On the Symbolic Generalization of Likes and Dislikes

Sean Hughes, Dermot Barnes-Holmes, and Pieter Van Dessel,

Ghent University

João Henrique de Almeida

Universidade Federal De Sao Carlos

Ian Stewart

National University of Ireland Galway

Jan De Houwer

Ghent University

Authors Note

SH, DBH, JDH, and PVD, Department of Experimental Clinical and Health Psychology, Ghent University. JHA, Universidade Federal De Sao Carlos, Brazil. IS, National University of Ireland Galway. This research was conducted with the support of a postgraduate scholarship from the Irish Research Council for Science, Engineering and Technology (IRCSET) to the first author. The manuscript was prepared with the support of Methusalem Grant BOF16/MET_V/002 to JDH, an Odysseus I grant of the Research Foundation – Flanders (FWO) to DBH, an FWO postdoctoral grant to PVD, and grant FAPESP 2014/01874-7 to JHA. Correspondence concerning this article should be sent to sean.hughes@ugent.be.

Abstract

Evaluative generalization refers to the fact that when evaluative responses towards a (focal) stimulus are established or changed, people change how they respond to non-focal stimuli as well. Whereas evaluative generalization between perceptually similar stimuli has been firmly established, the available evidence for symbolic evaluative generalization is less conclusive and limited to one possible type of relation (i.e., similarity). In this paper we offer a new set of procedures that can be used to systematically investigate symbolic evaluative generalization effects. We use these procedures to showcase how evaluative responses towards a focal stimulus can propagate to other stimuli when they are related on the basis of symbolic similarity, opposition, or comparison. These effects were evident when self-report, implicit, approach-avoidance, and behavioral choice measures were employed. Implications for theories of evaluative generalization are discussed and future directions outlined.

Keywords: Attitudes, Evaluative Learning, Generalization, Functional-Cognitive

On the Symbolic Generalization of Likes and Dislikes

Evaluation is at the core of our psychological lives. It not only guides our judgments and decisions, but often dictates how we treat our friends and family, as well as other individuals and groups. Evaluations bias what we remember, influence the politicians we vote for, musicians we listen to, and products we consume. We therefore need to understand how, when, and why evaluations are established and what factors play a role in their change.

One such factor - the *generalization* of evaluations - may explain why evaluative learning can exert such a powerful and far-reaching influence on behavior. Evaluative generalization refers to the fact that once evaluative responses towards a focal stimulus are established or changed, people often emit similar responses to non-focal stimuli that are related to that focal object. Unlike (evaluative) conditioning effects (see Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010), these changes in liking are not due to the mere spatio-temporal relation between focal and non-focal objects but rather to a different type of relationship. Most research on evaluative generalization can be divided into one of two categories: perceptual or symbolic.[1]

Evaluative Generalization Along Perceptual Dimensions

The generalization of evaluative responses is often based on the fact that focal and non-focal stimuli share *perceptual* properties with one another. This type of generalization plays a role in many social, cognitive, and clinical phenomena. Take the resemblance effect in social psychology: evaluations of strangers are often influenced by how much they physically resemble people we already know. Such generalizations tend to occur automatically without the perceiver's awareness or intent, increase in strength as the level of similarity between faces grows, and influence our judgments and decision making (see Gawronski & Quinn, 2013; Verosky & Todorov, 2010; Zebrowitz, White, & Wieneke, 2008).

[1] When we refer to a focal stimulus we are referring to a stimulus whose evaluative properties influence the evaluative properties of non-focal stimuli. In other words, the focal stimulus is the source of the change in valence and the non-focal stimuli are the target of the change in valence.

Now consider the 'guilt-by-association' effect: evaluations of one individual can generalize to an entire group whenever they share physical properties such as age, ethnicity, or gender (Hütter, Kutzner, & Fiedler, 2014). The very same goes for novel consumer products: a mere physical similarity between a known (valenced) object and an unknown (neutral) object is often enough for valence to transfer from one to the other (e.g., Fazio, Eiser, & Shook, 2004). In clinical psychology, we have known since Watson and Rayner (1920) that conditioned fears towards an aversive stimulus (e.g., white rat) readily generalize to perceptually related (e.g., white and furry) stimuli. Recent work indicates that this is also true for anxiety and chronic pain (e.g., Lissek et al., 2014; Meulders & Vlaeyen, 2013).

Evaluative Generalization Along Symbolic Dimensions

Evaluative responses can also generalize when focal and non-focal stimuli are *symbolically* related. Relatively early on in their development, humans acquire the ability to generate, use, and respond to symbols in the world around them (Deacon, 1997). Whereas perceptual generalization effects are - by definition - heavily dependent on the degree of physical overlap between stimuli, generalizations based on symbolic relations are often free from any such constraints. These relations allow stimuli to be connected in a near infinite number of ways and for the psychological properties of one stimulus to influence how people behave towards others. For instance, imagine you are told that a brown sticky substance is poisonous and afterwards learn that this substance is similar to a green liquid, and that the liquid is similar to a white gas (Brown Solid-*Similar*-Green Liquid-*Similar*-White Gas). You may generalize what you have learned about the first stimulus (solid) to the last stimulus (gas), even though they share no physical properties.

Many phenomena in psychological science may qualify as symbolic evaluative generalization effects. Person perception is a prime example: personality traits that define one individual often influence stereotypes and evaluations of others even when those individuals

share no physical similarity (e.g., Skowronski, Carlston, Mae, & Crawford, 1998). In marketing, prior evaluations of one consumer product (e.g., Sony television) frequently bias evaluations of other products that are released under the same brand name (e.g., Sony headphones) (i.e., the 'brand extension' effect; Ratliff et al., 2012; Völckner & Sattler, 2006). This can occur despite the fact that the products are physically dissimilar and the individual has no prior experience with the novel items released under that brand name. Hence the symbolic relation between the stimuli (i.e., the fact that they are both 'Sony' products) may account for this finding.

The ability to symbolically relate stimuli drastically expands the remit of evaluative generalization and enables humans to transfer what they have learned about the evaluative properties of one stimulus to another even when they are physically unrelated. Unlike perceptual generalization effects (where physical overlap is crucial), symbolic relations can be established in many different ways.

A Systematic Investigation of Symbolic Evaluative Generalization

Despite the importance of symbolic evaluative generalization, there has been little systematic research on this topic, both at the conceptual and empirical levels. At the conceptual level, we put forward the proposal that symbolic evaluative generalization has occurred if three conditions are met: (1) a change in liking took place, (2) that is an instance of generalization, and (3) that is due to a symbolic relation between stimuli. Hence, symbolic generalization excludes (1) changes in non-evaluative properties (e.g., arousal) or specific emotional responses (e.g., fear), (2) changes in liking that are instances of conditioning (i.e., due to a direct or indirect relation between the spatio-temporal properties of the focal and

non-focal stimuli)[2], and (3) changes in liking that are due to perceptual relations between focal and non-focal stimuli.

Importantly, at the empirical level, virtually all studies on symbolic evaluative generalization fail to meet one or more of these criteria. For instance, some studies fail to control for perceptual overlap between stimuli (e.g., the focal [Reemo*lap*] and non-focal items [Bosaa*lap*] often share the same name-ending; Ranganath & Nosek, 2008). Others fail to exclude an impact of direct or indirect spatio-temporal relations (e.g., Ratliff et al., 2012) whereas still others focused on specific emotions rather than liking (e.g., Bennett, Meulders, Baeyens, & Vlaeyen, 2015). Past work has also tended to focus on a single type of symbolic relation (i.e., similarity). Generalization is usually based on the fact that one brand, individual, or group is *similar* to another. Yet stimuli can be related in many other (non-similarity based) ways which determine both the direction and magnitude of the generalization effect. Given the profound impact that relation type has on changes in liking for focal stimuli (e.g., Unkelbach & Fiedler, 2016; Zanon, De Houwer, Gast, & Smith, 2014), it is somewhat surprising to see that this factor has not yet been taken into account. In short, although symbolic evaluative generalization seems to play a role in many different phenomena, previous research on this topic is limited in several ways.

