When People Co-occur with Good or Bad Events: Graded Effects of Relational Qualifiers on Evaluative Conditioning

Sean Hughes, Yang Ye, Pieter Van Dessel, and Jan De Houwer

Ghent University

Authors Note

SH, YY, PVD, and JDH, Department of Experimental Clinical and Health Psychology, Ghent University. This research was conducted with the support of Grant BOF16/MET_V/002 of Ghent University to JDH. PVD is supported by a Post-Doctoral Fellowship of the Research Foundation – Flanders (FWO). Correspondence concerning this article should be sent to sean.hughes@ugent.be.

Abstract

Studies on evaluative conditioning show that a change in liking can occur whenever stimuli are paired. Such instances of attitude change are known to depend on the type of relation established between stimuli (e.g., "*Bob is a friend of Mike*" vs. "*Bob is an enemy of Mike*"). Research has so far only compared assimilative and contrastive relational qualifiers (e.g., friend vs. enemy). For the first time, we compared the effect of non-oppositional qualifiers on attitude change in a EC procedure (e.g., "*Bob causes Positive Outcomes*" vs. "*Bob predicts Positive Outcomes*"). Differential effects of non-oppositional relational qualifiers were observed on explicit and implicit evaluations. We discuss the implications of our findings for attitude research, theories of attitude change, and for optimizing evaluative conditioning for changing attitudes in applied settings.

Keywords: Evaluative Conditioning, Cognitive Balance, Relational Qualifiers, Implicit

When People Co-occur with Good or Bad Events: Graded Effects of Relational Qualifiers on Evaluative Conditioning

Whether at parties or in the supermarket, workplace, or gym, we continually encounter a stream of unknown individuals and are faced with the task of forming impressions of them. These interpersonal preferences can be formed in a variety of ways, from information acquired through direct contact (Hütter, Kutzner, & Fiedler, 2014), to observations (Castelli, De Dea, & Nesdale, 2008), instructions (Gawronski, Ye, Rydell, & De Houwer, 2014), or deductions (Walther, 2002). One particularly well-known pathway for establishing new or changing existing social attitudes is evaluative conditioning (EC). In a typical EC study, a neutral conditioned stimulus (CS) (e.g., an unknown person) comes to be liked or disliked when it is paired with a positive or negative unconditioned stimulus (US) (e.g., a pleasant or unpleasant image or statement). Hence, EC research can inform us about the conditions under which, and the ways in which, spatio-temporal co-occurences influence interpersonal preferences towards known and unknown groups and individuals (e.g., Gawronski, & Walther, 2008; Olson, & Fazio, 2006; for a meta-analysis see Hofmann, De Houwer, & Perugini, Baeyens, & Crombez, 2010). In the real-world, EC may help to explain a variety of phenomena relevant to person perception, from the fact that persuasive messages (CS) communicated by attractive individuals (positive US) are liked more than messages communicated by unattractive sources (negative US; Tormala & Briñol, 2015), the transfer of negativity from a stigmatized group to associated individuals and groups (i.e., the *sins of the father* effect; Pryor, Reeder, & Monroe, 2012), to the *kill-the-messenger* effect, where the person transmitting a message often acquires its evaluative properties (Walther, Nagengast, & Trasselli, 2005).

More often than not, EC researchers have focused on the effects of mere stimulus pairings. Yet, in real-life, people usually encounter *relational information* that suggests why stimuli co-occur. For instance, we might learn that, although Bob and Mike are both paired with positive outcomes, the former causes those outcomes whereas the latter prevents them. Similarly, we might learn that a work colleague is loathed by the boss or that a politican continually links themselves to a pro-environmental policies despite the fact that their behavior indicates otherwise. In each case, *relational information,* or information about the specific way in which co-occuring stimuli are related has clear implications for the evaluative implications of stimulus co-occurrences.

With this in mind, a growing number of EC studies have begun to combine relational information with stimulus pairings and demonstrated that this information moderates the direction and magnitude of EC effects. For instance, Fiedler and Unkelbach (2011), paired a series of novel faces with one another: one face was always neutral whereas the second was positive or negative. Participants were told (or actively construed) that the faces were friends or enemies and asked, following an EC phase, to provide impressions of them. When faces were framed as friends, participants showed an assimilative EC effect (i.e., they liked the individual paired with a positive face and disliked the person paired with a negative face). Yet, people who related the faces as enemies showed a reversed or attenuated EC effect. Similarly, others have found that when people are told that a novel individual (CS) loathes cute kittens (US) (Förderer & Unkelbach, 2012), a fictitious creature starts a horrible noise or stops a good one (Moran & Bar-Anan, 2013), or that positive statements about Bob are invalid (Peters & Gawronski, 2011), EC effects are reversed or attenuated compared to situations where no such information is provided. These findings suggest that the specific way in which stimuli (people) are related exerts a powerful impact on the direction and magnitude of interpersonal preferences established via EC. These effects have been observed for both deliberative (i.e., explicit) and spontaneous (i.e., implicit) evaluations. Whereas explicit evaluations are typically assessed using questionnaires and self-report ratings, implicit evaluations are captured using indirect procedures such as the Implicit Association Test (IAT; Greenwald, Mcghee, & Schwartz, 1998) and the Affect Misattribution Procedure (Payne, Cheng, Govorun, & Stewart, 2005). These latter evaluations are often emitted under the various conditions of automaticity (e.g., quickly and without control) and can either converge or diverge from their explicit counterparts, and either independently or additively explain future 'downstream' behavior in ways that explicit evaluations do not (for an overview see Gawronski & Payne, 2010).

Implications of a Relational Approach to EC

The growth in studies examining how relational information moderates EC effects has theoretical significance for researchers interested in EC in particular and social psychology more generally.

Implications for EC. Theorising on EC has long been dominated by single-process associative stimulus-stimulus (S-S; e.g., Walther, Gawronski, Blank, & Langer, 2009) and stimulus-response (S-R) models (e.g., Sweldens, Van Osselaer, & Janiszewski, 2010). Changes in liking due to stimulus pairings were thought to be mediated by the automatic formation and activation of associative links between represents of a CS and US. However, the failure of such accounts to accommodate a range of findings has recently led to the emergence of two alternatives (see Hofmann et al., 2010, for a review). First, a single-process propositional perspective argues that all EC effects are mediated propositions about the CS-US relation. Unlike associations, which have no truth-value or content, propositions specify the way in which stimuli are related and can therefore be true or false (De Houwer, 2009, 2014; Mitchell, De Houwer, & Lovibond, 2009). The second alternative is offered by dual-process accounts which argue that EC effects can be mediated by associative and/or propositional processes (e.g., Gawronski & Bodenhausen, 2011). It is typically assumed that explicit evaluations reflect so-called propositional processes (e.g., rule-based inferences, Gawronski, Brannon, & Bodenhausen, 2016) whereas implicit evaluations reflect so-called associative processes (e.g., spread of activation of associations formed on the basis of feature similarity and spatio-temporal contiguity, Gawronski et al., 2016).

Researchers have attempted to distinguish between propositional and dual-process accounts by focusing on whether relational information moderates the impact of stimulus pairings on implicit evaluations, and in particular, whether this depends on the frequency of CS–US pairings (see Hu, Gawronski, & Balas, 2016). Propositional and dual process theories take a slightly different stance on the role that relational information plays in EC effects. Whereas propositional accounts argue that propositions about the CS-US relationship can have a direct impact on implicit evaluations (i.e., without any involvement of associative representations), dual-process accounts only allow for an indirect impact of propositions on implicit evaluations, that is, via the impact that propositions have on associations.

The two accounts also differ in their assumptions about the role that repeated co-occurrences play in EC effects. Studies that involve the delivery of relational information before (Moran & Bar-Anan, 2013), during (Förderer & Unkelbach, 2012), or after pairings (Zanon, De Houwer, Gast, & Smith, 2014) lead to a situation where the individual is confronted with two pieces of information: repeated co-occurrences and instructions or words that operate on, and typically reinterpret the meaning of, those co-occurrences. Propositional accounts might argue that co-occurrences lead to one proposition being formed and validated on the basis of experience (‘CS co-occurs with the US’) whereas relational qualifiers (e.g., the instruction that things that co-occur are opposite) generate a second proposition via instructions (‘CS is opposite to the US’). If the latter is considered more valid and relevant than the former, then CS evaluations should reflect the meaning of this reversed relation rather than mere co-occurrences. In contrast, dual-process accounts argue that relational information should moderate the impact of pairings on implicit evaluations to a much lesser extent when the number of CS-US co-occurrences is low (e.g., 8 pairings), rather than high (e.g., 24 pairings). Given the assumption that CS-US co-occurrences have a direct impact on the strength of associations, with many CS-US co-occurrences, CS-US associations will be strong irrespective of the type of relational information. Hence, whenever the CS and US co-occur frequently, implicit evaluations should reflect “unqualified co-occurrence effects irrespective of the manner in which the co-occurring stimuli are related” (Hu et al., 2016, p.19). As we will discuss below, the current study investigated the combined effects of relational information and CS-US co-occurrences on implicit and explicit evaluations and thus put these two competing accounts to the test.

