure order, such that the relational information only moderated pIAT performance if the task was completed after the ratings (see Supplementary Materials). Follow-up comparisons indicated that the opposite reaction condition differed from the genuine (p = .01) and faked reaction (p = .002) conditions, while the latter two conditions did not differ from one another (p = .54). Similar to Experiments 1-2, pIAT scores indicated a clear preference for the CS_{pos} over the CS_{neg} when the model's reactions were said to be genuine (M = 0.18, SD = 0.36), t(59) = 3.91, p < .001, d = 0.50, [0.23, 0.77], BF_{10} = 193.94, as well as when they were said to be faked (M = 0.23, SD = 0.40), t(54) = 4.18, p < .001, d = 0.56, [0.28, 0.85], BF_{10} = 210.64. However, pIAT scores did not differ from zero in the opposite reaction condition (M = -0.02, SD = 0.40), t(57) = -0.42, p = .34, d = 0.05, [-0.20, 0.31], BF_{10} = 0.20.

Table 3

Mean Differential Ratings and pIAT Scores per Condition (Experiment 3)

Variable	Relational information					
	Genuine reaction		Faked reaction		Opposite reaction	
Ratings (difference between CSs)	10.67	(6.68)	6.54	(7.58)	-4.02	(9.97)
pIAT scores	0.18	(0.36)	0.23	(0.40)	-0.02	(0.40)

Note. Values between parentheses indicate standard deviations.

Discussion

In line with predictions, the OEC effect indexed by self-reported ratings was moderated by relational information, with a large standard effect in the genuine reaction condition, a smaller standard effect in the faked reaction condition, and a reversed effect in the opposite reaction condition. Relational information also moderated pIAT scores when the pIAT was completed after the ratings: although the effect again did not differ between the genuine and faked reactions conditions, it was attenuated (but not reversed) in the opposite reaction condition.

Experiments 4a-4b

Experiment 4a sought to replicate Experiment 3 and to address the potential role of demand compliance. Given that the relational information in Experiments 1-3 was salient, and both the ratings and the pIAT were clearly concerned with evaluations, many participants may have inferred that the researchers wanted them to evaluate the stimuli in line with the relational information. Therefore, it is possible that the effects obtained thus far were the result of participants complying with this perceived researcher demand rather than reporting how they actually felt about the CSs.

On the one hand, it seems unlikely that the effects obtained in Experiments 1-3 were the simple product of demand compliance, given that these effects were still present when demand compliant participants were excluded (see Supplementary Materials). On the other hand, replicating the findings of Experiment 3 under conditions that are less likely to evoke demand compliance would provide even stronger evidence for the above claim.

We therefore carried out an experiment using a modified procedure designed to draw attention away from the relational information as well as from the fact that our main interest was in the evaluation of the cookies. In Experiment 4a, participants were told that they were taking part in a pilot study with the aim of selecting videos for future research. The relational information was mentioned only briefly and was no longer followed by a manipulation check that emphasised the importance of that information. The self-reported evaluative ratings were also now buried in a long list of otherwise irrelevant distractor questions about the videos.

An additional experiment (Experiment 4b) was conducted to explore the possibility that the high rates of demand awareness observed in Experiment 4a (see below) were at least partially due to participants having completed the pIAT (which clearly focused on their evaluations of the cookies) prior to answering the demand awareness question. Experiment

4b was therefore identical to Experiment 4a, with the exception that participants only completed the self-report ratings (i.e., there was no pIAT). Because Experiment 4b was conducted solely to explore whether demand awareness would remain high in the absence of the pIAT, we report Experiment 4a in detail below and mention only the noteworthy points of Experiment 4b (see Supplementary Materials for full methods and results of Experiment 4b).

Method

Participants and Design

After replacing participants with incomplete data ($n = 30$), our sample consisted of 239 participants (86 women, $M_{age} = 25.1$, $SD_{age} = 6.3$, range: 18-50 years) recruited via Prolific Academic in exchange for €1.80. The design was identical to Experiment 3, with the exception that evaluative measure order was not counterbalanced (i.e., participants first provided ratings and subsequently completed the pIAT).

Stimuli

The same videos were used as in Experiment 3. Target stimuli in the pIAT again consisted of six versions of each CS name. Unlike in Experiment 3 they were not rotated, as this made it difficult to ensure that they were equally close to both response labels. Instead, they were presented in lower- or uppercase and in regular, bold, or italic font.

Procedure

Participants were informed that they would take part in a pilot study that would allow us to select videos for future research. Therefore, they would watch a series of videos and answer questions about those videos. Participants were also told that we had asked the person in the videos (a) "to clearly display whether he liked or disliked the cookies (in other words, to show his genuine reaction to each cookie)", (b) "to fake that he liked a cookie or disliked a cookie (in other words, we told the person which reaction he should show to each cookie)", or (c) “to show the opposite of how he actually felt about the cookies (in other words, that he should pretend to like cookies that he actually disliked and the other way around)”.

After they had watched the videos, participants were reminded that we needed their honest answers to optimise our future research. They then answered 17 questions, most of which were distractor questions (e.g., about the visual quality of the videos or the model’s perceived age). Interspersed within these items were four questions that assessed stimulus evaluations: participants were asked how much they thought they would like each cookie, and how pleasant or unpleasant they considered each cookie to be (on scales from -4 to +4 with 0 as a neutral point).

After completing the pIAT, participants were again asked some questions about the experiment itself, including a number of questions related to demand. They were asked to indicate what they believed the researchers had expected them to do (demand awareness). In addition, they rated to what extent their responses had been based on their true feelings (for the ratings) and on responding quickly and accurately (for the pIAT), on trying to go along with the researchers’ goals or hypothesis (demand compliance), and on trying to go against the researchers’ goals or hypothesis (reactance).

Results

Data Preparation and Analytic Strategy

We excluded data from participants who had error rates above 30% across the pIAT ($n = 2$), above 40% for any test block ($n = 12$). No participants responded faster than 300 ms on more than 10% of trials.[6] The final sample of Experiment 4a consisted of 225 participants (84

[6] A cut-off of 300 rather than 400 ms was pre-registered for Experiment 4a because the data of Experiment 3 suggested that this lower cut-off was more appropriate for the simplified version of the pIAT.

women, M_{age} = 25.2, SD_{age} = 6.3) (Experiment 4b: n = 211, 56 women, M_{age} = 25.4, SD_{age} = 6.2). The analytic strategy was identical to that of Experiment 3.

Hypothesis Testing

Self-Reported Evaluations. Table 4 shows the means and standard deviations per condition for Experiments 4a-4b. Self-reported OEC effects were moderated by relational information, $F(2,222) = 46.90$, $p < .001$, $\eta^2_p = 0.30$, [0.21, 0.37], $BF_{10} > 10\,000$. All three conditions differed from each other (genuine-faked: $p = .014$, genuine-opposite: $p < .001$, faked-opposite: $p < .001$). The pattern in Experiment 4b was slightly different, in the sense that the genuine and faked reaction groups did not differ significantly, $p = .11$.