The Current Research

This lack of systematic research might (in part) be due to a lack of procedures to study this phenomenon, most prominently, procedures that control for the impact of perceptual relations and direct pairings. In this paper, we introduce a set of procedures that are designed to establish symbolic evaluative generalization. Our procedures not only allow one to

[2] Note that conditioning can be due to a direct or indirect relation between the spatio-temporal properties of stimuli. A direct spatio-temporal relation takes into account only the spatio-temporal properties of the two related stimuli whereas an indirect spatio-temporal relation also takes into account the spatio-temporal properties of other stimuli. For instance, stimuli that are repeatedly paired on a screen are directly related whereas stimuli that never co-occur but do co-occur with a common third stimulus are indirectly related in a spatio-temporal manner.

minimize the impact of perceptual similarity and direct pairings but also provide a way of manipulating how stimuli are related. Using these procedures, we conducted a series of studies that together provide the first systematic investigation of symbolic evaluative generalization effects.

Across six experiments we examined the impact of symbolic similarity, opposition, and comparative relations on explicit and implicit evaluations. All studies followed the same basic format which we will briefly preview here. We first established two symbols as contextual cues meaning ‘Same’ and ‘Opposite’ (Experiments 1-4) or ‘More than’ and ‘Less than’ (Experiments 5-6). During a subsequent training phase, we presented these cues onscreen along with two other stimuli. By reinforcing the selection of a certain cue in the presence of specific stimuli we set out to achieve two outcomes: (a) establish a positive or negative valence for a focal stimulus and (b) relate this focal stimulus to other non-focal group members (Experiments 1-2), fictitious brand products (Experiment 3-4), or potential prizes (Experiment 5-6) (see Figures 1 and 2). We then indexed evaluative responding using self-report, indirect (Implicit Association Task [IAT; Greenwald, McGhee, & Schwartz, 1998], Implicit Relational Assessment Procedure [IRAP; Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010], evaluative priming), approach-avoidance, and behavior choice tasks. Using a variety of indices allowed us to test the generality and robustness of our findings. Taken together, our studies provide new information on how the type of relationship between stimuli can moderate the direction and/or magnitude of generalization effects while controlling for perceptual overlap as well as direct or indirect pairings.

Experiments 1-4

Experiments 1-4 set out to model symbolic evaluative generalization in the context of social psychology (groups of novel stimuli [Pokémon]: Experiments 1 and 2) and consumer

psychology (Brands: Experiment 3-4). [3] We first established two symbolic similarity relations between focal and non-focal stimuli (Relationship 1: Pokémon 1-2-3, Relationship 2: Pokémon 4-5-6*).* Thereafter we established a valence for the two focal stimuli and tested for evaluative generalization using self-report and implicit measures (Experiment 1: IAT; Experiment 2: IRAP; Experiment 3: IAT and an evaluative priming task; Experiment 4: IAT and an approach-avoidance task).[4]

If symbolic generalization occurs then two outcomes should be observed. First, a valence established for a focal stimulus (e.g., Pokémon 1) should generalize to non-focal stimuli (e.g., Pokémon 2 and 3) despite the fact that the latter share no physical resemblance to Pokémon 1, nor were they ever paired with valenced images in the past. Second, the direction of evaluative change should depend on the symbolic relationship established between the focal (Pokémon 1) and valenced stimuli, as well as the focal and non-focal stimuli. If Pokémon 1 is opposite to positive images, then it should be, along with Pokémon 2 and 3, evaluated negatively. The reverse should be true for Pokémon 4, 5, and 6. Importantly, by manipulating the nature of the symbolic relation we demonstrate that these effects cannot be due to mere stimulus pairings. If the observed changes in liking were merely due to the pairing of stimuli then we would expect an assimilative effect, such that the non-focal stimuli should always acquire the same valence as the valenced stimuli that the focal stimuli were originally paired with. A mere pairings account would be hard-pressed to explain how the

[3] We employed Pokémon characters in our studies for several reasons. They have already been used in social psychology research as a means to successfully study attitude formation via conditioning (e.g., Olson & Fazio, 2001). During piloting it was also relatively easier to select a series of Pokémon that were neutral in valence compared to other stimulus sets (e.g., human faces). These stimuli were also perceptually different to one another which limits the possibility of perceptual generalization as an alternative explanation for our findings.

[4] As outlined above we established symbolic relations using contextual cues that were established in the laboratory. We decided on our procedures for two reasons. First, these procedures are widely used in the learning psychology literature and have consistently proven effective in generating the types of symbolic relations we were interested in (for a review see Hughes & Barnes-Holmes, 2016). Second, by establishing the meaning of the contextual cues in the laboratory, and by using these cues to relate previously unknown (abstract) stimuli with a single psychological property (valence), we sought to ensure tight experimental control over symbolic evaluative generalization.

valence of the non-focal stimuli was opposite to that of the valenced stimuli that focal items were paired with.

Method

Participants and Design. A total of 59, 64, 111, and 61 students participated in exchange for either a chocolate bar, course credits, or a monetary reward (Experiment 1: mean age = 21, *SD* = 6; Experiment 2: mean age = 20, *SD* = 6; Experiment 3: mean age = 20, *SD* = 4; Experiment 4: mean age = 21, *SD* = 5). The raw data, experimental scripts, and analysis files are available on the Open Science Framework website (https://osf.io/r6bgw). Sample size for all experiments was determined before any data analysis and was based on the availability of participants at the time of data collection. Final sample size (after participant exclusions noted below) allowed us to reliably observe an effect of $\eta^2_p = 0.036$ with a power (1- β) of .80 and α = .05 in the critical repeated measures ANOVAs for all experiments (Experiment 1: $\eta^2_p = 0.030$; Experiment 2: $\eta^2_p = 0.036$; Experiment 3: $\eta^2_p = 0.017$: Experiment 4: $\eta^2_p = 0.029$). In these studies, we report all measures, manipulations, and exclusions.

The block order of the indirect procedure and evaluative task order were counterbalanced across participants. In Experiment 3, we also counterbalanced the type of relation between focal stimulus and valenced stimuli (Opposition/Similarity) as well as order of indirect procedure (IAT/evaluative priming task).

Materials. In Experiments 1 and 2, six Pokémon characters were selected to serve as focal and non-focal stimuli (i.e., *yamask*, *unown*, *trubbish*, *klink*, *tynamo*, and *roggenrola*) while five pleasant and five unpleasant images served as valenced stimuli.[5] In Experiment 3, nonsense words served as focal and non-focal stimuli (i.e., *Pardal*, *Zatte*, *Ettalas*, *Ciney*,

[5] IAPs images 1710, 2352 and three additional images sourced from the internet served as positive images while IAPs images 1300, 9181, 9300, 9405, 2800 served as negative images. We also assigned each of the Pokémon characters a nonsense word as a name (e.g., Trubist, Sawg, Roggen, Omidit, Yamask, Tynam). This name was printed underneath the Pokémon during the learning and evaluative measure phases so that participants could easily identify the stimuli.

Witkap, and *Gageleer*). In Experiment 4, a different set of nonsense words served as focal and non-focal stimuli (i.e., *Vekte, Cuglo, Empeya, Pliaga, Rowth, Ambik*). In Experiments 3-4, five positive (*delicious*, *fresh*, *tasty*, *sweet*, *yummy*) and five negative adjectives (*disgusting*, *stale*, *nasty*, *sick*, *rotten*) served as valenced stimuli.[6] In all experiments, two arbitrary symbols were used as contextual cues.