Implications beyond EC. The central idea of the current research, that propositional relations between a novel stimulus (e.g., a person) and a liked or disliked stimulus (e.g., a good or bad behavior) can influence the formation of attitudes towards the novel stimulus, is in-line with at two major classes of theories of social psychology: balance theories and attribution theories.

Balance theories (e.g., Cartwright & Harary, 1956; Feather, 1971; Heider, 1958; Read, Vanman, & Miller, 1997; for a review, see Crandall, Silvia, N’gbala, Tsang, & Dawson, 2007) suggest that attitude formation and change are guided by cognitive consistency principles towards a state of balance that consists of logically or affectively uniform impressions of others. One of the most well-known is Heider’s (1958) P-O-X model, where he defined balance or unbalance in terms of positive or negative relationships (denoted as ‘+’or ‘–’, e.g., liking and disliking) among a triad of three units: the person “P”, a comparison person “O”, and a third comparison unit “X” (e.g., an object, an idea, an event, or another person). For example, if P likes O (‘P+O’), and O likes X (‘O+X’), then the system of relationships among the three units will be balanced if P also likes X (‘P+X’) and unbalanced if P dislikes X (‘P–X’). According to Heider, attitude formation follows the balance principle in that the newly formed attitude tends to be in balance with existing attitudes or relationships. For example, “my friend’s friend is my friend”, and “my enemy’s friend is my enemy”.

One can directly apply Heider’s (1958) P-O-X model to the analysis on how sameness vs. opposition CS-US relationships can influence EC (e.g., Fiedler & Unkelbach, 2011; Förderer & Unkelbach, 2012; Hu et al., 2016; Moran & Bar-Anan, 2013). According to Heider’s model, a person (P) should like or dislike a CS (X) paired with a US (O) on the basis of (1) the pre-existing positive or negative relationship (i.e., liking or disliking) between the person and the US, and (2) the relationship (e.g., sameness, opposition) between the CS and the US. For example, to the extent that the person likes (‘+’) a pleasant US, the person should dislike (‘–’) a CS that stops (‘–’) the US, and like (‘+’) a CS that causes (‘+’) the US. This prediction is consistent with existing findings about how positive and negative CS-US relationships influence EC on both explicit attitudes (e.g., Moran & Bar-Anan, 2013) and implicit attitudes (e.g., Hu et al., 2016).

While Heider’s (1958) balance theory is typically applied to the analysis of attitude formation and change in tradic relationships, other aspects of the balance are overlooked. For example, an important aspect of balance is the notion that affect (e.g., attitude) should be consistent with perception (see Crandall et al., 2007). Heider himself has discusssed this type of affect-perception consistency in his theory about attribution (Heider, 1944, 1958, 1988; also see Weiner, 1986, 1995). According to these theories, the effect of a piece of new information (e.g., Bob helps an elderly person to cross the street) on attitudes (e.g., liking of Bob) is modulated by the attribution of responsibility for the event (i.e., helping an elderly) to the focal person (i.e., Bob). According to Heider (1958), higher levels of attribution should result in a stronger “unit relation” between Bob and the event. The affective feeling that a person develops towards Bob (i.e., attitudes) on the basis of the event should be consistent with the strength of the unit relation between Bob and the event. In other words, the more one perceives Bob as responsible for a positive event, the more the person should like Bob, attaining affect-perception consistency (also see Duval & Duval, 1983; Read, Vanman, & Miller, 1997). Surprisingly, few studies has tested these hypotheses on the effects of attribution on attitude formation (for exceptions, see Weiner, 1979, 1986, 1995), especially in the case of implicit attitudes (i.e., automatic feelings). To the extent that different relational qualifiers (e.g., cause vs. predict vs. unrelated) may imply various degrees of attribution of responsibility of US occurrence to the CS, the current research fills in the gap between the evidence and theories on attribution and attitude formation.

The Current Research

Although it is clear that relational qualifiers can moderate evaluations, and have implications for theorising inside and outside of EC, surprisingly little is known about whether certain qualifiers moderate the impact of pairings on liking to a *greater* or *lesser* extent than others. Prior research only compared relational qualifiers that refer to opposite types of relations (e.g., friend/enemy; cause/prevent; start/stop; synonym/antonym). Although some have argued that "EC effects depend crucially on a distinction between assimilative predicates (i.e., an "is similar to"–relations) and contrastive predicates (i.e., "is different from"–relations) employed during encoding CS-US pairings" (Unkelbach & Fiedler, 2016, p.415), we take a much broader position. We agree that EC effects depend on how the individual relates CS to US, yet, we believe that many types of relations should influence EC effects, even those that do not strictly involve assimilation or contrast.

To the best of our knowledge, no study has ever compared different types of non-oppositional qualifiers in order to determine the *extent* to which they moderate the impact of pairings on explicit and implicit evaluations. With this in mind, we conducted a pilot study to identify three relations that moderated the impact of statements on liking to a strong ("*object* *causes valenced event*"), medium ("*object predicts valenced event*"), and weak ("*object is unrelated to valenced event*") extent. We then assessed if those same relations would moderate the impact of stimulus pairings on liking in a similar way. Participants were first provided with relational information about the meaning of paired events and then encountered actual co-occurrences for themselves. Thereafter we administered self-report questionnaires as well as an indirect measure (IAT) of evaluation. Doing so allowed us to examine whether the same graded pattern of likes and dislikes would emerge for both explicit and implicit evaluations. Given that the predictions of propositional and dual-process models regarding the impact of relational information deviate more clearly when the number of CS-US pairings is high, we made sure to present each CS-US pair many times. Observing an impact of relational information on EC would not only extend our knowledge about the role of relational information in EC but would also highlight the link between EC and research on attribution in person perception (e.g., Heider, 1958).

Pilot Study

We conducted an initial pilot study wherein two hundred and twelve participants (110 men, 102 women) ranging from 17 to 72 years ($M = 32.2$, $SD = 11.1$) were presented with a survey consisting of 68 statements that each contained three basic elements: a generic object (always referred to as 'the object'), a relational qualifier (e.g., 'causes'), and a generic subject (always referred to as "positive events" or "negative events"). There were thirty-four different types of relational qualifiers, each paired once with 'positive events' and again with 'negative events' (see Supplementary Materials). By holding the object and valenced event constant, and varying the relation between them, we examined whether, and to what extent, explicit evaluations of the object differed depending on the way in which stimuli were related. Analyses revealed that the type of relation moderated the strength of evaluative transfer from US to CS, both for statements that produced positive, $F(33, 211) = 74.46$, $p <$

.001, $\eta^2_p = .26$, and negative evaluations, $F(32, 211) = 91.79$, $p < .001$, $\eta^2_p = .30$. The vast majority of statements gave rise to an explicit evaluation that significantly differed from neutral ($p < .05$), and the magnitude of those ratings differed as a function of type of relation involved (see Supplementary Materials). Descriptively, the most extreme evaluations (both positive and negative) were those in which the source *caused*, *owned*, or was *hierarchically related* to the valenced event in some way. The least extreme evaluations (both positive and negative) were those in which the source was *physically related*, *conditionally related*, or *unrelated* to the valenced event.

Main Experiment

Our pilot study revealed that the specific way in which an object and valenced events are related influences both the direction and magnitude of attitude change. Based on these data, we selected three relations that moderated evaluative responding in a strong (causal, ‘causes’), medium (predictive, ‘predicts’), and weak (unrelated, ‘is unrelated to’) manner and examined how the type of relation moderates EC in the context of person perception.[1] We therefore used these relations and presented participants with a short story about two unknown individuals (CSs) who were said to either cause, predict, or be unrelated to the presence of positive or negative words (USs) in a subsequent task. Participants were then exposed to repeated pairings of the person’s photo and valenced words. Thereafter they provided self-report ratings and completed an IAT. Based on our pilot study, we expected a graded pattern of attitude change that depended on the type of relationship established between individuals and valenced events.

Method

[1] Analyses from the pilot study revealed that causal, predictive, and unrelated relations led to positive evaluations that significantly differed from one another, $F(2, 211) = 440.78$, $p < .001$, $\eta 2p = .68$, as well as negative evaluations that significantly differed from one another, $F(2, 211) = 178.39$, $p < .001$, $\eta 2p = .46$. Post-hoc comparisons of the positive ratings indicated that ‘causal’ relations led to more extreme evaluations than ‘predictive’ or ‘unrelated’ relations (all *ps* < .001), and that ‘predictive’ produced more extreme ratings than ‘unrelated’ relations (all *ps* < .01).