Similar to before, scores in the genuine reaction group ($M = 4.09$, $SD = 2.96$) indicated a strong preference for the CS_{pos} over the CS_{neg}, $t(74) = 11.99$, $p < .001$, $d = 1.38$, [1.06, 1.70], $BF_{10} > 10\,000$. The effect in the faked reaction group ($M = 2.84$, $SD = 2.54$) also indicated a (smaller) preference for the CS_{pos} over the CS_{neg}, $t(77) = 9.88$, $p < .001$, $d = 1.12$, [0.83, 1.40], $BF_{10} > 10\,000$. Finally, scores in the opposite reaction group ($M = -0.74$, $SD = 3.82$) did not differ significantly from zero, $t(71) = -1.64$, $p = .053$, $d = 0.19$, [-0.04, 0.43], $BF_{10} = 0.87$. That is, we found no evidence for a standard nor for a reversed OEC effect in this group.

Automatic Evaluations (pIAT). pIAT scores were not significantly moderated by relational information type, $F(2,222) = 2.75$, $p = .066$, $\eta^2_p = 0.02$, [0.00, 0.06], $BF_{10} = 0.53$. When pIAT block order and stimulus identity were included in the model, the effect of relational information became significant but was still weak, $p = .041$, $BF_{10} = 0.84$ (see Supplementary Materials). Follow-up comparisons indicated that none of the conditions differed from each other (genuine-faked: $p = .51$, genuine-opposite: $p = .21$, faked-opposite: $p = .07$). pIAT scores indicated a preference for the CS_{pos} over the CS_{neg} when the model's reactions were said to be genuine ($M = 0.14$, $SD = 0.47$), $t(74) = 2.60$, $p = .006$, $d = 0.30$, [0.07, 0.53], BF_{10} = 5.73, as well as when they were said to be faked (*M* = 0.19, *SD* = 0.42), *t*(77) = 4.00, *p* < .001, *d* = 0.45, [0.22, 0.68], BF_{10} = 143.87. However, the OEC effect was eliminated in the opposite reaction condition (*M* = 0.02, *SD* = 0.45), *t*(71) = 0.42, *p* = .66, *d* = 0.05, [-0.18, 0.28], BF_{10} = 0.10.

Table 4

Mean Differential Ratings and pIAT Scores per Condition (Experiments 4a-4b)

Variable	Relational information					
	Genuine reaction		Faked reaction		Opposite reaction	
Ratings (Experiment 4a)	4.09	(2.96)	2.84	(2.54)	-0.74	(3.82)
Ratings (Experiment 4b)	3.90	(2.62)	3.01	(2.75)	-0.27	(4.25)
pIAT scores (Experiment 4a)	0.14	(0.47)	0.19	(0.42)	0.02	(0.45)

Note. Values between parentheses indicate standard deviations. As the pIAT was not included in Experiment 4b, pIAT scores are available only for Experiment 4a.

Demand Awareness and Compliance. Most participants reported they were *aware* of the researcher demand: 63% of the genuine reaction group, 73% of the faked reaction group, and 78% of the opposite reaction group indicated that they believed that “the researchers wanted me to *evaluate* the cookies while taking into account the *instructions* given to the person in the videos (i.e., to combine what I saw in the videos with the information I received about the person’s instructions)”. The results of Experiment 4b further suggested that these high rates of demand awareness were not reduced when participants had not completed the pIAT (which clearly related the cookies to evaluative categories). In fact, *more* participants were considered demand aware in Experiment 4b than in Experiment 4a, $\chi^2(1) = 5.34$, *p* = .02, and the majority (67%) of these participants reported that they had already identified the researcher demand before encountering the exploratory questions.

Next, we checked whether excluding participants who selected the midpoint or higher on the demand *compliance* questions influenced the results. The pattern of self-report findings outlined above did not change after excluding participants who reported they were demand compliant on self-reports (n = 52), with one exception: the effect in the opposite reaction group became significantly smaller than zero (p = .03), indicating a reversed OEC effect (similar to the opposite reaction group in Experiment 3).

Excluding participants who reported demand compliance for the pIAT (n = 94) resulted in the effect of relational information on pIAT scores becoming non-significant. Note that the number of participants who reported demand compliance for the pIAT was surprisingly large relative to previous experiments; a closer inspection of responses to this question suggested that participants did not interpret it as we had intended.[7]

Discussion

OEC effects emerged on self-report measures and were again moderated by relational information (although the reversal of those effects in the opposite reaction group was less evident than in Experiment 3). OEC effects also emerged on the pIAT but these effects did not vary reliably as a function of relational information. Finally, even though we tried to reduce the potential influence of demand awareness and compliance, most participants were nonetheless aware of the researcher demand. Importantly, however, we once again found that excluding demand compliant participants did not reduce the impact of relational information on self-reports.

[7] Specifically, “responding quickly and accurately” could also be interpreted as going along with the researchers’ perceived goal and thus prompt participants to report a high level of “demand compliance” for the pIAT if they responded quickly and accurately. In line with this idea, 78% of those who selected the highest value for the demand compliance scale *also* did so for the scale assessing the extent to which they had responded based only on speed and accuracy.

Analyses on Combined Data

Although there were a number of procedural differences between the five experiments, we decided to perform analyses on the combined data as this allowed us to (a) test the effects of relational information with increased power; (b) include demand compliance and contingency memory in the models in order to test whether they moderated the effects; and (c) investigate whether demand compliance differed across experiments (i.e., whether the changes to the procedure in Experiments 4a-4b successfully reduced demand compliance).

Data Preparation

The data from all five experiments were combined into one large dataset. In order to standardise the values for the dependent variables (because the rating scales and the pIAT varied across studies), the explicit difference scores and the pIAT scores were scaled for each experiment. In order to be able to compare demand compliance in Experiments 4a-4b (where it was assessed via a numerical scale) to Experiments 1-3 (where it was assessed via a categorical response), all participants were coded as “Not demand compliant” if they indicated “No” in Experiments 1-3 or scored below the midpoint of the scale in Experiments 4a-4b, and as “Possibly demand compliant” if they indicated “Yes” or “I don’t know” in Experiments 1-3 or indicated the midpoint or higher in Experiments 4a-4b.

Analytic Strategy

Two subsets of the data were used to test specific hypotheses. First, the data of the genuine and faked reaction groups from all five experiments were used to test whether evaluative ratings (Experiments 1-4b; n = 711) and pIAT scores (Experiments 1-4a; n = 571) differed as a function of whether the model’s reaction was said to be genuine or faked. For both variables, we conducted a first ANOVA testing the effects of relational information, task order, experiment, and the interactions of relational information with both task order and experiment (as well as pIAT block order and its interaction with relational information for the pIAT scores). We also calculated "inclusion" Bayes Factors for the terms in this model, which reflect the evidence in favour of including a specific term in the model across "matched" models (i.e., all models that did not include any interactions with the term of interest but did include the underlying main effects if the term of interest was itself an interaction term). A second ANOVA further included demand compliance, contingency memory, and their interactions with relational information (the results are reported in the Supplementary Materials and will only be discussed here if they affect the interpretation of the main results).