Procedure. Participant provided informed consent. Overall there were three phases: contextual cue training, relational training, and evaluative measures (see Figure 1).

Contextual Cue Training. Training consisted of two types of trials designed to establish one symbol as meaning 'Same' and another meaning 'Opposite'. On each trial, two pictures that were similar (e.g., two images of dogs) or opposite (e.g., arrows pointing up and down) were presented at the top along with two symbols at the bottom of the screen (see Figure 2). On 'similarity' trials, choosing the "Same" symbol in the presence of two pictures that were similar caused "Correct" to appear for 1000ms. All stimuli then disappeared and the next trial began after 1000ms. Selecting the incorrect symbol caused "Incorrect" to appear. Feedback remained until participants emitted the correct response. All stimuli then disappeared and the next trial began after 1000ms. A broadly similar pattern of responding was required for 'opposition' trials. Critically however, selecting the 'Opposite' cue when shown two pictures that were opposite (e.g., fire/water, black/white) caused 'Correct' to appear whereas selecting the 'Same' cue caused 'Incorrect' to appear.

In Experiments 1-3, participants completed a minimum of one and a maximum of three blocks of trials. The number of trials within a block could vary from 20 to 80 depending on task performance. Within each block the location of the two symbols and trial-type was varied in a quasi-random order. Participants who emitted twenty consecutively correct responses moved on to a block of test trials. Test trials were identical to training trials.

[6] Note that the nonsense words used in Experiment 3 were actually the names of regional Belgian beers. Critically, pre-testing indicated that no-one was familiar with these names prior to the study.

However, novel images were presented and no feedback was provided. Successfully completing the test block required the correct cue be selected on 20/24 trials. Failure to do so resulted in re-exposure to training and testing until either (a) the mastery criterion was met or (b) three training and testing sequences were completed. Attaining the mastery criteria resulted in progression to relational training while failure resulted in participants being thanked, debriefed, and dismissed. [7]

Relational Training. Relational training aimed to establish two similarity relations: *Pokémon 1- 2- 3* and *Pokémon 4- 5- 6* (Experiments 1-2) or *Brand 1- 2- 3* and *Brand 4- 5- 6* (Experiments 3-4) (see Figure 3). Training consisted of four stages, each comprised of a minimum of one and a maximum of three blocks of trials. The number of trials within a block could vary from 20 to 60 depending on task performance.

During stage one, four relations were trained (e.g., *Pokémon/Brand 1-Same-Pokemon /Brand 2*; *Pokémon/Brand 4-Same-Pokémon/Brand 5*; *Pokémon/Brand 1-Opposite-Pokémon/Brand 5*; *Pokémon/Brand 2-Opposite-Pokémon/Brand 4*). For instance, when Pokémon 1 and 2 (or Pokémon 4 and 5) were displayed together, selecting the 'Same' cue caused "Correct" to appear. When an incorrect response was made error feedback appeared and progression required the correct cue to be selected. Alternatively, when Pokémon 1 and 5, or Pokémon 2 and 4 were presented together, selecting the 'Opposite' cue was reinforced and the 'Same' cue punished. Finally, in addition to the four relations, and to ensure that the cue meaning ('Same' and 'Opposite') remained salient throughout the task, cue training trials were also interspersed within each block.

[7] In Experiment 4, contextual cue training blocks consisted of a fixed set of 24 trials. Participants completed a minimum of one and a maximum of four such blocks (depending on their accuracy criteria). They then proceeded to a test block of 24 trials and progressed to the relational test phase regardless of test performance. Relational training blocks also consisted of blocks of 24 trials. Participants were exposed to a minimum of one or a maximum of four such blocks in each of the four relational training stages (each stage was followed by a test block). Failure to pass a relational test led to re-exposure to stages 3 and 4 and another relational test. Thereafter participants proceeded to the evaluative measures regardless of performance.

The second phase trained four new relations (e.g., *Pokémon/Brand 2-Same-Pokemon/Brand 3*; *Pokémon/Brand 5-Same-Pokémon/Brand 6*; *Pokémon 2/Brand -Opposite-Pokémon/Brand 6*; *Pokémon/Brand 3-Opposite-Pokémon/Brand 5*). In the third phase, participants were re-exposed to all of the previously trained relations. In the fourth phase, valence was established for the focal-stimulus in each relation by presenting it with a valenced image (Experiments 1-2) or a valenced word (Experiments 3-4) and requiring one of the two cues to be selected. Specifically, either Pokémon 1/Brand 1 and a positive image or word, or Pokémon 4/Brand 4 and a negative image or word were presented at the top along with the two cues at the bottom of the screen. In Experiments 1-2, choosing the 'Opposite' cue resulted in "Correct" being displayed whereas selecting the 'Same' cue produced error feedback. In Experiment 3, we manipulated the type of relation between valenced and focal stimulus, such that half the participants were reinforced for relating the two as same whereas the other half were reinforced for relating them as opposite. In Experiment 4, all participants were reinforced for relating stimuli as similar to one another. In each case, progression required that participants emit 20 correct responses on a training block followed by at least 20 out of 24 correct responses on a test block. Failure to do so resulted in re-exposure to the task; failure after three training and testing blocks resulted in being thanked, debriefed, and dismissed. [8]

Evaluative measures. In the IAT of Experiment 1, there were seven blocks. The first block (20 practice trials) required participants to sort images of two Pokémon into their respective categories, with one (e.g., Pokémon 3) assigned to the left ('E') key and the other (e.g., Pokémon 6) with the right ('I') key. In the second block (20 practice trials) they assigned negative words to the 'Bad' category using the left key and positive words to the

[8] Note that half of the participants completed a single session of contextual cue and relational training while the other half returned for a second session of training (and the evaluative measures) on the following day. We adopted this design in order to determine if the amount of training provided influenced the strength of evaluative responding. This factor was not found to do so – and as such – was not included in the analyses reported below.

'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 (e.g., Pokémon 3 and 'negative' words with the left key and Pokémon 6 and 'positive' words with the right key). The fifth block (20 trials) partially reversed key assignments, with Pokémon 3 now assigned to the right and Pokémon 6 with the left key. The sixth and seventh blocks (20 and 40 trials respectively) required participants to categorize Pokémon 6 with negative words and Pokémon 3 with positive words. Eight positively and negatively valenced adjectives were employed as one set of attribute stimuli (*happy, love, pleasure, fun, joy, kind, friendly, wonderful* vs. *sick, horrible, disgusting, sad, terrible, awful, pain, vomit*) while four images of Pokémon 3 and 6 at different orientations served as the second set. The IAT in Experiment 4 was similar with the exception of the stimuli employed. Brand 3 and 6 now served as category and attribute stimuli whereas the following valence stimuli served as attributes of the 'Good' and 'Bad' categories (fantastic, nice, brilliant, and pleasant vs. horrible, terrible, awful, unpleasant).

The IRAP of Experiment 2 consisted of a minimum of one and a maximum of three pairs of practice blocks followed by a fixed set of three pairs of test blocks. Each block consisted of twenty four trials that presented one of two stimuli (either Pokémon 3 or 6) at the top of the screen, a second stimulus (positive or negative adjective) in the middle of the screen, and two response options ("True" or "False") at the bottom of the screen (see Figure 4). This stimulus configuration resulted in four "trial-types": *Pokémon 3-Positive*; *Pokémon 3-Negative, Pokémon 6-Positive,* and *Pokémon 6-Negative*. Trial presentation and onscreen location of the two response options was varied in a quasi-random order. During one block of trials they had to respond as if Pokémon 3 was negative and Pokémon 6 was positive, whereas during a second block, they had to respond the opposite way, acting as if Pokémon 3 was positive and Pokémon 6 was negative. These two response contingencies were alternated

across successive blocks. The IRAP commenced with a pair of practice blocks. Participants progressed from the practice to the test blocks when they met accuracy (at least 75% accuracy) and latency criteria (median latency of less than 2000ms) on a successive pair of practice blocks. Failure to meet these criteria resulted in re-exposure to another pair of practice blocks until those criteria were either achieved or a maximum of four pairs of practice blocks were completed. Test blocks were similar to practice blocks with the exception that no performance criteria were required to progress from one block to the next.