Participants and design. A total of 180 English-speaking volunteers (115 men, 65 women) ranging from 18 to 39 years (*M* = 28.3, *SD* = 5.6) participated online via the Prolific Academic website (https://prolific.ac) in exchange for a monetary reward (€1.50). The experiment was programmed in Inquisit 4.0 and hosted via Inquisit Web (Millisecond Software, Seattle, WA). The experiment involved a single-factor between-subjects design (*Relation Type*: causal vs. predictive vs. unrelated), with self-reported and implicit evaluations as the main dependent variables. Three method variables were manipulated between participants: evaluative task order (self-reports vs. IAT first), IAT block order (learning [EC] phase consistent vs. inconsistent first) and stimulus assignment (which person was paired with positive or negative words). The sample size was determined prior to data collection. We stopped data-collection whenever at least 180 participants had completed all measures of the experiment to ensure that we would have sufficient statistical power to detect medium effects even after data cleaning (necessary sample size = 154 to have power = 0.80 to find a medium effect of η^2_p = 0.06 at alpha = 0.95). Note that the study designs and data-analysis plans for the pilot and main experiment are available on the Open Science Framework website (https://osf.io/me9dh/). We report all manipulations and measures used in the study. All data were collected without intermittent data analysis. The data analytic plan, experimental scripts, and data are available at the above link.

Materials

Stimuli. Two male faces were selected from the CAL/PAL Face Database (Minear & Park, 2004) to serve as CSs, based on previous studies conducted in our lab suggesting that they have a neutral valence. Ten positive (*fantastic, wonderful, honest, kind, brave, amazing, nice, pleasant, selfless, great*) and ten negative adjectives (*horrible, nasty, violent, terrible, hated, disgusting, mean, unpleasant, stupid, bully*) served as USs.

IAT. Automatic evaluative responding was measured using an IAT. Images of the two male faces served as one set of target stimuli and the words "Good" and "Bad" as another. Eight positively valenced and eight negatively valenced adjectives served as one set of attribute stimuli (*confident, friendly, cheerful, loyal, generous, loving, funny, warm* versus *liar, cruel, evil, ignorant, manipulative, rude, selfish, disloyal*) and images of the two faces served as the second set.

Procedure

Participants were first provided with a general overview of the experiment and then asked for their informed consent. The study consisted of four phases: stimulus pre-ratings, EC, evaluative measures, and exploratory questions.

Stimulus pre-ratings. Participants were presented with the ten positive and ten negative adjectives one at a time on the screen and asked to indicate how positive or negative they considered them to be using a scale ranging from −5 (negative) to +5 (positive) with 0 as a neutral point. Pre-ratings of the positive ($M = 3.72$, $SD = 0.81$) and negative USs ($M = -3.73$, $SD = 0.81$) significantly differed from one another, $t(179) = 67.79$, $p < .001$.

Causal instructions. Prior to the EC phase, participants in the causal, predictive, and unrelated conditions were each given a unique set of instructions. Those in the causal relation condition were given the following information:

"*In the next part of the study you will see a person and a word on the screen. The people you will see are actually two participants from a previous study. We gave them the opportunity to earn some extra money by helping us design this experiment. Specifically, their task was to decide what words you (and the other participants in this experiment) will see (i.e., they selected the words you are about to encounter). Please pay careful attention to the images of the two people and the words they are causing you to see*".

After each block of trials these instructions were repeated: "*Remember: the two people you see during the experiment are directly responsible for the words you are seeing (they caused them to appear)*."

Predictive instructions. Participants in the predictive relation condition received different information:

"*In the next part of the study you will see a person and a word on the screen. The people you will see are actually two participants from a previous study. We gave them the opportunity to earn some extra money by helping us design this experiment. Specifically, they are going to help you predict what type of words you (and the other participants in this experiment) will see. They did not choose the words you are going to see - we (the experimenters) did. The images of people will simply be present onscreen so that you can predict what type of word will appear. Please pay careful attention to the images of the two people and the words they predict will occur."*

Then after each block they were told the following: "*Remember: the two people you see during the experiment are not responsible for the words you are seeing (we the experimenters are). The people you are seeing simply predict what types of words will appear onscreen."*

Unrelated instructions. Participants in the unrelated relation condition were told:

"*The people you are going to see are actually two participants from a previous study. We are going to show you the words that they saw during their experiment. Important: we (the experimenters) are responsible for the words you will see in this experiment. The people you see are in no way related to the words (other than the fact that they [just like you] had to view them)*."

Then they were reminded after each block that: "*The two people you see during the experiment are unrelated to the words you are seeing. We, the experimenters, are responsible for the words that are appearing onscreen.*"

Note that in all three conditions participants could not proceed from the instruction to EC phase until they could accurately report the relation between the CS and US specified in the instructions they had received.

EC. The EC procedure consisted of three blocks of twenty trials (60 trials total). Each trial simultaneously presented a neutral male face (CS) on the left and valenced adjective (US) on the right sides of the screen for 2500ms. During each block one of the faces (CS1) was presented with one of ten different positively valenced adjectives whereas the second face (CS2) was presented with one of ten different negative adjectives. Hence, each CS was paired with a US for a total of 30 trials, which is more than the number of pairings that was used in the high-frequency pairing conditions of earlier studies (i.e., 24 pairings in Hu et al., 2016). The order of trials was random and the inter-trial interval was 1250ms. Assignment of the CSs to positive or negative USs was counterbalanced across participants.

Self-reported ratings. Self-reported ratings were assessed using four different semantic differential scales. On each trial, one of the two CSs was presented and participants were asked to indicate their general impression of that stimulus using a scale ranging from −5 to +5 with 0 as a neutral point. The four end-points of the scales were as follows: *Negative-Positive, Pleasant-Unpleasant, Good-Bad, I Like It-I Don't Like It*. A mean evaluative rating was calculated for each CS by averaging scores from these four scales.

IAT. We assessed whether the combination of relational information and stimulus pairings led to a change in automatic evaluative responding using the IAT. Participants were informed that a series of images and words would appear one-by-one in the middle of the screen and that their task was to categorize those items with their respective target (Face 1 or

Face 2) or attribute categories (‘Good’ and ‘Bad’) as quickly and accurately as possible. They were told that the two individuals they had previously encountered (targets) as well as the words “Good” and “Bad” (attributes) would appear on the upper left and right sides of the screen and that stimuli could be assigned to these categories using either the left (‘E’) or right keys (‘I’). Each trial started with the presentation of a fixation cross for 200ms in the middle of the screen, followed immediately by a target or attribute stimulus. If the participant categorized the image or word correctly – by selecting the appropriate key for that block of trials – the stimulus disappeared from the screen and the next trial began. In contrast, an incorrect response resulted in the presentation of a red “X” which remained on-screen until the correct key was pressed. Overall, each participant completed seven blocks of trials. The first block of 20 practice trials required them to sort images of two individuals into their respective categories, with Face 1 assigned to the right (‘I’) key and Face 2 to the left (‘E’) key. On the second block of 20 practice trials, participants assigned negative words to the “Bad” category using the left key and positive words to the “Good” category using the right key. Blocks 3 and 4 (20 and 40 trials, respectively) involved a combined assignment of target and attribute stimuli to their respective categories. Specifically, participants categorized Face 1 and positive words using the right key and Face 2 and negative words using the left key. The fifth block of 20 trials reversed the key assignments, with Face 2 now assigned to the right key and Face 1 to the left key. Finally, the sixth and seventh blocks (20 and 40 trials respectively) required participants to categorize Face 1 with negative words and Face 2 with positive words. Participants then completed the following exploratory questions.

Confidence ratings. Participants were asked to indicate how confident they were about their opinion of each of the two individuals using a scale ranging from −5 (Not confident at all) to +5 (Confident) with 0 (Somewhat Confident) as a mid-point.

Instruction memory. We assessed whether participants could recall the relation between the CSs and USs using the following statement: "*At the beginning of the experiment we told you something about the relationship between the positive and negative words and the two men. What did we tell you?*". Participants could indicate one of the following four options: (a) The two men caused the positive or negative words to appear, (b) the two men predicted when the positive or negative words would occur, (c) the two men were unrelated to the words that appeared, or (d) I don't remember.

Belief in the instructions. We assessed whether participants believed the instructions we provided using the following question: "*To what extent did you believe us when we told you (at the beginning of the experiment) that the two men you saw CAUSED/PREDICTED/ WERE UNRELATED TO the positive and negative words that appeared in this experiment?*". Note that participants in the three conditions only saw the relational qualifier that applied in their particular case. They could respond using a scale ranging from 1 (*I did not believe you at all*) to 9 (*I completely believed you*) with 5 (*I somewhat believed you*) as a mid-point.