Second, the data of the genuine, faked, and opposite reaction groups were used to test whether evaluative ratings (Experiments 3-4b; $n = 609$) and pIAT scores (Experiments 3-4a; $n = 398$) differed as a function of the three types of relational information. ANOVAs with the same terms as described above were run for this subset.

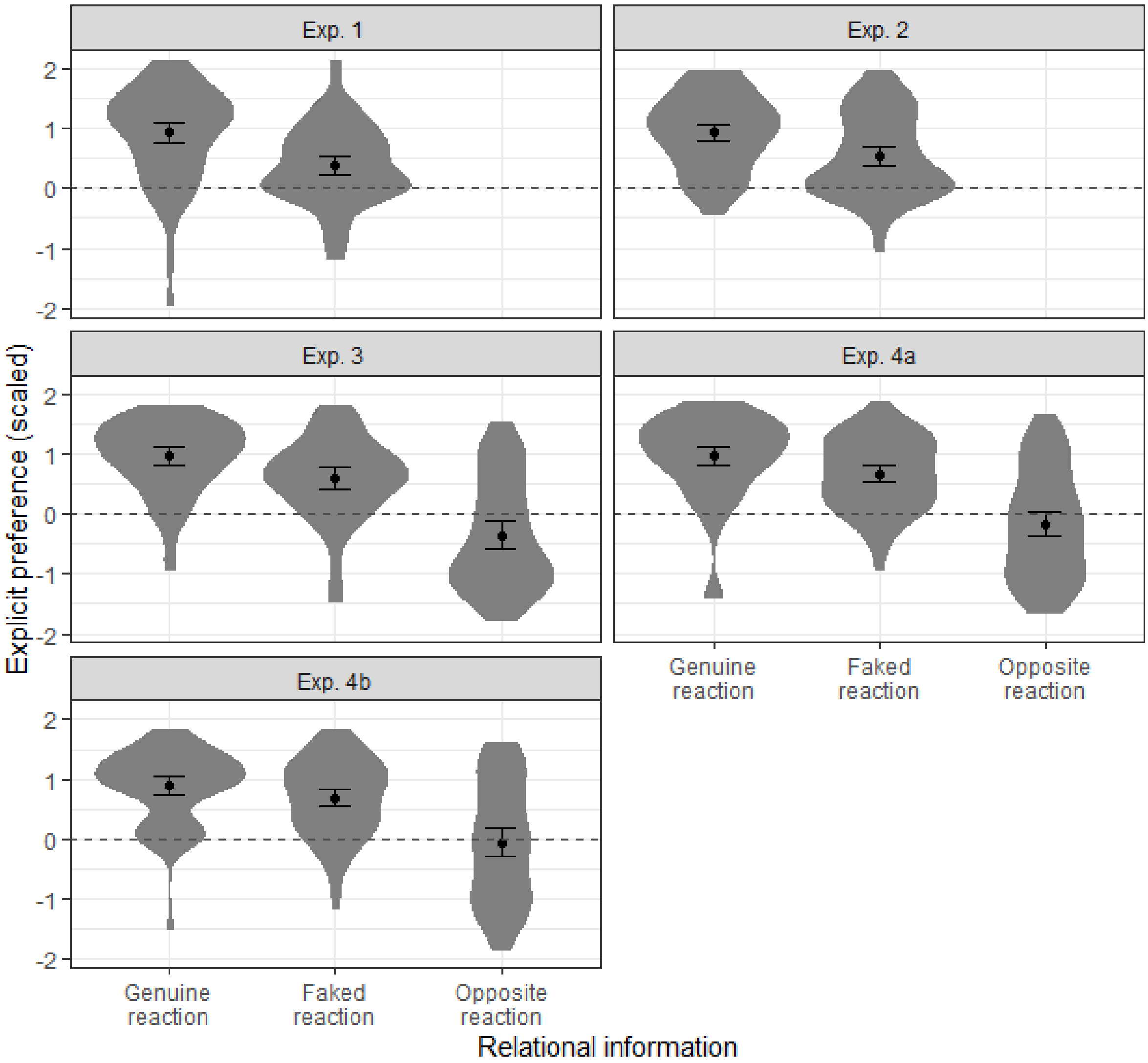
Self-Reported Evaluations

Figure 1 shows the means, confidence intervals, and distributions of the scaled explicit difference scores as a function of relational information for all five experiments. When the genuine and faked reaction groups (Experiments 1-4b) were compared, there was a main effect of relational information, $F(1,699) = 47.65$, $p < .001$, $\eta^2_p = 0.06$, [0.04, 0.09], $BF_{10} > 10\ 000$, such that the OEC effect was larger in the genuine reaction group than in the faked reaction group. In addition, there was a main effect of task order, $F(1,699) = 18.37$, $p < .001$, $\eta^2_p = 0.03$, [0.01, 0.05], $BF_{10} = 2124.97$, such that the OEC effect was larger when the ratings were completed first. No other effects were significant. Interestingly, both the overall OEC effect as well as the impact of relational information on the OEC effect emerged only for participants who had correct contingency memory. Finally, the effect of relational information was not qualified by whether participants reported demand compliance.

When all three groups (Experiments 3-4b) were included, there was again a main effect of relational information, $F(2,597) = 25.22$, $p < .001$, $\eta^2_p = 0.08$, [0.05, 0.11], $BF_{10} > 10\,000$. However, it interacted with task order, $F(2,597) = 8.79$, $p < .001$, $\eta^2_p = 0.03$, [0.01, 0.05], $BF_{10} = 1.31$, such that there was a large effect of relational information when the ratings were completed first, $F(2,511) = 124.34$, $p < .001$, and a smaller but still significant effect when the pIAT was completed first, $F(2,86) = 9.86$, $p < .001$. The effect of relational information was also qualified by experiment, $F(4,597) = 3.89$, $p = .004$, $\eta^2_p = 0.025$, [0.001, 0.030], $BF_{10} = 0.54$, such that the effect of relational information was more pronounced in Experiment 3 than in Experiments 4a-4b. No other effects were significant, although the larger model again indicated that only participants with correct contingency memory showed the expected effects, while the effect of relational information was not qualified by demand compliance.