The evaluative priming task of Experiment 3 consisted of one practice and two test blocks. The non-focal stimuli directly (Brand 2 and Brand 4) or indirectly (Brand 3 and Brand 6) related to the focal stimuli served as primes while four positive and four negative adjectives served as targets. Each trial began with the presentation of a fixation cross (500ms) followed by a brand name (200ms). A target stimulus was then presented and remained onscreen until a response was emitted. The next trial began after a 1000ms ITI. The task began with six practice trials during which one of the brand names was randomly assigned as a prime and either a positive or negative word as a target. Two blocks of 32 test trials were then administered during which each of the four primes were randomly presented with one of the four positive and one of the four negative targets. The order of these trials was randomized within each block. For the evaluative priming task, the four non-focal brands served as prime stimuli (*Brands 2, 3, 5,* and *6*) while four positive (*delicious, tasty, yummy, nice*) and four negative words (*disgusting, rotten, sick, vomit*) served as target stimuli.

Self-reported ratings were assessed in all experiments using a series of Likert scales. Each Pokémon character or Brand was presented once and participants provided their general impression of the stimulus using a scale ranging from -4 (Negative) to +4 (Positive) with 0 as a neutral point in Experiments 1-3 or -5 to +5 in Experiment 4.

Approach-avoidance task. Participants performed a manikin task (De Houwer, Crombez, Baeyens, & Hermans, 2001) that was designed to measure participants' approach-avoidance tendencies towards Brand 3 and Brand 6. In this task, participants are asked to move a manikin representing themselves on the screen by pushing the up or down arrow keys of the keyboard. Evidence suggests that this task triggers spontaneous distance-regulation tendencies to valenced stimuli (Krieglmeyer, Deutsch, De Houwer, & De Raedt, 2010) and produces stronger effects than many other variants measuring approach-avoidance tendencies (for evidence see Krieglmeyer & Deutsch, 2010). Before performing the task, participants were shown Brand 3 and Brand 6 and they were asked to indicate which of the two brands they liked the most. Participants were then instructed that they would perform a task in which they would see these two brands as well as a stick figure (manikin). Participants were asked to respond as quickly as possible whenever they saw one of the brands by moving the manikin towards or away from it. Participants were further asked to adhere to four guidelines. First, they should approach the brand that they just indicated that they liked the most (or the least, for participants who started with an incompatible stimulus-response mapping block) by moving the manikin towards it. Second, they should avoid the brand that they just indicated that they liked the least (or the most for participants who started with an incompatible stimulus-response mapping block) by moving the manikin away from it. Third, they should imagine themselves to be the manikin while performing the actions and finally, they should respond as quickly as possible to the brands without making too many mistakes. Participants completed one compatible and one incompatible response-mapping block, each consisting of 41 trials. Block order was counterbalanced across participants. Consistent with Krieglmeyer et al. (2010), each trial consisted of the presentation of the manikin in either the upper or the lower half of the screen, followed by the presentation of a stimulus (Brand 3/6) after 750ms. After participants pressed the appropriate key, the manikin moved up or down the screen

(towards or away from the stimulus). Participants had to press the key three times to gradually move the manikin all the way across the screen. The screen turned black 50ms after the third response. The inter-trial interval was 1000ms. If participants made an incorrect response, an error message appeared immediately after the first key press for 500ms.

Additional questions. After the evaluative measures, additional questions were asked. First, to ensure that the cues acquired the same meaning as we expected, a question probing for contextual cue meaning was administered. We also assessed whether people relied on the cues when evaluating the Pokémon or Brand and whether their responses were based on demand compliance (e.g., "*you just indicated that you either liked or disliked the various brands. On what did you base your responses: (a) what I learned during the experiment, (b) on what I thought the researcher wanted me to say, (c) on some other factor than what I learned or what the researcher wanted me to say, (d) I don't know*"). In Experiment 3, after participants were informed that the experiment was over, they were presented with six small boxes that were identical in shape, size, and color, each with the name of one brand printed on the top. As a reward for participating, they were offered the opportunity to select three "samples" of the brand products to take home with them. After participants made their choice they were debriefed, thanked, and dismissed.

Results

Analytic Strategy. A series of repeated measures analysis of variance (ANOVAs) and post-hoc *t*-tests were carried out on the self-reported ratings, IAT, IRAP, and approach-avoidance scores (*dependent variables*). Our analyses had two goals: to determine if the valence of evaluative responding depended on the relationship established between (a) focal and valenced stimuli by the cues and (b) whether this valence would generalize from focal to non-focal stimuli (*independent variables*).

Data Preparation

Explicit and Implicit measures. We excluded the data from participants who (1) failed contextual cue training or relational training (Experiment 1: 2 participants; Experiment 2: 13 participants; Experiment 3: 13 participants; Experiment 4: 5 participants), (2) did not return for the second session of training (Experiment 2: 5 participants), (3) failed the IRAP (Experiment 2: 2 participants), or (4) had incomplete data due to a computer error (Experiment 1: 2 participants; Experiment 3: 3 participants). This led to the removal of 4, 20, 16, and 5, participants from Experiments 1-4, respectively, and a final sample of 55, 44, 95, and 56 in those respective cases. Finally, a total of 3, 3, 5, and 1 participant in Experiments 1, 2, 3, and 4 respectively were demand compliant. Removing their data did not significantly influence the reported findings. Therefore their data were retained in all analyses reported below. Details of how the IAT, IRAP, priming, and approach-avoidance data were transformed can be found in the supplementary materials.

Hypothesis Testing. ***Self-Reported Ratings.*** Evaluative ratings from Experiments 1-2, and the opposition condition of Experiment 3 were submitted to a 2 (*US Valence:* Positive vs. Negative) x 3 (*Stimulus:* Focal Stimulus vs. Directly Related vs. Indirectly Related Non-Focal Stimulus) repeated measures ANOVA. Analyses revealed a main effect of US Valence in all Experiments, (Experiment 1: $F(1, 54) = 16.94$, $p < .001$, $\eta^2_p = 0.24$, 95% CI [0.07; 0.41]; Experiment 2: $F(1, 43) = 38.13$, $p < .001$, $\eta^2_p = 0.47$, 95% CI [0.24; 0.62]; Experiment 3: $F(1, 45) = 40.90$, $p < .001$, $\eta^2_p = 0.48$, 95% CI [0.25; 0.62]). Experiment 2, but not Experiments 1 or 3, also revealed a main effect of Stimulus, $F(2, 43) = 7.61$, $p < .001$, $\eta^2_p = 0.15$, 95% CI [0.01; 0.37]. We did not observe an interaction between Stimulus and Valence in any of these studies (all *ps* > .1).

One-sample t-tests indicated that participants disliked the focal stimulus that was symbolically opposite to positive images or words (i.e., Pokémon or Brand 1), (Experiment 1: $t(54) = 3.70$, $p < .001$, $d = 0.49$, 95% CI [0.22; 0.78]; Experiment 2: $t(43) = 6.45$, $p < .001$, *d*

= 0.97, 95% CI [0.61; 1.33]; Experiment 3: $t(45) = 6.11$, $p < .001$, $d = 0.90$, 95% CI [0.55; 1.24]). They also disliked the first non-focal stimulus (Pokémon or Brand 2), (Experiment 1: $t(54) = 3.88$, $p < .001$, $d = 0.52$, 95% CI [0.24; 0.80]; Experiment 2: $t(43) = 5.70$, $p < .001$, $d = 0.86$, 95% CI [0.51; 1.19]; Experiment 3: $t(45) = 4.43$, $p < .001$, $d = 0.65$, 95% CI [0.33; 0.97]), and the second non-focal stimulus in that same relation: (Pokémon or Brand 3), (Experiment 1: $t(54) = 3.10$, $p = .003$, $d = 0.42$, 95% CI [0.14; 0.69]; Experiment 2: $t(43) = 5.45$, $p < .001$, $d = 0.82$, 95% CI [0.47; 1.16]; Experiment 3: $t(45) = 4.02$, $p < .001$, $d = 0.59$, 95% CI [0.28; 0.90]).