Demand. Demand compliance was assessed via the following item: "*Earlier we asked you to indicate how positive/negative you felt about the two faces. On what basis did you give those ratings?*". Participants could select from the following possibilities: (a) "I responded based on what I thought the experimenter wanted me to do", (b) "I responded based on what I learned about the stimuli earlier in the experiment", (c) "I responded based on my feelings about the stimuli", or (d) "I do not know why I responded the way I did".

Reactance. Reactance was assessed using a two-step procedure. First, participants were asked "Do you think we tried to influence how you felt about the two faces?" and could respond either "yes" or "no". If they selected "yes" then they were asked a follow-up question: "To what extent did you resist our attempts to influence how you felt?". They could respond using a scale ranging from 1 (*I did not resist your attempts to influence how I felt at* *all*) to 9 (*I very strongly resisted your attempts to influence how I felt*) with 5 (*I slightly resisted your attempts to influence how I felt*) as a mid-point.

Results

Analytic Strategy

To determine whether self-reported and implicit evaluations of the two face stimuli (dependent variables) differed as a function of the type of relation (causal, predictive, and unrelated) established between the faces and valenced words by the instructions and EC procedure (independent variables), a series of ANOVAs and post-hoc tests were carried out on the rating and IAT data.

Data Preparation

IAT. Following the recommendations of Greenwald, Nosek, and Banaji (2003), response latency data were prepared using the D4 scoring algorithm. The resulting D4 IAT scores reflect the difference in mean response latency between the critical blocks divided by the overall variation in those latencies. The IAT score was calculated so that positive values reflected a response bias for the face paired with positive stimuli (i.e., CS1) relative to the face paired with negative stimuli (i.e., CS2) whereas negative values indicated a reverse pattern of responding. According to our data analysis plan, IAT data were removed for participants who (a) had error rates above 30% when considering all IAT blocks or above 40% for any one of the critical IAT test blocks (7 participants; i.e., 3.9%), or (b) responded faster than 400 ms on more than 10% of the IAT trials (8 participants; 4.4%). This left us with 165 participants for analyses on the IAT data.

Preliminary analyses. Thirteen individuals (7.2%) failed to recall the relation between the CS and US during the instruction memory question whereas seven (3.9%) reported that they were demand compliant. Including or removing these individuals did not change the statistical significance of the effects reported below. As such, all participants were included.

Hypothesis Testing

Self-reported ratings. Submitting evaluative ratings to a 3 (*Relation Type*: Causal vs. Predictive vs. Unrelated) × 2 (*Stimulus*: CS1 vs. CS2) × 2 (*Evaluative Task Order*: Self-Reports vs. IAT first) mixed ANOVA revealed a main effect for Stimulus, $F(1, 172) = 226.77$, $p < .001$, as well as two-way interactions between Stimulus and Relation Type, $F(2, 172) = 9.89$, $p < .001$, and between Stimulus and Evaluative Task Order, $F(2, 172) = 10.87$, $p < .001$. Overall, participants rated CS1 positively ($M = 2.19$, $SD = 1.95$) and CS2 negatively ($M = -1.78$, $SD = 2.19$), $t(179) = 13.77$, $p < .001$, $d = 1.03$, $BF_1 > 10^5$. These ratings were more extreme when the self-reported ratings were completed before (CS1: $M = 2.69$, $SD = 1.86$; CS2: $M = -2.39$, $SD = 2.12$) relative to after the IAT (CS1: $M = 1.56$, $SD = 1.88$; CS2: $M = -1.04$, $SD = 2.04$). Critically, the extremity of self-reported ratings differed depending on the type of relation established between the CS and US. Planned comparisons revealed a bigger difference in ratings in the causal relation condition (CS1: $M = 2.81$, $SD = 1.91$; CS2: $M = -2.53$, $SD = 2.12$, $d = 1.37$) than in the unrelated relation condition (CS1: $M = 1.58$, $SD = 1.90$; CS2: $M = -0.90$, $SD = 2.04$, $d = 0.70$), $t(117) = 4.22$, $p < .001$, $d = 0.77$, $BF_1 = 843.81$, or the predictive relation condition (CS1: $M = 2.18$, $SD = 1.87$; CS2: $M = -1.93$, $SD = 2.11$, $d = 1.11$), $t(118) = 1.77$, $p = .04$, $d = 0.32$, $BF_1 = 1.51$. We also observed a bigger difference in ratings in the predictive relation condition than in the unrelated relation condition, $t(119) = 2.49$, $p = .007$, $d = 0.46$, $BF_1 = 6.25$. CS1 and CS2 ratings differed from zero in the causal, predictive, and unrelated conditions (*ps* < .001).

IAT. Submitting IAT scores to a 3 (*Relation*) × 2 (*IAT Block Order*: Learning [EC] phase Consistent vs. Inconsistent) between-subjects ANOVA revealed signficant main effects for Relation, $F(2, 152) = 3.22$, $p = .04$, and IAT Block Order, $F(1, 152) = 18.07$, $p < .001$.

Overall, the IAT scores revealed a preference for CS1 over CS2 (*M* = 0.28, *SD* = 0.40), *t*(163) = 9.00, *p* < .001, *d* = 0.70. This difference was more extreme whenever participants began the task in a manner that was compatible (*M* = 0.41, *SD* = 0.33) relative to incompatible (*M* = 0.15, *SD* = 0.40) with the stimulus pairings in the EC procedure. Critically, the magnitude of the IAT effect differed depending on the type of relation established between the CS and US. Participants who were told that the CSs caused the USs had higher IAT scores (*M* = 0.38, *SD* = 0.34, *d* = 1.12) than those who were informed that the CSs predicted (*M* = 0.24, *SD* = 0.39, *d* = 0.62), *t*(109) = 2.04, *p* = .02, *d* = 0.39, BF_1 = 2.47, or were unrelated to the valenced words (*M* = 0.20, *SD* = 0.44, d = 0.47), *t*(107) = 2.40, *p* = .01, *d* = 0.46, BF_1 = 5.09. Participants in the predictive relation condition did not have higher IAT scores than those in the unrelated relation condition, *t*(106) = 0.48, *p* = .31, *d* = 0.09, BF_0 = 3.27.

Confidence ratings. A mixed ANOVA revealed an effect of Stimulus, *F*(1, 177) = 15.46, *p* < .001, but no main or interaction effects of Relation, *F*(2, 177) = 0.07, *p* =.94. Overall, participants were somewhat confident in their ratings (*M* = 1.30, *SD* = 2.88) and more so for CS1 (*M* = 1.61, *SD* = 2.96) than for CS2 (*M* = 0.99, *SD* = 3.18).

Belief in the instructions. A between-subjects ANOVA revealed an effect of Relation, *F*(2, 175), = 9.43, *p* < .001, such that participants were more inclined to believe the instructions when they indicated that the CS was unrelated to the US (*M* = 7.12, *SD* = 2.31) than when they were told the CS predicted (*M* = 5.46, *SD* = 2.69) or caused the US (*M* = 5.21, *SD* = 2.77). Belief in the instructions was positively correlated with difference in evaluative ratings of CS1 and CS2 in the causal (*r* = 0.32, *p* = .01), and predictive conditions (*r* = 0.36, *p* = .005) and negatively correlated with those same ratings in the unrelated condition (*r* = –0.37, *p* = .004). Belief in the instructions only correlated (negatively) with

IAT scores in the predictive condition ($r = -0.27$, $p = .05$) but not in the causal ($r = 0.03$, $p = .83$) or unrelated conditions ($r = -0.13$, $p = .34$).

Reactance. A between-subjects ANOVA revealed a main effect of Relation, $F(2, 175) = 4.62$, $p = .011$, such that reactance was lower in the causal ($M = 3.84$, $SD = 2.83$) than predictive ($M = 4.72$, $SD = 2.76$) or unrelated ($M = 5.37$, $SD = 2.59$) conditions (although only the difference between the causal and unrelated condition reached significance, $p = .003$). Correlational analyses revealed that higher levels of reactance were associated with lower magnitude of the preference for CS1 over CS2 on evaluative ratings in the causal ($r = -0.43$, $p <.001$), predictive, ($r = -0.48$, $p < .001$) and unrelated conditions ($r = -0.30$, $p = .02$), and that this was also true for IAT scores (but only for those in the causal condition: $r = -0.32$, $p = .02$ and not the predicitive, $r = -0.08$, $p = .56$, or unrelated conditions, $r = -0.04$, $p = .78$).

General Discussion

A growing body of work reveals that relational qualifiers (i.e., cues that specify how other stimuli are related) can exert a powerful influence on interpersonal preferences formed via spatio-temporal pairings. We examined for the first time whether certain qualifiers moderate these EC effects to a greater extent than others. We first conducted a pilot study with evaluative statements and selected three that allow for a transfer of liking in a strong (causal), medium (predictive), and weak (unrelated) fashion. We then investigated if those same relations would also moderate the impact of stimulus pairings on person perceptions. Results were consistent with this hypothesis: causal relations gave rise to bigger changes in attitudes than predictive (on both explicit and implicit measures) and unrelated relations (on both explicit and implicit measures), whereas predictive relations produced larger effects than unrelated relations (on explicit but not implicit measures).