Figure 1

Means and Distributions of Scaled Explicit Difference Scores in Experiments 1-4b



Automatic Evaluations (pIAT)

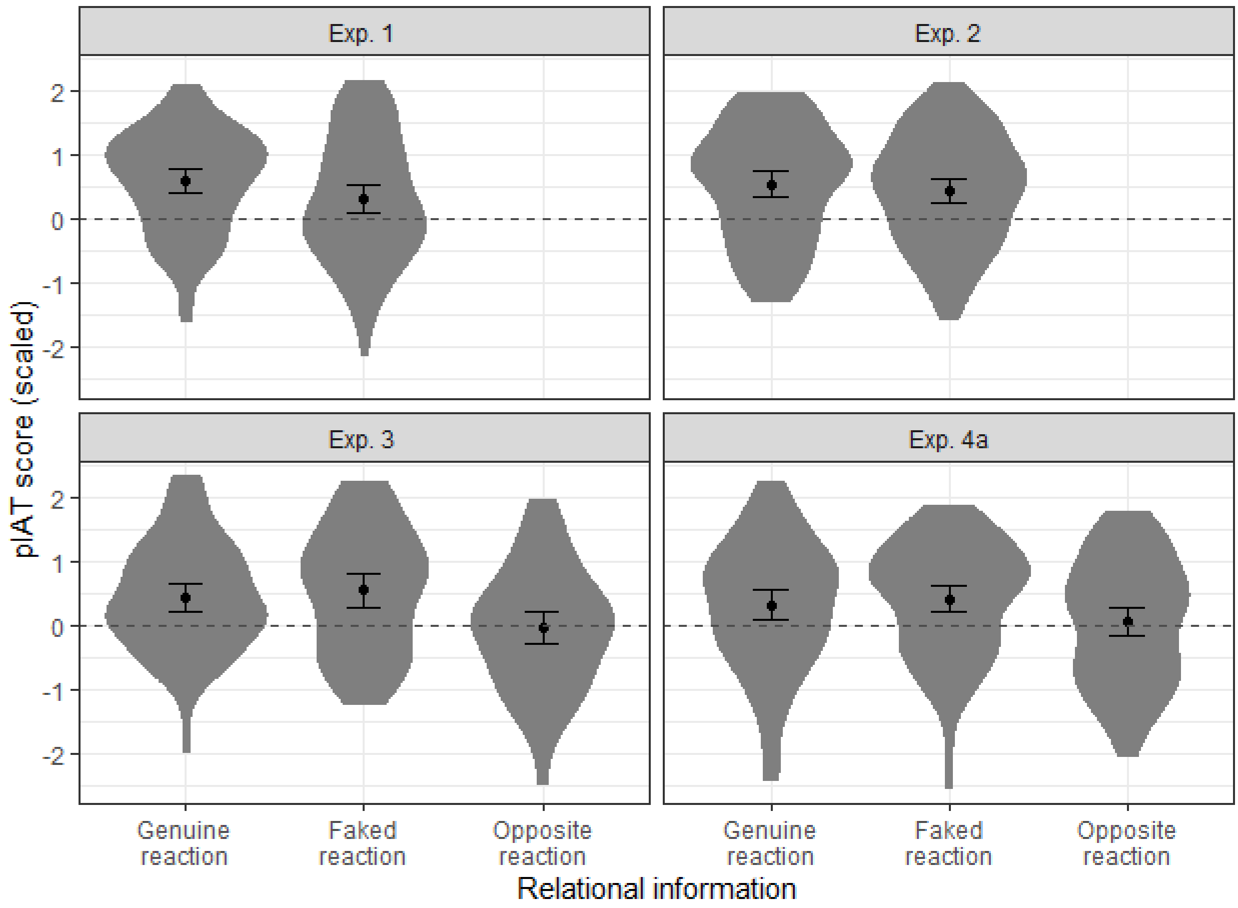
Figure 2 shows the means, confidence intervals, and distributions of scaled pIAT scores (Experiments 1-4a) as a function of relational information. When only the genuine and faked reaction groups were included, there was no main effect of relational information, $F(1,559) = 0.55$, $p = .46$, $\eta^2_p = 0.001$, [0.00, 0.01], $BF_{10} = 0.13$ (note that the BF suggests evidence *for* the null hypothesis). There were only main effects of task order, $F(1,559) = 9.36$, $p = .002$,

η^2_p = 0.02, [0.004, 0.038], BF_{10} = 1.39, such that the OEC effect was larger when the ratings were completed first, and of pIAT block order, $F(1,559) = 38.05$, $p < .001$, η^2_p = 0.06, [0.03, 0.10], BF_{10} > 10 000, such that the OEC effect was larger when the compatible block was completed first. No other effects were significant. However, OEC effects were once again found to emerge only when participants had correct contingency memory.

When all three groups were included, there was no main effect of relational information, $F(2,386) = 1.62$, $p = .20$, η^2_p = 0.008, [0.00, 0.03]. However, relational information interacted with task order, $F(2,386) = 4.76$, $p = .009$, η^2_p = 0.02, [0.003, 0.052], such that there was a clear effect of relational information when the ratings were completed first, $F(2,300) = 14.74$, $p < .001$, but no effect of relational information when the pIAT was completed first, $F(2,83) = 0.15$, $p = .86$. There was also an interaction between relational information and experiment, $F(2,386) = 4.05$, $p = .02$, η^2_p = 0.02, [0.002, 0.046], such that there was a clear effect of relational information in Experiment 3 but only a trend in Experiment 4a. The results of the Bayesian analysis (which assigns more weight to main effects and less to interaction effects) diverged, with a BF_{10} of 120.34 for the main effect of relational information and *BF*s of 0.42 and 0.26 for its interaction with task order and experiment, respectively. Finally, there was a main effect of pIAT block order, $F(1,386) = 47.23$, $p < .001$, η^2_p = 0.11, [0.06, 0.16], BF_{10} > 10 000, such that pIAT scores were larger when the compatible block was completed first. Once again, OEC effects were found only for participants with correct contingency memory.

Figure 2

Means and Distributions of Scaled pIAT Scores in Experiments 1-4a



Demand Compliance

Demand compliance with regard to the ratings differed significantly across the first three experiments, $\chi^2\,(4) = 11.47$, $p = .02$, mostly due to more participants reporting demand compliance in Experiment 3. A comparison of the recoded demand compliance values for Experiment 3 and Experiments 4a-4b (which aimed at reducing this demand compliance) suggested that fewer participants reported demand compliance in Experiments 4a-4b relative to Experiment 3, $\chi^2\,(1) = 6.14$, $p = .01$. However, this result should be interpreted very cautiously, as the phrasing of the questions and their response formats varied.

Demand compliance with regard to the pIAT did not differ significantly across the first three experiments, $\chi^2\,(4) = 3.03$, $p = .55$. Surprisingly, the analysis suggested that many more participants reported some degree of demand compliance in Experiment 4a than in Experiment 3, $\chi^2(1) = 19.76$, $p < .001$ (but see Footnote 7 for an argument that this question was not interpreted as we intended by a number of participants).