In contrast, participants liked the focal stimulus that was symbolically opposite to negative images or words (Pokémon or Brand 4), (Experiment 1: $t(54) = 4.02$, $p < .001$, $d = 0.54$, 95% CI [0.26; 0.82]; Experiment 2: $t(43) = 4.03$, $p < .001$, $d = 0.61$, 95% CI [0.28; 0.92]; Experiment 3: $t(45) = 8.31$, $p < .001$, $d = 1.23$, 95% CI [0.84; 1.61]). They also liked the first non-focal stimulus: (Pokémon or Brand 5), (Experiment 1: $t(54) = 3.99$, $p < .001$, $d = 0.54$, 95% CI [0.25; 0.82]; Experiment 2: $t(43) = 5.57$, $p < .001$, $d = 0.84$, 95% CI [0.49; 1.18]; Experiment 3: $t(45) = 6.80$, $p < .001$, $d = 1.00$, 95% CI [0.65; 1.36]), and the second non-focal stimulus in that same relation (Pokémon or Brand 6), (Experiment 1: $t(54) = 2.46$, $p = .017$, $d = 0.33$, 95% CI [0.06; 0.60]; Experiment 2: $t(43) = 5.79$, $p < .001$, $d = 0.87$, 95% CI [0.52; 1.21]; Experiment 3: $t(45) = 5.51$, $p < .001$, $d = 0.81$, 95% CI [0.47; 1.14]).

In the similarity condition of Experiment 3, a main effect emerged for Stimulus, $F(2, 49) = 7.07$, $p = .001$, $\eta^2_p = 0.13$, 95% CI [0.04; 0.39]. A main effect also emerged for US Valence in Experiments 3-4, (Experiment 3: $F(1, 49) = 197.55$, $p < .001$, $d = 0.80$, 95% CI [0.69; 0.88]; Experiment 4: $F(1, 55) = 153.11$, $p < .001$, $d = 0.74$, 95% CI [0.60; 0.89]), as well as a two-way interaction between Stimulus and US Valence, (Experiment 3: $F(2, 49) = 7.58$, $p = .001$, $d = 0.13$, 95% CI [0.04; 0.39]; Experiment 4: $F(2, 55) = 8.31$, $p < .001$, $d = 0.13$, 95% CI [0.03; 0.24]). Participants liked Brand 1, (Experiment 3: $t(49) = 16.05$, $p <$

.001, d = 2.27, 95% CI [1.74; 2.79]; Experiment 4: $t(55) = 17.34$, $p < .001$, d = 2.32, 95% CI [1.81; 2.82]), Brand 2, (Experiment 3: $t(49) = 15.29$, $p < .001$, d = 2.16, 95% CI [1.65; 2.67]; Experiment 4: $t(55) = 9.20$, $p < .001$, d = 1.23, 95% CI [0.88; 1.57]), and Brand 3, (Experiment 3: $t(49) = 6.06$, $p < .001$, d = 0.86, 95% CI [0.53; 1.18]; Experiment 4: $t(55) = 7.11$, $p < .001$, d = 0.95, 95% CI [0.63; 1.26]). They disliked Brand 4, (Experiment 3: $t(49) = 11.41$, $p < .001$, d = 1.61, 95% CI [1.19; 2.03]; Experiment 4: $t(55) = 13.96$, $p < .001$, d = 1.87, 95% CI [1.43; 2.29]), Brand 5, (Experiment 3: $t(49) = 8.19$, $p < .001$, d = 1.16, 95% CI [0.79; 1.51]; Experiment 4: $t(55) = 5.76$, $p < .001$, d = 0.77, 95% CI [0.47; 1.07]), and Brand 6, (Experiment 3: $t(49) = 10.66$, $p < .001$, d = 1.51, 95% CI [1.09; 1.91]; Experiment 4: $t(55) = 7.18$, $p < .001$, d = 0.96, 95% CI [0.64; 1.27]) (see Table 1).

Finally, we examined whether the type of relationship established between stimuli (similarity vs. opposition) in Experiment 3 moderated evaluative responding. We first reverse scored ratings from the opposition condition and then compared them to those obtained in the similarity condition using a 2 (*Relation Type*: similarity vs. opposition) x 2 (*US Valence*) x 3 (*Stimulus*) ANOVA. In addition to the previously mentioned effects, analyses revealed a main effect for Relation Type, $F(1, 94) = 3.92$, $p = .05$, $\eta^2_p = 0.04$, 95% CI [0.00; 0.14], a two-way interactions between US Valence and Relation Type, $F(1, 94) = 6.01$, $p = .02$, $\eta^2_p = 0.06$, 95% CI [0.00; 0.17], as well as Stimulus and Relation Type, $F(2, 94) = 4.45$, $p = .04$, $\eta^2_p = 0.05$, 95% CI [0.00; 0.19]. Overall, there was a trend suggesting that similarity relations led to larger evaluative responses than those produced via opposition relations. This effect was larger for the focal compared to non-focal stimuli. That said, it was only evident for two stimuli in the first (Brands 1 and 2) and none in the second relation (Brands 4, 5, 6).

Table 1. Mean, standard deviations, and 95% confidence intervals for self-reported ratings of focal and non-focal stimuli in Experiments 1-4. Positive and negative refer to the valenced images that the focal stimulus was presented together with.

	Experiment 1	Experiment 2	Experiment 3		Experiment 4
Relation Type	*Opposition*		*Opposition*	*Similarity*	*Similarity*
	M SD	*M SD*	*M SD*	*M SD*	*M SD*
Focal Stimulus (*Positive*)	-1.2 (2.3)	-1.8 (1.9)	-1.9 (2.1)	3.2 (1.4)	4.0 (1.7)
(*Pokémon1/Brand1*)	[-1.81; -0.59]	[-2.30; -1.30]	[-2.51; -1.29]	[2.81; 3.59]	[3.56; 4.45]
Non-Focal Stimulus	-1.1 (2.1)	-1.4 (1.6)	-1.5 (2.3)	2.8 (1.3)	2.9 (2.4)
(*Pokémon 2/Brand 2*)	[-1.66; -0.55]	[-1.82; -0.98]	[-2.17; -0.84]	[2.44; 3.16]	[2.27; 3.53]
Non-Focal Stimulus	-0.9 (2.2)	-1.4 (1.7)	-1.5 (2.5)	2.0 (2.3)	2.8 (2.9)
(*Pokémon 3/Brand 3*)	[-1.48; -0.32]	[-1.85; -0.95]	[-2.22; -0.78]	[1.36; 2.64]	[2.04; 3.56]
Focal Stimulus (*Negative*)	1.2 (2.2)	1.2 (1.9)	2.2 (1.8)	-2.9 (1.8)	-3.7 (1.9)
(*Pokémon 4/Brand 4*)	[0.62; 1.78]	[0.69; 1.70]	[1.68; 2.72]	[-3.40; -2.40]	[-4.20; -3.20]
Non-Focal Stimulus	1.2 (2.1)	1.5 (1.8)	2.0 (2.0)	-2.4 (2.0)	-2.2 (2.9)
(*Pokémon 5/Brand 5*)	[0.65; 1.75]	[1.02; 1.98]	[1.42; 2.58]	[-2.95; -1.85]	[-2.96; -1.44]
Non-Focal Stimulus	0.8 (2.3)	1.7 (1.9)	1.9 (2.4)	-2.7 (1.8)	-2.7 (2.8)
(*Pokémon 6/Brand 6*)	[0.19; 1.41]	[1.2; 2.2]	[1.21; 2.59]	[-3.20; -2.20]	[-3.43; -1.97]