Our results not only demonstrate that relational qualifiers can moderate interpersonal preferences formed via pairings but also raise questions as to why certain types of relations are more conductive than others. Although we did not have *a priori* predictions about which relations would be the most conductive (before the pilot study), in hindsight, the observed differences do make sense. Let us start with the fact that causal relations allow for a bigger transfer of valence than predictive relations. This might be related to the fact that causality not only denotes that a person is related to valenced events but that the former is directly responsible for the latter. In other words, causality is a directional (cause→effect) relation that is highly reliable and unambiguous (the relation between events and responsibility for the valenced outcome is clear). It is also singular in the sense that only the person (and not someone else) is responsible for the occurrence of the valenced event. In contrast, prediction merely implies a temporal relation between a person and events and does not ensure that he/she is causally related to the valenced outcome. Although a predictive relation itself can also be reliable and unambiguous, the extent to which the person is causally responsible for the outcome is ambiguous (i.e., it could be the cause or merely a correlate). This additional ambiguity about the relation between object and event may be crucial in explaining the observed differential effect of causal and predictive relations.

We also obtained evidence for attitude change even when participants learned that a person was *unrelated* to valenced events. An explanation in terms of manipulation failure seems unlikely given that participants reported that they believed (and could remember) the instructions provided. Another possibility is that being told or learning that a person and valenced outcomes are unrelated is - in some sense - ambiguous because pairing those stimuli does relate them in some way. In other words, repeated co-occurences could provide the basis for the formation of either a proposition about the CS-US relation (i.e., 'Bob co-occurs with positive events') or an association between representations of those stimuli. From a propositional perspective, (some) participants might consider this experience-based proposition more valid than that formed on the basis of relational information (i.e. 'Bob is unrelated to positive events'), which may guide their downstream evaluative judgements. From a dual-process perspective, high frequency co-occurrences between stimuli might lead to strong mental assocations that over-ride propositions formed via instructions when these propositions are considered invalid.

This finding that 'unrelated' relations also result in changes in liking could underpin phenomena elsewhere in social psychology. Take the '*kill the messenger'* effect where the individual (CS) delivering a message (US) acquires the valence of that message even though he is unrelated to the content of the message he is carrying (see also Wegner, Wenzlaff, Kerker, & Beattie, 1981). Or research on prejudice and stereotyping which also suggests that persuasive arguments can backfire when they include negations. For instance, "thinking about stereotyped groups or individuals in counter-stereotypical terms (e.g., "*old people are good drivers*") is more effective in reducing unwanted stereotyping than attempts to negate an existing stereotype (e.g., "*it is not true that old people are bad drivers*") (Gawronski, Deutsch, Mbirkou, Seibt, & Strack, 2008, p.376). Hence, our results are consistent with past work showing that the denial of a relation creates a relation that allows for the formations of likes and dislikes.

Future Directions and Implications

The current work opens up several new directions for theorizing and research on EC and persuasion. It also provides concrete recommendations for optimizing attitude change in applied settings.

Theoretical implications

EC. Our findings raise questions about single-process associative accounts which argue that EC effects are mediated by the automatic formation and activation of associative links between a CS and US. If this were the case then explicit and implicit evaluations should have simply reflected the valence of the US a CS was repeated paired with, and not the additional information about how those stimuli are related. They also speak against dual-process accounts given that they predict a dissociation between explicit and implicit evaluations, such that the former should have mainly been driven by relational information concerning the CS-US relation whereas the latter should have been driven by the valence of the US a CS was repeatedly paired with (see Hu et al., 2017; Peters & Gawronski, 2011). We found no evidence of this and instead observed a clear impact of relational information on implicit and explicit evaluations (even for high frequency CS-US repetitions).

Our results are consistent with single-process propositional accounts arguing that EC effects are mediated by the formation and activation of propositions about the CS-US relation. Information about the CS-US relation moderated implicit and explicit evaluations in a manner that was consistent with the evaluative meaning of the relational information rather than the mere valence of the co-occurring US. Moreover, and in contrast to Hu et al. (2017) (but in-line with Zanon et al., 2014), we obtained such findings without the need to present relational information during the CS-US pairing phase. In the study of Hu et al., relational information only impacted implicit evaluations whenever it was presented on a trial-by-trial basis during EC and not when it was presented before the pairings. They argued that the "*qualifying effects of relational information on implicit evaluations depend on the mental integration of co-occurrence and relational information during the encoding of CS–US pairings*" (p.29). The current findings suggest that presenting relational information prior to stimulus pairings does not always lead to the independent processing of co-occurrence and relational information, and thus, ineffective integration of the two pieces of information during the encoding of CS–US pairings. Rather it can bias EC effects in a similar way to when such information is provided during the encoding stage.

Person perception. Our findings provide support to Heider's notion of affect-perception consistency (see Crandall, 2007), showing a graded correspondence between the newly formed attitude towards a target individual and the perception of relations (i.e., cause, predict, unrelated) between the individual and positive or negative events. These findings are also consistent with Heider's (1944, 1958, 1988) theory of attribution (also see Weiner, 1986, 1995), insofar as different relational qualifiers (i.e., cause vs. predict vs. unrelated) imply various degrees of attribution of responsibility of the events to the target individual. In our case, it may be that when the target individual *caused* the positive or negative events, it led to higher-degrees of attribution of responsibility than when the same individual merely predicted or was unrelated to those events. Such reasoning is in-line with Heider's (1958, pp. 10–11) proposal that different transitive verbs (e.g., 'give') can imply different types of causal and responsibility attributions (for related findings see Walther et al. 2011)[2].

Persuasion. Our findings are also relevant to research on persuasion. On the one hand, theories such as the Heuristic-Systematic Model (Chaiken, Liberman, & Eagly, 1989) and the Elaboration-Likelihood Model (e.g., Petty & Cacioppo, 1986) view the mental processes that mediate changes in liking due to pairings as primitive in nature. The former (HSM) would consider pairings as a heuristic cue (i.e., an environmental cue which elicits an information processing strategy based on simple rules, schemas, or prior knowledge) (e.g., pairings automatically elicit a rule such as "*stimuli which co-occur are similar in valence*"). The latter (ELM) typically relegates pairings to to the peripheral route of attitude change, and sees it as a potential input for, but rarely a type of, argument in itself (Petty & Brinol, 2014). On the other hand, the Unimodel of persuasion allows for "persuasive evidence to be presented in an

[2] The current work is also relevant to literature on spontaneous trait inferences (STI) and transference (STT). Some authors argue that STI and STT reflect two distinct cognitive processes (STT is underpinned by associations whereas STI are driven by attributions; Carlston & Skowronski, 2005). Others argue that both STI and STT are based on simple automatic associative processes (e.g., Brown & Bassili, 2002) (for a more detailed discussion, see Orghian, Garcia-Marques, Uleman, & Heinke, 2015). It could be that the nature of the relation between people, or between a person and event or object, also moderates STI and STT effects.

unlimited number of forms and variations", (Bohner, Erb, & Siebler, 2008; p.172) including forms that do not involve words or sentences, such as pairings (Kruglanski & Thompson, 1999). In short, certain theories of persuasion tend to view pairings as a simple heuristic cue whereas others allow for the possibility that pairings could function as a persuasive argument (for more see De Houwer & Hughes, 2016).

Relational EC effects, like those presented here, involve more than the mere pairing of stimuli. Participants are given extra verbal information about the pairings (e.g., relational qualifiers, instructions) and/or provided with the motivation and opportunity to think about the pairings in some way (e.g., make online judgements). This combination of verbal information and pairings (and/or requirement to effortfully think about them) may lead people to treat pairings, not as a heuristic cue, but rather as a 'regularity-based argument' (e.g., "*the CS causes Cancer*" is an argument for disliking the CS). If so, then (relational) EC and persuasion may have more in common than traditionally believed: both can involve changes in liking due to arguments but differ in the way that those arguments are delivered and constructed (e.g., either via stimulus pairings [EC] or words and sentences [persuasion]).

It is also likely that relational EC (at least in the studies conducted thus far) elicits systematic processing (HSM) (or creates the conditions where elaboration is high; ELM), such that people attempt to thoroughly understand what they have been told or seen via careful attention, thinking, and reasoning. For instance, imagine that people see "CS Loves US" onscreen during the EC phase. They may carefully consider why it is that the CS loves the US and whether this is a good or bad reason to like or dislike the CS itself. Repeated presentations of this information might also provide the opportunities to elaborate on the argument, while pre-task instructions likely increase their motivation to do so.