Finally, Experiments 4a-4b also required participants to report to what extent their evaluative ratings had been based only on their true feelings. Therefore, combining the data from these two experiments allowed us to test the impact of relational information (genuine vs. faked vs. opposite) on evaluative ratings in a sufficiently large subsample of participants who had indicated the highest value (9) on this scale ($n = 210$). In this subsample, there was a large main effect of relational information on evaluative ratings, $F(2,204) = 41.30$, $p < .001$, $\eta^2_p = 0.29$, [0.20, 0.36], $BF_{10} > 10\ 000$.

Discussion

The analyses on the combined data largely confirm the conclusions of the individual experiments. First, all three types of relational information moderated self-reported evaluations, and this pattern emerged regardless of whether participants reported demand compliance as well as in a subsample of participants who reported responding based only on their true feelings. Second, only the opposite reaction information attenuated automatic evaluations (if measured after participants had completed the self-reports). Finally, both self-reported and automatic OEC effects emerged only for participants who correctly remembered the contingencies between the CSs and the model's reactions.

General Discussion

Social learning research reveals that evaluations can be formed or changed by simply observing others as they interact with stimuli in the environment. One subtype of social learning, OEC, involves a change in liking that is due to pairing a stimulus (CS) with a model's reaction (US). Although prior research provided clear evidence for OEC effects, it did not resolve the debate of whether those effects are mediated by associative or inferential mental-processes (Baeyens et al., 1996, 2001).

Across five experiments we tested an important prediction of an inferential account, namely that relational information would moderate OEC effects. We repeatedly found that OEC effects are sensitive to the perceived nature of the relationship between a stimulus and the model’s reactions. When participants were informed that a model’s reactions were *genuine,* strong OEC effects emerged: after watching a model react positively to one stimulus and negatively to another, participants preferred the former over the latter, as reflected in their self-reported (ratings) and automatic evaluations (pIAT). When they learned that the model’s reactions to the cookie were *faked*, OEC effects still consistently emerged. While self-reported OEC effects were reduced compared to those in the genuine reaction group (analyses on the data pooled across experiments supported this conclusion), pIAT scores did not differ between these groups, with Bayesian analyses of the pooled data providing evidence *for* the null hypothesis. Finally, attenuated (or even reversed) self-reports and attenuated pIAT effects were obtained when participants were informed that the model was displaying the *opposite* reaction to what he actually felt, although the impact of this relational information on the pIAT depended on participants having already completed the self-reports.

Taken together, our results suggest that although OEC effects were influenced by relational information, this influence was often not as strong as expected. In the next section we highlight the similarity of our findings to earlier findings in EC research and discuss how theoretical explanations of those earlier findings might thus also apply to our results.

Theoretical Implications

Our findings exhibit several strong similarities to prior work on EC. First, although the faked reaction information implied that the model’s expressions were unrelated to the valence of the CSs, it reduced self-reported OEC effects only to some extent and failed to attenuate pIAT effects. This resembles studies showing EC effects when CSs and USs are said to be unrelated (Hughes, Ye, Van Dessel, et al., 2019) or even when participants are instructed to actively minimize, prevent, or suppress the impact of CS-US pairings (suggesting that pairings may have an uncontrollable impact on behaviour; Balas & Gawronski, 2012; Gawronski et al., 2014, 2015). Second, although self-reported EC effects are usually reversed when oppositional relational information is provided, participants often still show some impact of CS-US co-occurrences over and above their specific relation (Heycke & Gawronski, 2019; Hütter & Sweldens, 2018; Kukken et al., 2019). In the current studies, the reversal of self-reported OEC effects in the opposite reaction condition was also rather weak and not even significant in two out of three cases. Finally, the fact that automatic evaluations were merely attenuated in this condition also mirrors prior EC research (e.g., Hughes, Ye, & De Houwer, 2019; Moran & Bar-Anan, 2018). Given this similarity in results, it can be interesting to consider how those results have shaped theoretical thinking in EC research before considering the theoretical implications of our results for OEC research.

First, it has been pointed out that the residual uncontrollable impact of pairings on liking could be taken as evidence against propositional accounts of EC (e.g., Gawronski et al., 2014). This conclusion rests on the assumption that people have full control over the propositions they form and the inferences that they make on the basis of those propositions. However, it has also been argued that a residual uncontrollable impact of pairings on liking can be explained by a propositional perspective if one assumes that once a proposition has been formed (e.g., “stimulus A co-occurred with a positive US”), it can be retrieved automatically and influence evaluations (Gawronski et al., 2014; De Houwer, 2018). If this pairing-based proposition conflicts with a proposition that takes relational information into account (e.g., “stimulus A is opposite to a positive US”), this could explain why oppositional information often fails to fully reverse EC effects. Moreover, if the former proposition is easier to retrieve than the latter under automaticity conditions (e.g., when one has to respond quickly), then this assumption can also account for the finding that relational information often has an even weaker impact on automatic evaluations. Finally, it has been suggested that the mean may conceal individual differences in terms of whether participants take relational information into account (De Houwer et al., 2020; Moran et al., 2016).

Second, the aforementioned findings have also been related to dual-process theories of EC. These assume that both propositional and associative processes are involved in EC (e.g., Gawronski & Bodenhausen, 2018; McConnell & Rydell, 2014; for a recent overview of theoretical accounts, see Corneille & Stahl, 2019). Dual-process theories can explain why pairings still have an impact over and above relational information by assuming that EC effects are determined by the combined influence of associations (which are formed based on the pairings) and propositions (which are formed based on combining information from the pairings and the instructions). Moreover, because dual-process accounts generally also assume that the two types of processes differ in terms of their impact on self-reported and automatic evaluations, they can account for the finding that automatic evaluations are often less sensitive to relational information.

These same explanations can also be extended to the current work on OEC. On the one hand, a purely propositional or inferential account would be able to explain the current results by making assumptions similar to those mentioned above. For example, participants may have formed the proposition “the CS_{pos} was followed by a positive reaction” based on the videos and then automatically retrieved this proposition while evaluating the CS_{pos}. Even after receiving the faked reaction information, such a proposition may still have exerted a strong influence, especially considering that this information did not offer any clear implications for inferring the valence of the CSs (i.e., it did not imply that the stimuli were equal in valence and therefore participants may not have formed any other proposition to base their evaluations on). Although the opposite reaction information did allow participants to infer the valence of the CSs (i.e., “the CS_{pos} is bad”), the proposition based on the observations could still have influenced evaluations, especially if it was easier to retrieve automatically because it did not require combining multiple pieces of information.[8] There may also have been individual differences in terms of how participants used the relational information. Based on the distributions of self-reports in Experiments 1-2, we initially speculated that different participants used the faked reaction information in different ways; however, subsequent experiments did not suggest clear individual differences in this group. Interestingly, the opposite reaction group in Experiments 3-4b did seem to include some participants showing a reversed rating effect but others showing no or even a standard effect (see Figure 1).