IAT (Experiments 1, 3, and 4). A one sample t-test revealed that IAT scores in Experiment 1 (i.e., where an opposition relation was established between focal and valanced stimuli) did not indicate a preference for Pokémon 6 over Pokémon 3 (M = 0.09, SD = 0.51), $t(54) = 1.36$, $p = .18$, $d = 0.18$, 95% CI [-0.05; 0.34]. In contrast, the IAT score of participants in the opposition condition of Experiment 3 did reveal an automatic preference for Brand 6 over Brand 3 (M = -0.25, SD = 0.32), $t(18) = -3.30$, $p = .004$, $d = 0.76$, 95% CI [0.24; 1.26]. In Experiments 3 and 4 (where a similarity relation was established between focal and valence stimuli) IAT scores revealed an automatic preference for Brand 3 over Brand 6, (Experiment 3: $t(23) = 6.28$, $p < .001$, $d = 1.28$, 95% CI [0.73; 1.82]; Experiment 4: $t(55) = 4.28$, $p < .001$, $d = 0.57$, 95% CI [0.29; 0.85]).

Note that, when IAT scores of Experiments 1 and 4 were submitted to a one-way ANOVA with IAT block order as a between groups variable a main effect emerged for IAT Block Order, (Experiment 1: $F(1, 53) = 5.39$, $p = .02$, $\eta^2_p = 0.09$, 95% CI [0.01; 0.18];

Experiment 4: $F(1, 54) = 8.15$, $p = .006$, $\eta^2_p = 0.13$, 95% CI [0.01; 0.29]). Participants who began the IAT in a manner that was consistent with prior training (consistent block first) showed an automatic evaluative bias towards Pokémon 6 relative to 3 (Experiment 1) or Brand 3 over 6 (Experiment 4) and this effect differed from zero, (Experiment 1: $t(23) = 2.86$, $p = .009$, $d = 0.59$, 95% CI [0.15; 1.01]; Experiment 4: $t(25) = 5.36$, $p < .001$, $d = 1.05$, 95% CI [0.56; 1.53]). Those who began the task in a manner that was inconsistent with prior training showed no such effect, (Experiment 1: $t(30) = 0.45$, $p = .65$, $d = -0.08$, 95% CI [-0.43; 0.27]; Experiment 4: $t(29) = 1.36$, $p = .19$, $d = 0.25$, 95% CI [-0.12; 0.61]).

IRAP (Experiment 2). If relational training was successful, then participants should show a similar pattern of automatic evaluative responses as seen on the IAT from Experiment 1 (i.e., a preference for Pokémon 6 over Pokémon 3).Submitting *D*-IRAP scores to a 2 (*Trial-Type*: positive, negative) x 2 (*Stimulus*: Pokémon 3, Pokémon 6) repeated measures ANOVA revealed an interaction between Trial-Type and Stimulus, $F(1, 39), = 44.67$, $p < .001$, $\eta^2_p = 0.53$. Comparing scores on the *Pokémon 3-Positive* and *Pokémon 6-Positive* trial-types revealed that participants endorsed the belief that Pokémon 6 was positive ($M = .39$, $SD = .38$) to a greater extent than they endorsed Pokémon 3 as being positive ($M = -.17$, $SD = .39$), $t(39) = 9.13$, $p < .001$, $d = 1.44$, 95% CI [1.14.; 1.74]. However, participants did not endorse the belief Pokémon 3 was negative ($M = .22$, $SD = .32$) to a greater extent than they endorsed Pokémon 6 as being negative ($M = .18$, $SD = .36$), $t(39) = 0.69$, $p = .50$, $d = 0.11$, 95% CI [-0.18; 0.35]. Finally, a one-sample t-test revealed that the overall *D*-IRAP score differed from zero, with participants showing a relative preference for Pokémon 6 over Pokémon 3 ($M = .15$, $SD = .27$), $t(39) = 3.62$, $p < .001$, $d = 0.57$, 95% CI [0.23; 0.90].

Evaluative Priming Task (Experiment 3). Submitting priming data to a 2 (Type of *Primes: Directly related to focal stimuli, indirectly related to focal stimuli*) x 2 (*Relation Type*) repeated measures ANOVA revealed no main or interaction effects (all *ps* > .2).

Behavioral Choice Task (Experiment 3). We defined a choice as being correct when Brand 1, 2, and 3 were selected by those in the similarity condition and Brand 4, 5, and 6 were selected by those in the opposition condition (the selection of any other combination of brands was defined as an incorrect choice). Seventeen participants did not take part in the task. Of the remaining seventy-nine participants, sixty three (80%) selected all three correct brands in-line with what they learned during relational training while the other sixteen (20%) made a partially correct response (selecting one or two of the correct options). Participants selected the three correct brands significantly more often than would be expected by chance, exact binomial p (one-tailed) < .001.

Approach-avoidance task (Experiment 4). If relational training was successful then we should observe an approach tendency towards Brand 3 and an avoidance tendency towards Brand 6. Participants' mean approach-avoidance task latencies were submitted to a one-tailed paired *t*-test. We observed a significant difference between latencies in compatible (M = 626ms, SD = 101ms) and incompatible blocks (M = 651ms, SD = 94ms), $t(43) = 1.98$, $p = .027$, $d = 0.30$, 95% CI = [4ms, Inf], indicating a preference for Brand 3 over Brand 6.

Discussion

Experiments 1-4 provide evidence for symbolic evaluative generalization. Participants first learned that a focal stimulus was symbolically similar to several non-focal stimuli. The focal stimulus was then paired with, but actually related as opposite (Experiments 1-3) or similar to (Experiment 4), positive or negative images. When participants learned that a Pokémon character or Brand was opposite to positive images they rated that character (and symbolically related characters) negatively. Yet when a Pokémon or Brand was opposite to negative images that character (and symbolically related characters) was rated positively. These effects were observed on explicit ratings, implicit evaluations measures (IAT and IRAP, but not evaluative priming task), and a measure of automatic approach-avoidance

tendencies. Experiment 3 also found that symbolic relations biased decision making: participants selected the non-focal brands during a choice test even though they had never been directly related with valenced events in the past.

Experiments 5-6

In Experiments 5 and 6 we set out to show that our procedures can also be used to study the impact of symbolic relations on the *magnitude* of liking. Towards this end, cue training was modified to establish two symbols as meaning 'More than' and 'Less than'. These cues were then used to generate a comparative relation comprised of five prizes (*Prize 1 < 2 < 3 < 4 < 5*). During a later phase, participants learned that they would receive a small 'prize' for taking part and that by selecting the first stimulus in the above relation (*Prize 1*) their prize money would increase by 1 cent. Selecting the second stimulus (*Prize 2*) allowed them to increase their earnings by either 25 cents (Experiment 5) or 1 euro (Experiment 6). Thereafter, self-report, implicit, and behavioral choice measures were administered.

We anticipated three outcomes. First, self-reported ratings should increase linearly as a function of a stimulus' location within the relation. Stimuli which were never paired with money (Prizes 3, 4, and 5) should be evaluated more positively than stimuli which reliably and consistently increased prize winnings (Prize 1 or Prize 2). Second, this linear increase in evaluation should also be evident at the implicit level. Third, training a small number of relations may cause people to 'act as if' those stimuli are related in several untrained ways, and it is this learning process (known as 'arbitrarily applicable relational responding' in the learning psychology literature) that underpins symbolic evaluative generalization effects (for more see Hughes & Barnes-Holmes, 2016a). In Experiment 5, we exposed participants to a relational test to determine if directly training a small number of relations would cause them to 'derive' or infer that stimuli were related in a number of novel and untrained ways.