The idea of regularity-based arguments leads to several new empirical possibilities. First, it may be that pairings - even in the absence of extra verbal instructions or information

– can serve as an argument indicating that a CS is similar to a US, and as a result, leads to a change in liking. This is not to say that pairings always function in this regard: when motivation and ability to process pairings are low, it is likely that the mere fact that stimuli are paired does indeed function as a heuristic cue. However, increasing people's motivation and opportunities to effortfully operate on the pairings, may cause them to treat pairings as a simple, regularity-based, argument and this argument drives their evaluations (i.e., pairings may function as both heuristic cue and persuasive argument). Second, regularity-based (EC) and persuasive arguments may differ in their respective strength, such that merely stipulating that a regularity exists (e.g., 'CS co-occurs with a US') could lead to a weaker change in liking than when people are explicitly given information about the evaluative properties of a stimulus (e.g., 'the CS tastes good'). Third, one could manipulate the conditions known to influence ability and motivation to think to see if this influences the types of effects observed here and elsewhere in the relational EC literature (e.g., perceived relevance of the information, cognitive load when reading the instructions or during the EC phase). These questions seem like fruitful avenues for future investigation and would serve to link two areas in attitude research (persuasion and EC) that have rarely interacted with one another.

Future directions. First, it is important to realise that we only focused on three (causal, predictive, and unrelated) relations in our main study for pragmatic purposes (i.e., to demonstrate that relational qualifiers differ in their capactity to moderate EC effects). Yet our relational survey contains many more relations that can be used by researchers to pose theoretical and empirical questions about EC. For instance, recent work found that under certain conditions, co-occurrence information can exert a stronger impact than relational information (e.g., Moran et al., 2016). In addition to comparing the main effects of co-occurrences and relational information, it would be interesting to examine how both variables interact, that is, the way in which the impact of co-occurrences depends on the nature of the relational qualifier[3].

Second, it has been argued that people also form propositonal beliefs about the way in which a CS and US are related even in situations where a CS and US are paired in the absence of additional relational information (i.e., as is the case in traditional EC studies; De Houwer, 2009, Mitchell et al., 2009). Our research highlights the possibility that subtle procedural parameters in EC studies (e.g., the particular stimuli used, as well as the ordering, timing, and nature of stimulus presentations) could influence beliefs about *how* the CS and US are related. It would thus be interesting to examine in future research how people spontaneously construct beliefs about CS-US relations, that is, even when explicit relational information is absent). Doing so could provide better insight into how people symbolically respond to regularities (such as stimulus pairings) in their environment (De Houwer & Hughes, 2016).

One could also examine which aspects of the EC procedure influence beliefs about how CS and US are related. For instance, consider Unkelbach and Fiedler (2016) who recently found that by forcing people to make judgments during the EC phase (e.g., asking them whether they liked the CS or US more) they could alter the way in which those stimuli were being related (CS is better than US) without the need to introduce specific words or instructions (e.g., '*better or worse than*'). Such a manipulation could be interfaced with the relations identified in this paper to examine whether stimulus relations can be altered in more subtle ways, and whether the same graded pattern of evaluations would emerge if judgments

[3] In our pilot survey causal, predictive, and unrelated relations produced strong, medium, and weak positive evaluations. Yet causal produced strong whereas predictive and unrelated relations produced weak changes in negative evaluations. Thus it seems that the impact of relational information on evaluations due to verbal statements may depend on the nature of the valenced events. Future work could explore whether this is also true for the impact of relational information on evaluations produced via regularities in the environment as well.

led the individual to view the object as a cause, predictor, and as unrelated to valenced events.

The current research might be the first to examine the modulating effects of attribution of responsibility on EC effects. Although our work is limited to the context of person perception, it is possible that attribution plays a general role in EC, regardless of whether the CS is a person or a non-human stimulus (e.g., a neutral symbol). In the latter case, individuals who are exposed to CS-US pairings can still make attributions about the extent to which the CSs are responsible for the occurrence of USs. In Heider's (1958) words, this type of causal attribution will lead to the perception that CS and US belong together, forming a unit relation. Such relation could be the key mediator for changes in liking or disliking of the CSs. Future research should further examine the potential role of causal attribution in EC. Finally, processing models have recently emerged that might help parse out the distinct processes that contribute to EC effects (e.g., Hütter, Sweldens, Stahl, Unkelbach, & Klauer, 2012). Future work could use these models to determine to what extent the effect of relational qualifiers on liking is driven by associative and/or propositional processes, and in so doing, contribute to the ongoing debate between single-process (propositional) and dual-process perspectives on EC (for recent work in this vein see Hütter & De Houwer, 2017). Other work could introduce manipulations that differentially influence propositional and associative processes, and determine if such manipulations impact the types of findings reported here. In the current paper we created one such condition (high frequency CS-US co-occurrences) that should have led to a stronger impact of associative than propositional processes on implicit evaluations according to dominant dual-process theories of evaluation (for reasons why, see Hu et al., 2016). Under this condition, however, we found evidence that implicit evaluations were still dependent on the type of relation between stimuli, which argues against such models. Future work could explore other conditions that might parse between different evaluation theories. For instance, as one reviewer pointed out, the current study and nearly every other study examining the impact of relational qualifiers on EC effects have provided participants with time, resources, and the goal of attending to that relational information (i.e., optimised the conditions under which propositional learning is often assumed to operate). Thus the question remains: would the impact of relational information on EC effects still be evident when these conditions were undermined or removed (e.g., where participants are unaware of the pairings, or relational qualifiers were presented under cognitive load, or in an incidental fashion, as seems to be the case in the Olson & Fazio, 2001, paradigm)? It could be that relational information only moderates EC effects when the individual has the motivation and opportunity to process that information during the *learning* phase (note a moderation effect might still be evident even if there is no motivation or opportunity during the *retrieval* or *expression* phases). When replicating this work, researchers could ask participants how they construed the relationship between paired stimuli. Doing so would provide another check to ensure that our manipulation was successful (one that does not depend on the evaluative measures), and additional evidence that the propositional relation between stimuli mattered.

Applied implications. Finally our findings have implications for applied research on EC. So far, applied EC researchers have mainly relied on *unqualified* stimulus pairings as a means to promote vegetable consumption in children (e.g., Havermans & Jansen, 2007), as well as increased condom use (Ellis, Homish, Parks, Collins, & Kiviniemi, 2015), or exercising behaviors in college students (Antoniewicz & Brand, 2016). Although certainly effective, the addition of relational qualifiers that push evaluations in the intended direction (positive or negative valence) and that boost the magnitude of those evaluations might improve the aforementioned interventions. The current findings suggest that applied researchers looking to augment the effectiveness of their interventions may do so through the addition of specific relational qualifiers that direct the propositions that people form about the CS-US relation. Although those same researchers could simply inform people about the CS-US relation in the absence of stimulus pairings, there may be reasons why an EC approach might supersede verbal instructions alone. To illustrate why, imagine that a researcher wants to change evaluations of a novel individual and tells participants that "*Bob is a good guy*". In some cases this may lead to reactance given that people feel coerced into liking Bob by an outside force. Instead, simply pairing Bob with positive images might influence their preferences in subtle, even irrational ways without evoking reactance (De Houwer & Hughes, 2016). Our research suggests that researchers can boost the effects of pairings by using verbal or non-verbal elements that imply a specific type of relation between the stimuli. In other words, by learning to shape the relational beliefs that people form on the basis of stimulus pairings we have new tools to influence preferences in subtle but precise ways. Finally, much work in social and personality psychology creates relations between verbs and people (e.g., in vignettes) in order influence judgements and evaluations of individuals and groups. The current work (particularly the relational survey presented in the Supplementary Materials) highlights many relations that could be fed into, and refine, such protocols.

Constraints on Generality. The current findings were obtained using an online sample of participants recruited from a website where site-wide demographic information indicates that the population is to some extent WEIRD (Western, Educated, Industrialized, Rich, and Democratic; Henrich, Heine, & Norenzayan, 2010). That said, given the relatively simple nature of the manipulation (instructions followed by stimulus pairings) we expect that these findings should also be replicable in the very same samples used in most evaluative learning research (e.g., student samples recruited in a laboratory context). Although antecdotal, we believe that the success of our relational manipulation is likely linked to the manipulation check implemented prior to (and during) the EC phase that ensured that the relational information was both salient and task relevant for participants. We have no reason to believe that the results depend on other characteristics of the participants, materials, or context.