On the other hand, our findings could also be explained by assuming the involvement of both propositional and associative processes (i.e., a dual-process perspective). Because propositional processes are assumed to play a role, such a perspective could explain why relational information moderated OEC effects. In addition, by assuming that CS-US associations also contribute to OEC effects, it would be able to explain (a) why the impact of relational information was relatively weak and observations still influenced evaluations to some extent in the faked and opposite reaction conditions, and (b) why the impact of relational information was especially weak for automatic evaluations (which dual-process accounts assume to be more sensitive to associations relative to self-reported evaluations; e.g., Gawronski & Bodenhausen, 2018; McConnell & Rydell, 2014). In sum, the current

[8] This explanation is similar to that proposed by episodic memory models (e.g., Stahl & Aust, 2018), which deal with how information is encoded, maintained, and retrieved. Specifically, these models suggest that speeded evaluations are less likely to reflect the valence implied by a specific relation because this requires integrating two pieces of information (in our case, the observed reactions and the relational information). Note that episodic models can also explain why the pIAT was more sensitive to relational information if participants had already completed the ratings, as the rating task entails rehearsal of the integrated CS evaluations.

findings cannot distinguish between these two classes of accounts. However, we should note that it seems unlikely that any single set of data would be able to do so: almost any pattern of results could probably be accommodated by either perspective depending on the additional assumptions that one makes (see also De Houwer et al., 2020).

Nevertheless, just like similar findings in the EC literature had a profound impact on the debate about the mental processes underlying EC, our findings strongly constrain mental models of OEC. Most importantly, the evidence for an impact of relational information on EC played a crucial role in providing support for the idea that propositional processes are involved in producing this effect. Similarly, whereas we cannot exclude the possibility that associations play some role in OEC, the current findings do argue against a purely associative account. This conclusion seems especially relevant for theories of observational conditioning: in the broader literature on observational (fear) conditioning, many researchers assume that CS-US associations mediate these effects (e.g., Askew & Field, 2008; Field, 2006; Heyes, 2012; Olsson & Phelps, 2007; Reynolds et al., 2015, 2018). An associative account of observational conditioning may be able to explain the reduced OEC effect in the faked reaction condition by assuming that participants paid less attention to the videos (because they had already been told the reactions were faked) and that this reduced the opportunity for associations to be formed (see also Baeyens et al., 2001).[9] However, it is not clear how this account would be able to explain the eliminated and in some cases reversed OEC effects in the opposite reaction condition. The current findings can hence inform future theorizing about observational conditioning (and other social learning phenomena that also involve observing a model's emotional reaction in the presence of stimuli) by requiring mental-process accounts to specify at least a partial role for propositional processes.

[9] We thank an anonymous reviewer for drawing our attention to this alternative explanation.

Limitations and Future Directions

The work presented in this paper has several limitations which could inform future research on observational (evaluative) conditioning. First, we always provided relational information *before* the OEC phase. However, past work has found that relational information moderates EC differently depending on when it is presented (e.g., Hu et al., 2017; Zanon et al., 2014). Future work could therefore manipulate when relational information is provided (e.g., either before, during, or after the observation phase) and examine if this impacts OEC effects. Importantly, providing relational information only after the observation phase has been completed would also allow one to exclude the possibility that a reduced OEC effect is simply due to reduced attention during observation of the model (see above).

A second limitation has to do with our measure of automatic evaluations. Although we mainly wished to include a more automatic measure in addition to self-reports, and the pIAT seemed a suitable candidate, this task does limit the conclusions we can draw with regard to the underlying mental processes. Specifically, a differential impact of the relational information on the pIAT and self-reports could be due to differences in automaticity. However, there are also other structural differences between the two tasks (such as CS evaluations being assessed in a relative vs. nonrelative manner), leaving open the possibility that we would have observed a more similar pattern on both measures if their structural fit had been better (see Payne et al., 2008). Therefore, using two tasks that differed only in terms of a specific automaticity condition (e.g., speed) or even using a single task that allows one to disentangle automatic and non-automatic components of responses (i.e., a process dissociation approach; for a review, see Payne & Bishara, 2009) might have provided more insight into the mental processes driving the OEC effects. In addition, recent research has shown that whether the nature of CS-US relations influences IAT performance depends on which CSs are compared within the task: if the CSs differed with regard to the USs they were paired with but not with regard to their relation to those USs, the IAT only reflected the valence of the paired USs; if the CSs differed with regard to their relation to USs, the IAT actually did reflect this relation (Bading et al., 2019). In light of this prior research, it is perhaps not that surprising that the pIAT in our studies mainly reflected the valence of the model's reactions, as each participant received only one type of relational information. Future research on OEC could therefore include a measure of automatic evaluations that is more appropriate for drawing conclusions about the underlying mental processes.

Third, our results consistently indicated that when participants are told that someone faked their reactions, this person's behaviour toward stimuli still heavily influences evaluations of those stimuli. This is an interesting finding that warrants replication and further investigation, as it could have important implications for real-life situations in which people know that a model's behaviour may not be genuine (e.g., watching commercials or television programs). Further research is required to determine whether such observations would also influence other behaviours (such as which products people buy).

Fourth, many participants appeared to be demand *aware*. Our attempts to undermine demand awareness in Experiments 4a-4b were unsuccessful. Most likely, demand awareness was induced by basic elements of the procedure, such as the instruction to read and remember the relational information and the completion of the evaluative measures. Note, however, that the high levels of demand awareness do not necessarily mean that participants actually *complied* with this demand. We consider it unlikely that our results were driven by demand compliance because those results were still evident when participants who reported demand compliance were excluded. Furthermore, Experiments 4a-4b – where a lower percentage of participants reported demand compliance – largely replicated the self-report results, even when including only participants who indicated the highest possible score when asked to what extent they reported only their true feelings. An important caveat should be mentioned here, however: we do not know for certain that our exploratory questions constituted valid measures of demand awareness, compliance, and honesty. It remains a possibility that these questions were themselves sensitive to demand compliance (as in a study by Nichols and Maner, 2008, where a suspicion probe failed to detect awareness of a hypothesis of which participants were in fact aware). Therefore, we are limited in our interpretation of participants' responses to these questions. Future work could seek to provide even stronger evidence for the impact of relational information by using other paradigms and measures less susceptible to demand compliance.

Finally, regardless of whether one favours an inferential or a dual-process explanation of the current findings, applying an inferential account to OEC did lead us to identify an important boundary condition of these effects. As research on observational conditioning has generally been driven by an associative perspective, we believe that testing additional predictions of (single-process) inferential theories can help to further expand our knowledge about factors influencing the strength and direction of observational conditioning. In the current studies, our manipulations focused on one aspect of the observations: the relation between the model's reactions and the stimuli. Yet from an inferential perspective, still other propositions might be involved in forming evaluations based on someone else's behaviour. For example, whether an observer is influenced by a model's reactions should also depend on the relation between the model and the observer: the observer's liking of the model, their perceived similarity, and the extent to which the model is considered a relevant source might all moderate how strongly the model's reactions influence the observer's evaluations. Future research could test these and other predictions derived from an inferential perspective on observational (evaluative) conditioning.