Method

Participants. A total of 81 and 43 undergraduate students completed the study in exchange for a €5 payment (Experiment 5: mean age = 20 *SD* = 3; Experiment 6: mean age = 22, *SD* = 4). Our final sample size allowed us to reliably observe an effect of $\eta^2_p = 0.025$ (Experiment 5) and $\eta^2_p = 0.044$ (Experiment 6) with .80 power in the critical repeated measures ANOVAs.

Materials. The same set of nonsense words as in Experiment 3 served as the prize names (i.e., Pardal, Zatte, Ettalas, Ciney, and Witkap). Different quantities of money served as valenced stimuli (1 cent vs. 25 cents in Experiment 5 and 1 cent vs. 1 euro in Experiment 6). The same arbitrary symbols served as cues as in Experiments 1-4. However, this time the two cues were trained to mean 'More than' and 'Less than'.

Procedure. The study consisted of three phases: cue training, relational training, and evaluative measures. In Experiment 6 all participants completed two sessions of cue and relational training rather than one.

Contextual Cue Training. Training was similar to that used in Experiments 1-4. However, this time, one symbol was trained to mean 'More than' and a second as 'Less than'. Two pictures were presented at the top of the screen and two symbols on the bottom. To ensure that quantity - and not some other property - was the dimension along which the stimuli were related, a red circle was used as a discriminative stimulus for cue selection. Specifically, when presented with (a) an image of a red circle containing many items and (b) an image with a smaller number of items not in a circle, selecting the 'More than' cue was reinforced and the 'Less than' symbol was punished. The opposite pattern of responding was required on trials designed to establish the second symbol as meaning 'Less than'. Specifically, selecting the 'Less than' cue was reinforced when the picture containing the smaller number of items was enclosed in a red circle and the picture containing the larger number of items was not surrounded by a circle. Although training began with stimuli that

were physically related to one another the task subsequently used stimuli that were not physically related so that participants could learn that quantity was the dimension they should attend to. Participants could only progress once they met training and testing criteria.

Relational Training. A single comparative relation consisting of five nonsense words was generated using a similar procedure as in cue training (i.e. *Prize 1 < 2 < 3 < 4 < 5*) (see Figure 5). Training consisted of four phases, each comprised of a minimum of one and maximum of three blocks. Each block consisted of a minimum of 20 and maximum of 80 trials (depending on task performance). During phase one, three relations were established (*Prize 1 < 2; Prize 2 > 1; Prize 2 < 3*). The second phase of training was identical to the first with the exception that three additional relations were established (*Prize 3 < 4; Prize 4 < 5; Prize 5 > 4*) while the third phase exposed participants to all previously trained relations.

In phase four, we established a valence for Prize 1 and Prize 2 by making access to different amounts of money contingent on their selection. Participants were informed that they would receive a small monetary reward for taking part in the study and that they could increase this reward by clicking on one of the two nonsense words presented in the middle of the screen. A total of fifteen trials was then presented that displayed "number of opportunities remaining" in the upper left corner, the participant's overall prize winnings in the upper right corner, and Prizes 1 and 2 at the bottom of the screen. Selecting Prize 1 on any given trial always increased their total winnings by 1 cent while choosing Prize 2 added 25 cents (in Experiment 5) or 1 euro (in Experiment 6) to that same amount.

Relational Testing. Participants completed a relational test to determine whether a comparative relation was formed as predicted. On each trial, two Prize names were presented along with the 'More than' and 'Less than' cues and participants were asked to click on the symbol that describes the relationship between the two stimuli. Testing consisted of twelve trials, six of which presented Prize 2 and Prize 4 together, three of which presented Prize 2

and Prize 3 together, and three of which presented Prize 3 and Prize 4 together. Responding in accordance with the predicted comparative relation required participants to act as if (a) Prize 4 was more than Prize 3 or 2; (b) Prize 2 or Prize 3 were less than Prize 4 or (c) that Prize 3 was more than Prize 2 (no corrective feedback was given at any point). Participants were defined as having passed the test when they emitted a minimum of 10 out of 12 correct responses while those who did not were defined as having failed. [9]

Evaluative measures. Participants in Experiment 5 completed one of three IATs. The first IAT assessed automatic evaluative responding towards Prize 2 relative to 3; the second assessed responding towards Prize 3 relative to 4 while the third targeted Prize 2 relative to 4. In Experiment 6, automatic evaluative responding was assessed within rather than between participants using three "shortened" IATs, each comprised of five blocks (20 trials per practice block and 30 trials per test block).

Self-Report ratings of Prizes 2, 3, and 4 were assessed using a scale rating from -50 (negative) to +50 (positive) with 0 as a neutral point. A similar set of cue meaning and demand compliance questions was administered as before.

Questions. The same questions were used as in the previous studies. In Experiment 6 a behavioral choice task was also included to determine if the symbolic relations would impact decision making and selection judgments. Similar to Experiment 3, participants were presented with three identically shaped and sized boxes (with the names of Prizes 2, 3, and 4 printed on top) and asked to select one to add to their total prize earnings. Third, the time at which the relational test and behavioral choice tasks were administered was counterbalanced to determine if they had an influence on self-reported and IAT performance.

Results

[9] Although this pass criterion was lower than that typically seen in the learning psychology literature (100%), we have found that responding with at least 80% accuracy provides a useful means of distinguishing between participants who 'act-as-if' stimuli are symbolically related versus those who do not (see Hughes et al., 2016). It also shortens the overall number of training blocks encountered and thus time taken to complete the task.

Data Preparation. ***Participant exclusion.*** We excluded the data from participants who (1) failed contextual cue training or relational training (Experiment 5: 13 participants; Experiment 6: 4 participants), (2) did not return for the second session of training (Experiment 6: 1 participant), or (3) had incomplete data due to a computer error (Experiment 5: 4 participants; Experiment 6: 2 participants). This led to the removal of 17 and 7 participants from Experiments 5-6 respectively, and a final sample of 64 and 36 in those respective cases. Finally, one participant in both Experiments 5 and 6 indicated that they were demand compliant. Removing their data did not significantly influence the reported findings. Thus their data were retained for subsequent analysis.

IAT. The IAT was scored so that positive values reflected a relative response bias favoring the prize further along the comparative relation. For instance, when the relation between Prize 2 and Prize 3 was examined, positive scores reflected a bias favoring Prize 3 over 2. Yet when the relation between Prize 3 and Prize 4 was examined, a positive score indicates a bias for Prize 4 over 3 (a negative value for any of the three IATs indicated a reversed pattern of responding).

Relational test. 11 and 4 participants (17%; 11%) failed the relational test in Experiments 5 and 6, respectively. Given these small numbers it was not possible to compare the performance of those who passed or failed the test. Thus the data of all participants was retained for analysis.