Conclusion

Taken together, it seems that not all qualifiers are created equal: causal relations influence interpersonal preferences to a greater extent than predictive relations, and predictive relations lead to stronger effects than unrelated relations. This finding gives rise to a more nuanced perspective on the role of relational information in EC, further supporting the idea that it is not the mere fact that stimuli are paired, but how they are paired, that determines implicit and explicit evaluations. The current results extend our understanding of relational information on EC beyond the contrastive relations explored to date, and provide data on a wealth of relations that have yet to be explored. They provide further support for the role of propositional processes in EC, and introduce two new ideas: that attributional processes might play a role in EC, and that stimulus pairings combined with verbal information could represent ‘regularity-based arguments’ that are in some respects similar to persuasive arguments. Thus researchers interested in using qualifiers, whether for theoretical or applied reasons, need not only think about the presence or absence of such qualifiers but also the nature of qualifier they use.

References

Antoniewicz, F., & Brand, R. (2016). Learning to like exercising: evaluative conditioning changes automatic evaluations of exercising and influences subsequent exercising behavior. *Journal of Sport and Exercise Psychology*, *38*(2), 138–148.

Baeyens, F., Eelen, P., & Crombez, G. (1995). Pavlovian associations are forever: On classical conditioning and extinction. *Journal of Psychophysiology*, *9(2),* 127-141.

Bohner, G., Erb, H. P., & Siebler, F. (2008). Information processing approaches to persuasion. Integrating assumptions from the dual-and single-processing perspectives. *Attitudes and attitude change*, 161-188.

Brown, R. D., & Bassili, J. N. (2002). Spontaneous trait associations and the case of the superstitious banana. *Journal of Experimental Social Psychology*, *38*(1), 87-92.

Carlston, D. E., & Skowronski, J. J. (2005). Linking versus thinking: evidence for the different associative and attributional bases of spontaneous trait transference and spontaneous trait inference. *Journal of personality and social psychology*, *89*(6), 884-898.

Crandall, C. S., Silvia, P. J., N'Gbala, A., Tsang, J., & Dawson, K. (2007). Balance theory, unit relations, and attribution: The underlying integrity of Heiderian theory. *Review of General Psychology*, *11*, 12-30.

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

De Houwer, J. (2014). A propositional model of implicit evaluation. *Social and Personality Psychology Compass*, *8*(7), 342–353.

De Houwer, J., & Hughes, S. (2016). Evaluative conditioning as a symbolic phenomenon: On the relation between evaluative conditioning, evaluative conditioning via instructions, and persuasion. *Social Cognition*, *34*(5), 480–494.

Ellis, E. M., Homish, G. G., Parks, K. A., Collins, R. L., & Kiviniemi, M. T. (2015). Increasing condom use by changing people's feelings about them: An experimental study. *Health Psychology*, *34*(9), 941-950.

Fiedler, K., & Unkelbach, C. (2011). Evaluative conditioning depends on higher order encoding processes. *Cognition and Emotion*, *25*(4), 639–656.

Fincham, F. D., & Jaspars, J. M. (1980). Attribution of responsibility: From man the scientist to man as lawyer. In L. Berkowitz (Ed.), *Advances in experimental social psychology (Vol. 13, pp. 81-138)*. New York: Academic Press.

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

Gawronski, B., & Bodenhausen, G. V. (2011). The associativepropositional evaluation model: Theory, evidence, and open questions. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 44, pp. 59–127). Amsterdam, The Netherlands: Elsevier.

Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., & Strack, F. (2008). When "just say no" is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of experimental social psychology, 44(2)*, 370-377.

Gawronski, B., & Payne, B. K. (Eds.). (2010). *Handbook of implicit social cognition: Measuremmt, theory, and applications*. New York: Guilford Press.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology, 74(6)*, 1464-1480.

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.

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, *90*(1), 1-20.

Havermans, R. C., & Jansen, A. (2007). Increasing children's liking of vegetables through flavour–flavour learning. *Appetite*, *48*(2), 259–262.

Heider, F. (1944). Social perception and phenomenal causality. *Psychological Review*, *6*, 358–374.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: Wiley.

Heider, F. (1988). *The notebooks of Fritz Heider: Vol. 4. Balance theory* (M. Benesh-Weiner, Ed.). Munchen-Weinheim, Germany: Psycologie Verlags Union.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. Nature, 466(7302), 29.

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.

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,* 17-32.

Hughes, S., De Houwer, J., & Barnes-Holmes, D. (2016). The Moderating Impact of Distal Regularities on the Effect of Stimulus Pairings. *Experimental Psychology*, *63,* 20-44.

Hughes, S., Ye, Y., & De Houwer, J. (in prep). *Evaluative Conditioning Effects Are Modulated By The Nature of Contextual Pairings*.

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

Hütter, M., & Fiedler, K. (2016). *Editorial: Conceptual, Theoretical, and Methodological Challenges in Evaluative Conditioning Research*. *Social Cognition*, *34(5),* 343–356.

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

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: the attribution proces in social psychology. In L. Berkowitz (ed.), *Advances in experimental social psychology* (Volume 2, pp. 219-266). New York: Academic Press.

Jones, C. R., Fazio, R. H., & Olson, M. A. (2009). Implicit misattribution as a mechanism underlying evaluative conditioning. *Journal of Personality and Social Psychology*, *96*(5), 933-948.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (ed.), *Nebraska Symposium on Motivation (*Volume 15, pp. 192-238). Lincoln: University of Nebraska Press.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J.-E. L., Joy-Gaba, J. A., … others. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*(4), 1765-1785.

Mierop, A., Hütter, M., Stahl, C., & Corneille, O. (2018). Does attitude acquisition in evaluative conditioning without explicit CS-US memory reflect implicit misattribution of affect?. *Cognition and Emotion*, 1-12.

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 630–633.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32(02),* 183-198.

Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., Van Schie, K., Van Harmelen, A.-L., … Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods*, *45*(1), 169–177.

Moran, T., & Bar-Anan, Y. (2013). The effect of object–valence relations on automatic evaluation. *Cognition & Emotion*, *27*(4), 743–752.

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.

Orghian, D., Garcia-Marques, L., Uleman, J. S., & Heinke, D. (2015). A connectionist model of spontaneous trait inference and spontaneous trait transference: Do they have the same underlying processes?. *Social Cognition*, *33*(1), 20-66.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science*, *12*(5), 413-417.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: affect misattribution as implicit measurement. *Journal of personality and social psychology, 89(3)*, 277-293.

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.

Petty, R. E., & Brinol, P. (2014). The elaboration likelihood and metacognitive models of attitudes. *Dual-process theories of the social mind*, 172-187.

Tormala, Z.L. and Briñol, P. (2015) Attitude change and persuasion: past, present and future directions. In *The Cambridge Handbook of Consumer Psychology* (Norton, M.I. et al., eds), pp. 29–64, Cambridge University Press.

Unkelbach, C., & Fiedler, K. (2016). Contrastive CS-US relations reverse evaluative conditioning effects. *Social Cognition*, *34*(5), 413–434.

Walther, E., Gawronski, B., Blank, H., & Langer, T. (2009). Changing likes and dislikes through the back door: The US-revaluation effect. *Cognition and Emotion, 23(5),* 889-917.

Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer.

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York: Guilford Press.

Zanon, R., De Houwer, J., & Gast, A. (2012). Context effects in evaluative conditioning of implicit evaluations. *Learning and Motivation*, *43*(3), 155–165.

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.

SUPPLEMENTARY MATERIALS

Pilot Study

Method

Participants and design. Two hundred and thirty-two English-speaking volunteers participated online via the Prolific Academic website (https://prolific.ac) in exchange for a small monetary reward (€0.75). Nineteen were excluded because they failed to complete the survey. One was excluded for failing to meet the minimum age criterion (17 or older). This left a final sample of two hundred and twelve participants (110 men, 102 women) ranging from 17 to 72 years (M = 32.2, SD = 11.1).[4]

Materials. The survey consisted of 68 statements that each contained three basic elements: a generic object (always referred to as 'the object'), a relational qualifier (e.g., 'causes'), and a generic subject (always referred to as "positive events" or "negative events"). There were thirty-four different types of relational qualifiers each of which was paired once with 'positive events' and 'negative events' (see Appendix). The list of relations that comprised the survey were generated through a brain-storming session following a literature search for relational qualifiers that have previously been shown to influence evaluation.

Procedure

Relational survey. Participants were informed that they would encounter statements about objects and valenced events. Their job was to indicate how much they liked or disliked the object in each statement. Following these instructions they then completed the survey. They rated all 68 statements one after the other and were asked to treat each statement as independent from the others. Four statements were presented onscreen at any given time – one above the other. The presentation order of statements was randomized across

[4] Note that the study designs and data-analysis plans for this pilot experiment are available on the Open Science Framework website (https://osf.io/me9dh/). We report all manipulations and measures used in the study. All data were collected without intermittent data analysis. The data analytic plan, experimental scripts, and data are available at the above link.

participants, with the exception that the statement 'the object is the same as positive/negative events' always appeared at the beginning of the survey.[5]

Self-reported ratings. After reading a statement, participants indicated how much they liked or disliked the object using a scale ranging from 1 (*I dislike it extremely*) to 9 (*I like it extremely*) with 5 (*I neither like nor dislike it*) as a neutral mid-point.