Conclusion

Observing how others react to stimuli can influence our own evaluations of those stimuli. Yet we still know relatively little about the mental processes that mediate such observational evaluative conditioning effects. The work reported here offers evidence that the effects are sensitive to relational information (i.e., information about the relation between stimuli and the model's reactions), which supports the involvement of propositional processes in this type of social learning.

Declaration of Conflicting Interests

The authors declare that there are no conflicts of interest.

References

Askew, C., & Field, A. P. (2008). The vicarious learning pathway to fear 40 years on. *Clinical Psychology Review*, *28*(7), 1249–1265. https://doi.org/10.1016/j.cpr.2008.05.003

Bading, K., Stahl, C., & Rothermund, K. (2019). Why a standard IAT effect cannot provide evidence for association formation: The role of similarity construction. *Cognition and Emotion*, *34*(1), 128–143. https://doi.org/10.1080/02699931.2019.1604322

Baeyens, F., Eelen, P., Crombez, G., & De Houwer, J. (2001). On the role of beliefs in observational flavor conditioning. *Current Psychology*, *20*(2), 183–203. https://doi.org/10.1007/s12144-001-1026-z

Baeyens, F., Vansteenwegen, D., De Houwer, J., & Crombez, G. (1996). Observational conditioning of food valence in humans. *Appetite*, *27*(3), 235–250. https://doi.org/10.1006/appe.1996.0049

Balas, R., & Gawronski, B. (2012). On the intentional control of conditioned evaluative responses. *Learning and Motivation*, *43*(3), 89–98. https://doi.org/10.1016/j.lmot.2012.06.003

Bandura, A. (1971). *Social Learning Theory*. General Learning Press.

Bayliss, A. P., Frischen, A., Fenske, M. J., & Tipper, S. P. (2007). Affective evaluations of objects are influenced by observed gaze direction and emotional expression. *Cognition*, *104*(3), 644–653. https://doi.org/10.1016/j.cognition.2006.07.012

Bayliss, A. P., Paul, M. A., Cannon, P. R., & Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic Bulletin & Review*, *13*(6), 1061–1066. https://doi.org/10.3758/BF03213926

Bry, C., Treinen, E., Corneille, O., & Yzerbyt, V. (2011). Eye'm lovin' it! The role of gazing awareness in mimetic desires. *Journal of Experimental Social Psychology*, *47*(5), 987–993. https://doi.org/10.1016/j.jesp.2011.03.023

Capozzi, F., Bayliss, A. P., Elena, M. R., & Becchio, C. (2015). One is not enough: Group size modulates social gaze-induced object desirability effects. *Psychonomic Bulletin & Review*, *22*(3), 850–855. https://doi.org/10.3758/s13423-014-0717-z

Castelli, L., Carraro, L., Pavan, G., Murelli, E., & Carraro, A. (2012). The power of the unsaid: The influence of nonverbal cues on implicit attitudes. *Journal of Applied Social Psychology*, *42*(6), 1376–1393. https://doi.org/10.1111/j.1559-1816.2012.00903.x

Castelli, L., De Dea, C., & Nesdale, D. (2008). Learning social attitudes: Children's sensitivity to the nonverbal behaviors of adult models during interracial interactions. *Personality and Social Psychology Bulletin*, *34*(11), 1504–1513. https://doi.org/10.1177/0146167208322769

Corneille, O., Mauduit, S., Holland, Rob. W., & Strick, M. (2009). Liking products by the head of a dog: Perceived orientation of attention induces valence acquisition. *Journal of Experimental Social Psychology*, *45*(1), 234–237. https://doi.org/10.1016/j.jesp.2008.07.004

Corneille, O., & Stahl, C. (2019). 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. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior*, *37*(1), 1–20. https://doi.org/10.3758/LB.37.1.1

De Houwer, J. (2018). Propositional models of evaluative conditioning. *Social Psychological Bulletin*, *13*(3), e28046. https://doi.org/10.5964/spb.v13i3.28046

De Houwer, J., Van Dessel, P., & Moran, T. (2020). 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

De Houwer, J., & Vandorpe, S. (2010). Using the Implicit Association Test as a measure of causal learning does not eliminate effects of rule learning. *Experimental Psychology*, *57*(1), 61–67. http://dx.doi.org/10.1027/1618-3169/a000008

Fiedler, K., & Unkelbach, C. (2011). Evaluative conditioning depends on higher order encoding processes. *Cognition and Emotion*, *25*(4), 639–656. https://doi.org/10.1080/02699931.2010.513497

Field, A. P. (2006). Is conditioning a useful framework for understanding the development and treatment of phobias? *Clinical Psychology Review*, *26*(7), 857–875. https://doi.org/10.1016/j.cpr.2005.05.010

Förderer, S., & Unkelbach, C. (2012). Hating the cute kitten or loving the aggressive pit-bull: EC effects depend on CS–US relations. *Cognition and Emotion*, *26*(3), 534–540. https://doi.org/10.1080/02699931.2011.588687

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. https://doi.org/10.1177/0146167213513907

Gawronski, B., & Bodenhausen, G. V. (2018). Evaluative conditioning from the perspective of the associative-propositional evaluation model. *Social Psychological Bulletin*, *13(3),* e28024. https://doi.org/10.5964/spb.v13i3.28024

Gawronski, B., Mitchell, D. G. V., & Balas, R. (2015). Is evaluative conditioning really uncontrollable? A comparative test of three emotion-focused strategies to prevent the acquisition of conditioned preferences. *Emotion*, *15*(5), 556–568. http://dx.doi.org/10.1037/emo0000078

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. http://dx.doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. http://dx.doi.org/10.1037/0022-3514.85.2.197

Heycke, T., & Gawronski, B. (2019). Co-occurrence and relational information in evaluative learning: A multinomial modeling approach. *Journal of Experimental Psychology: General*, *149*(1), 104–124. https://doi.org/10.1037/xge0000620

Heyes, C. M. (2012). What's social about social learning? *Journal of Comparative Psychology*, *126*(2), 193–202. http://dx.doi.org/10.1037/a0025180

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

Hu, X., Gawronski, B., & Balas, R. (2017). Propositional versus dual-process accounts of evaluative conditioning: I. The effects of co-occurrence and relational information on implicit and explicit evaluations. *Personality and Social Psychology Bulletin*, *43*(1), 17–32. https://doi.org/10.1177/0146167216673351

Hughes, S., Ye, Y., & De Houwer, J. (2019). Evaluative conditioning effects are modulated by the nature of contextual pairings. *Cognition and Emotion*, *33*(5), 871–884. https://doi.org/10.1080/02699931.2018.1500882