Results

Submitting mean scores from the 34 statements that produced positive evaluations (see Table 1) to a repeated-measures Analysis of Variance (ANOVA) (with Relation Type as a within-subjects factor) revealed a main effect for Relation Type, $F(33, 211) = 74.46$, $p < .001$, $\eta^2_p = .26$. Likewise, submitting scores from statements that produced negative evaluations (see Table 2) to the ANOVA also revealed a main effect of Relation Type, $F(32, 211) = 91.79$, $p < .001$, $\eta^2_p = .30$. Inspection of Tables 1 and 2 indicates that the vast majority of statements gave rise to an explicit evaluation that significantly differed from neutral ($p < .05$), and that the magnitude of those ratings differed as a function of type of relation involved. Descriptively, the most extreme evaluations (both positive and negative) were those in which the source *caused*, *owned*, or was *hierarchically related* to the valenced event in some way. The least extreme evaluations (both positive and negative) were those in which the source was *physically related*, *conditionally related*, or *unrelated* to the valenced event.

Discussion

The specific way in which an object and valenced events are related, influences both the direction and magnitude of object evaluations. The results are interesting as such in that they document important differences in how relational qualifiers moderate the impact of statements on changes in liking. Within the context of the present research, however, we only

[5] We adopted this strategy in the hope that equivalence relations ("*Object Same Valence Event*") might serve as a 'baseline' relation against which participants might calibrate how they respond to all other relations in the survey.

utilized these data to identify three relations that moderated the transfer of valence in a strong (causal, ‘causes’), medium (predictive, ‘predicts’), or weak (unrelated, ‘is unrelated to’) way and used these relations in our main study.

Table 1. Means, standard deviations, *t*-test scores, and effect sizes for those CS-US relations from the relational survey that gave rise to positive evaluations. Items are rank ordered in terms of their respective effect sizes (with scores independently compared to the neutral [5] mid-point of the scale).

Statement	Positive Evaluation			
	M	*SD*	*T*	*d*
1. The object causes positive events.	7.83	1.22	$t(211) = 33.77$	2.32
2. I own the object and it is related to positive events.	7.67	1.23	$t(211) = 31.72$	2.18
3. The object is related to positive events.	7.01	0.93	$t(211) = 31.67$	2.16
4. The object is a part of positive events.	7.73	1.31	$t(211) = 30.77$	2.11
5. The object starts positive events.	7.57	1.23	$t(211) = 30.43$	2.09
6. The object stops negative events.	7.73	1.31	$t(211) = 30.37$	2.09
7. The object is more positive than other positive events.	7.76	1.39	$t(211) = 28.73$	1.97
8. The object is a member of positive events.	7.13	1.12	$t(211) = 27.74$	1.91
9. The object is a type of positive events.	7.30	1.25	$t(211) = 26.77$	1.84
10. The object, and only this object, predicts positive events.	7.33	1.30	$t(211) = 26.09$	1.79
11. The object prevents negative events from happening.	7.61	1.53	$t(211) = 24.81$	1.70
12. The object loves positive events.	7.29	1.38	$t(211) = 24.15$	1.66
13. The object is an effect of positive events.	6.98	1.27	$t(211) = 22.66$	1.56
14. The object is similar to positive events.	6.71	1.10	$t(211) = 22.58$	1.55
15. The object predicts positive events.	6.93	1.28	$t(211) = 22.00$	1.51
16. The object is the opposite of negative events.	7.16	1.48	$t(211) = 21.25$	1.46
17. The object refers to positive events.	6.68	1.16	$t(211) = 21.14$	1.45
18. The object is the same as positive events.	7.14	1.48	$t(211) = 21.00$	1.44
19. The object occurs before positive events.	6.55	1.18	$t(211) = 19.13$	1.31
20. The object (physically) approaches positive events.	6.60	1.23	$t(211) = 18.94$	1.30
21. The object was previously not related to positive events but now is.	6.54	1.19	$t(211) = 18.75$	1.29
22. The object is just one of many objects that predict positive events.	6.56	1.22	$t(211) = 18.57$	1.28
23. Others own the object and it is related to positive events.	6.64	1.31	$t(211) = 18.22$	1.25
24. The object is physically close to positive events.	6.57	1.30	$t(211) = 17.56$	1.21
25. The object (physically) avoids negative events.	6.61	1.39	$t(211) = 16.93$	1.16
26. The object occurs after positive events.	6.49	1.29	$t(211) = 16.68$	1.15
27. The object occurs at the same time as positive events.	6.48	1.38	$t(211) = 15.68$	1.08
28. The object is related to positive events only under some, but not under other, conditions.	5.88	1.00	$t(211) = 12.78$	0.88
29. The object is physically far away from negative events.	6.22	1.53	$t(211) = 11.56$	0.79
30. The object loathes negative events.	6.30	1.72	$t(211) = 10.96$	0.75
31. The object is not related to negative events.	5.93	1.29	$t(211) = 10.48$	0.72
32. The object is different from negative events.	5.69	1.17	$t(211) = 8.67$	0.59
33. The object is less positive than other positive events.	5.67	1.21	$t(211) = 8.09$	0.56
34. The object was previously related to negative events but now is not.	5.41	1.19	$t(211) = 4.96$	0.34

Table 2. Means, standard deviations, *t*-test scores, and effect sizes for those CS-US relations from the relational survey that gave rise to negative evaluations. Items are rank ordered in terms of their respective effect sizes (with scores independently compared to the neutral [5] mid-point of the scale).

Statement	Negative Evaluation			
	M	*SD*	*t*	*d*
1. The object causes negative events.	2.15	1.46	$t(211) = 28.38$	1.95
2. The object is more negative than other negative events.	2.13	1.58	$t(211) = 26.44$	1.82
3. The object is a type of negative events.	2.68	1.33	$t(211) = 25.40$	1.75
4. The object stops positive events.	2.26	1.58	$t(211) = 25.30$	1.74
5. The object is a member of negative events.	2.76	1.29	$t(211) = 25.13$	1.73
6. The object starts negative events.	2.32	1.57	$t(211) = 24.85$	1.71
7. The object prevents positive events from happening.	2.36	1.65	$t(211) = 23.22$	1.59
8. The object is the same as negative events.	2.81	1.49	$t(211) = 21.39$	1.47
9. The object is the opposite of positive events.	2.75	1.61	$t(211) = 20.37$	1.39
10. The object is similar to negative events.	3.29	1.22	$t(211) = 20.32$	1.39
11. The object loves negative events.	2.54	1.79	$t(211) = 20.01$	1.37
12. The object is a part of negative events.	3.01	1.46	$t(211) = 19.85$	1.36
13. The object is related to negative events.	3.17	1.43	$t(211) = 18.69$	1.28
14. The object loathes positive events.	2.75	1.77	$t(211) = 18.44$	1.27
15. Others own the object and it is related to negative events.	3.09	1.59	$t(211) = 18.29$	1.26
16. The object is an effect of negative events.	3.31	1.36	$t(211) = 18.11$	1.24
17. I own the object and it is related to negative events.	3.09	1.59	$t(211) = 17.46$	1.19
18. The object was previously not related to negative events but now is.	3.37	1.42	$t(211) = 16.65$	1.14
19. The object is physically close to negative events.	3.42	1.42	$t(211) = 16.15$	1.11
20. The object refers to negative events.	3.63	1.42	$t(211) = 14.07$	0.97
21. The object occurs after negative events.	3.73	1.45	$t(211) = 12.71$	0.87
22. The object (physically) approaches negative events.	3.74	1.51	$t(211) = 12.16$	0.84
23. The object occurs before negative events.	3.69	1.57	$t(211) = 12.15$	0.84
24. The object occurs at the same time as negative events.	3.70	1.61	$t(211) = 11.72$	0.81
25. The object (physically) avoids positive events.	3.60	1.75	$t(211) = 11.61$	0.79
26. The object is related to negative events only under some, but not other, conditions.	4.24	1.10	$t(211) = 10.02$	0.69
27. The object is physically far away from positive events.	4.04	1.55	$t(211) = 9.07$	0.62
28. The object is less negative than other negative events.	4.29	1.16	$t(211) = 8.98$	0.62
29. The object is not related to positive events.	4.42	1.37	$t(211) = 6.24$	0.43
30. The object was previously related to positive events but now is not.	4.46	1.46	$t(211) = 5.36$	0.37
31. The object is different from positive events.	4.65	1.34	$t(211) = 3.86$	0.27
32. The object is just one of many objects that predict negative events.	4.90	1.73	$t(211) = 0.84$	0.06
33. The object, and only this object, predicts negative events.	4.94	2.23	$t(211) = 0.37$	0.03