Hughes, S., Ye, Y., Van Dessel, P., & De Houwer, J. (2019). When people co-occur with good or bad events: Graded effects of relational qualifiers on evaluative conditioning. *Personality and Social Psychology Bulletin*, *45*(2), 196–208. https://doi.org/10.1177/0146167218781340

Hütter, M., & Sweldens, S. (2018). Dissociating controllable and uncontrollable effects of affective stimuli on attitudes and consumption. *Journal of Consumer Research*, *45*(2), 320–349. https://doi.org/10.1093/jcr/ucx124

Jones, B. C., DeBruine, L. M., Little, A. C., Burriss, R. P., & Feinberg, D. R. (2007). Social transmission of face preferences among humans. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1611), 899–903. https://doi.org/10.1098/rspb.2006.0205

King, D., Rowe, A., & Leonards, U. (2011). I trust you; hence I like the things you look at: Gaze cueing and sender trustworthiness influence object evaluation. *Social Cognition*, *29*(4), 476–485. https://doi.org/10.1521/soco.2011.29.4.476

Kukken, N., Hütter, M., & Holland, R. W. (2019). Are there two independent evaluative conditioning effects in relational paradigms? Dissociating the effects of CS-US pairings and their meaning. *Cognition and Emotion*, *43*(1), 170–187. https://doi.org/10.1080/02699931.2019.1617112

Manera, V., Elena, M. R., Bayliss, A. P., & Becchio, C. (2014). When seeing is more than looking: Intentional gaze modulates object desirability. *Emotion*, *14*(4), 824–832. http://dx.doi.org/10.1037/a0036258

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). Guilford.

Mineka, S., & Cook, M. (1993). Mechanisms involved in the observational conditioning of fear. *Journal of Experimental Psychology: General*, *122*(1), 23–38. http://dx.doi.org/10.1037/0096-3445.122.1.23

Mineka, S., Davidson, M., Cook, M., & Keir, R. (1984). Observational conditioning of snake fear in rhesus monkeys. *Journal of Abnormal Psychology*, *93*(4), 355–372. http://dx.doi.org/10.1037/0021-843X.93.4.355

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*(2), 183–198; discussion 198-246. http://dx.doi.org/10.1017/S0140525X09000855

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326. https://doi.org/10.1037/0033-2909.132.2.297

Moran, T., & Bar-Anan, Y. (2018). The effect of co-occurrence and relational information on speeded evaluation. *Cognition and Emotion*, *34*(1), 144–155. https://doi.org/10.31234/osf.io/hwaj4

Moran, T., Bar-Anan, Y., & Nosek, B. A. (2016). The assimilative effect of co-occurrence on evaluation above and beyond the effect of relational qualifiers. *Social Cognition*, *34*(5), 435–461. http://dx.doi.org/10.1521/soco.2016.34.5.435

Moran, T., Bar-Anan, Y., & Nosek, B. A. (2017). The effect of the validity of co-occurrence on automatic and deliberate evaluations. *European Journal of Social Psychology*, *47*(6), 708–723. https://doi.org/10.1002/ejsp.2266

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development*, *72*(3), 718–735. https://doi.org/10.1111/1467-8624.00311

Mumme, D. L., & Fernald, A. (2003). The infant as onlooker: Learning from emotional reactions observed in a television scenario. *Child Development*, *74*(1), 221–237. https://doi.org/10.1111/1467-8624.00532

Nichols, A. L., & Maner, J. K. (2008). The good-subject effect: Investigating participant demand characteristics. *The Journal of General Psychology*, *135*(2), 151–165.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*, *86*(5), 653–667. https://doi.org/10.1037/0022-3514.86.5.653

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience*, *10*(9), 1095–1102. http://dx.doi.org/10.1038/nn1968

Payne, B. K., & Bishara, A. J. (2009). An integrative review of process dissociation and related models in social cognition. *European Review of Social Psychology*, *20*(1), 272–314. https://doi.org/10.1080/10463280903162177

Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology*, 16–31.

Peters, K. R., & Gawronski, B. (2011). Are we puppets on a string? Comparing the impact of contingency and validity on implicit and explicit evaluations. *Personality and Social Psychology Bulletin*, *37*(4), 557–569. https://doi.org/10.1177/0146167211400423

Reynolds, G., Field, A. P., & Askew, C. (2015). Preventing the development of observationally learnt fears in children by devaluing the model's negative response.

Journal of Abnormal Child Psychology, *43*(7), 1355–1367. https://doi.org/10.1007/s10802-015-0004-0

Reynolds, G., Wasely, D., Dunne, G., & Askew, C. (2018). A comparison of positive vicarious learning and verbal information for reducing vicariously learned fear. *Cognition and Emotion*, *32*(6), 1166–1177. https://doi.org/10.1080/02699931.2017.1389695

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Skinner, A. L., Meltzoff, A. N., & Olson, K. R. (2017). “Catching” social bias: Exposure to biased nonverbal signals creates social biases in preschool children. *Psychological Science*, *28*(2), 216–224. https://doi.org/10.1177/0956797616678930

Skinner, A. L., Olson, K. R., & Meltzoff, A. N. (2020). Acquiring group bias: Observing other people’s nonverbal signals can create social group biases. *Journal of Personality and Social Psychology*, *119*(4), 824–838. http://dx.doi.org/10.1037/pspi0000218

Skinner, A. L., & Perry, S. (2020). Are attitudes contagious? Exposure to biased nonverbal signals can create novel social attitudes. *Personality and Social Psychology Bulletin*, *46*(4), 514–524. https://doi.org/10.1177/0146167219862616

Stahl, C., & Aust, F. (2018). Evaluative conditioning as memory-based judgment. *Social Psychological Bulletin*, *13*(3), e28589. https://doi.org/10.5964/spb.v13i3.28589

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

Treinen, E., Corneille, O., & Luypaert, G. (2012). L-eye to me: The combined role of Need for Cognition and facial trustworthiness in mimetic desires. *Cognition*, *122*(2), 247–251. https://doi.org/10.1016/j.cognition.2011.10.006

Weisbuch, M., & Ambady, N. (2009). Unspoken cultural influence: Exposure to and influence of nonverbal bias. *Journal of Personality and Social Psychology*, *96*(6), 1104–1119. https://doi.org/10.1037/a0015642

Weisbuch, M., Pauker, K., & Ambady, N. (2009). The subtle transmission of race bias via televised nonverbal behavior. *Science*, *326*(5960), 1711–1714. https://doi.org/10.1126/science.1178358

Zanon, R., De Houwer, J., Gast, A., & Smith, C. T. (2014). When does relational information influence evaluative conditioning? *The Quarterly Journal of Experimental Psychology*, *67*(11), 2105–2122. https://doi.org/10.1080/17470218.2014.907324