Hypothesis Testing. ***Self-Reported Ratings.*** A repeated measures ANOVA (with stimulus as a within-subjects factor) revealed a main effect of Stimulus, (Experiment 5: $F(2, 63) = 10.81$, $p < .001$, $\eta^2_p = 0.25$, 95% CI [0.09; 0.38]; Experiment 6: $F(2, 70) = 30.15$, $p < .001$, $\eta^2_p = 0.46$, 95% CI [0.31; 0.56]). Participants in Experiment 5 exhibited a small positive response towards Prize 2 ($M = 6.55$, $SD = 21.83$; $t(64) = 2.39$, $p = .02$, $d = 0.30$, 95% CI [0.05; 0.55]), a larger response to Prize 3 ($M = 10.88$, $SD = 21.09$; $t(63) = 4.13$, $p < .001$, $d =$

0.52, 95% CI [0.25; 0.78]), and an even larger response to Prize 4 (*M* = 23.09, *SD* = 23.05; *t*(63) = 8.02, *p* < .001, *d* = 1.00, 95% CI [0.69; 1.30]). Participants in Experiment 6 did not show an evaluative response towards Prize 2 (*M* = 0.36, *SD* = 20.32; *t*(35) = 0.11, *p* = .92, *d* = 0.02, 95% CI [-0.31; 0.34]), but did show a positive response to Prize 3 (*M* = 7.06, *SD* = 17.08; *t*(35) = 2.48, *p* = .02, *d* = 0.41, 95% CI [0.07; 0.75]), and an even larger response to Prize 4 (*M* = 28.44, *SD* = 13.90; *t*(35) = 12.27, *p* < .001, *d* = 2.05, 95% CI [1.46; 2.62]). Post hoc tests using the Bonferroni correction revealed that Prize 4 was rated more positively than either Prize 2 (Experiment 5: *p* < .001; Experiment 6: *p* < .001) or Prize 3 (Experiment 5: *p* = .008; Experiment 6: *p* < .001).

IAT. Participants in Experiment 5 did not display a relative preference for Prize 3 over 2, *t*(20) = 1.48, *p* = .16, *d* = 0.32, 95% CI [-0.12; 0.76], Prize 4 over 3, *t*(15) = 1.39, *p* = .18, *d* = 0.35, 95% CI [-0.16; 0.85], or for Prize 4 over 2, *t*(26) = 1.07, *p* = .29, *d* = 0.21, 95% CI [-0.18; 0.59]. However, participants in Experiment 6 did show a relative preference for Prize 3 over 2, *t*(35) = 2.39, *p* = .02, *d* = 0.39, 95% CI [0.06; 0.74], Prize 4 over 3, *t*(35) = 3.22, *p* = .003, *d* = 0.54, 95% CI [0.18; 0.88], and for Prize 4 over 2, *t*(35) = 4.72, *p* < .001, *d* = 0.79, 95% CI [0.41; 1.16].

Behavioral Choice Task (Experiment 6). Participants were provided with the option to choose one of three stimuli. We defined a choice as correct when Prize 4 was selected and incorrect when Prize 2 or 3 was selected. Overall, twenty six participants (72%) selected the correct stimulus while ten did not (28%). Participants selected the correct prize significantly more often than would be expected by chance, exact binomial *p* (one-tailed) < .01.

Discussion

Using our symbolic generalization procedures, we were able to demonstrate that symbolically relations can also influence the magnitude of evaluative generalization effects. Results revealed that the magnitude of evaluative transfer depended on the location of stimuli

in the symbolic comparative relation. Despite being the only item that consistently and reliably increased the amount of money people received, Prize 2 was liked less than Prize 3 while Prize 3 was liked less than Prize 4. A linear increase in automatic stimulus evaluation was also evident, such that Prize 4 was preferred to 3 or 2.

General Discussion

Evaluations can be formed or changed whenever people encounter regularities in their environment, such as the repeated presentation of a single stimulus (mere exposure), pairing of stimuli (evaluative conditioning), or whenever behavior is linked to its consequences (approach-avoidance) (see De Houwer, 2009). These newly formed or changed (focal) stimulus evaluations can then be transmitted to non-focal stimuli in situations where the former shares some perceptual or symbolic relation with the latter. This phenomenon (evaluative generalization) may explain how likes and dislikes come to exert a powerful and far-reaching impact on behavior.

At the conceptual level, we argue that strong evidence for symbolic evaluative generalization requires not only that changes in liking are observed, but also that focal and non-focal stimuli bear no physical similarity to one another, and that the change in liking for non-focal stimuli cannot be explained in terms of direct or indirect spatio-temporal relations between stimuli (i.e., conditioning). Until now few procedures existed that were capable of meeting the above criteria. We therefore introduced a set of methods that are capable of doing so, and used them to obtain repeated evidence for symbolic evaluative generalization effects. Specifically, we observed changes in liking even though (a) all three stimuli in the symbolic relation were physically dissimilar to one another (which argues against perceptual generalization), (b) the valence of the focal stimulus was established only after it was related to non-focal stimuli (which argues against direct conditioning) and (c) the direction of the change in liking of the non-focal stimulus was opposite to the valence of the valenced stimuli

with which it was indirectly related (which argues against indirect conditioning). In the following section, we consider how mental and functional theories account for these findings.

Symbolic Evaluative Generalization: A Theoretical Analysis

In addition to providing a conceptual, methodological, and empirical contribution to the literature on (symbolic) evaluative generalization, we believe that our results also have theoretical implications. There are two different levels at which to explain symbolic generalization effects (De Houwer, 2011; Hughes et al., 2016): (1) a mental level that aims to uncover the mental mechanisms that *mediate* the impact of the environment on liking and (2) a functional level that aims to describe those elements of the past and present environment that *moderate* evaluative responses. We turn our attention to these two levels of explanation.

Mental level of explanation. Much work on evaluative generalization has been carried out at the mental level of analysis where the goal is to identify the mental representations and processes that mediate between environment and (evaluative) behavior. In what follows we consider how three broad classes of mental models (associative, propositional, and dual-process) accommodate our findings.

Associative mental models. Single process associative models represent a broad class of mental models that share a common property (the mental association) but which vary in their respective complexity (e.g., unidirectional vs. bidirectional associations). Although it is impossible to prove or disprove such a broad class of models (Miller & Escobar, 2001), our results do place some constraints on them. For instance, many of these models could accommodate the propagation of valence from one stimulus to another in equivalence relations by claiming that chains of associations were formed in memory between stimuli (e.g., between Pokémon 1 → Positive Images; Pokémon 1 → Pokémon 2 → Pokémon 3). In our studies, however, we always established a linear chain of stimulus relations such that Stimulus 1 was related to Stimulus 2 which was then related to Stimulus 3. Only after this

training was valence established for Stimulus 1. A change in valence from Stimulus 1 to 3 cannot occur via unidirectional associations because Stimulus 3 followed, rather than preceded the presentation of Stimulus 1 and 2. Thus bidirectional associations are needed (e.g., Custers & Aarts, 2011; Elsner & Hommel, 2004). Yet even here the association formation mechanism must somehow be able to take into account the *type* of relation between stimuli (e.g., similarity, opposition, or comparison). For instance, during the training phase of Experiments 1-4, Pokémon or brand products were repeatedly paired with stimuli of one valence and yet people subsequently acted as if those same Pokémon or brands had acquired the opposite valence. Likewise, in Experiments 1-6, certain stimuli were repeatedly paired with one another (e.g., Pokémon 1 and 5; Brand 2 and 4) and yet participants later acted as if those same stimuli were opposite to, or greater than, one another in valence or value. Although we do not exclude the possibility that some single process associative models can model how stimuli are related rather than merely associated, our findings do seem to challenge many currently available models (e.g., Rescorla & Wagner, 1972; Pearce & Hall, 1980). In short, although some future associative model might account for our data, our results constrain current and future associative models of symbolic evaluative generalization. More specifically, a successful associative model of our findings should not only allow for bidirectional associations but also for an impact of different types of relational information.

***Propositional models*.** The current data fit well with propositional models that involve qualified links between mental representations in memory. Whereas associations convey only the strength with which representations are linked, propositions can specify the structure of stimulus relations (e.g., ‘X is opposite to Y’ or ‘X is larger than Y’). Likewise, while associations gradually develop with many experienced pairings, propositions can be rapidly formed on the basis of experience, instructions, inference, or deductive reasoning (De

Houwer, 2009, 2014). When combined, these two properties of propositions may account for the current set of findings. [10]

In Experiments 1-3, cue training likely gave rise to propositions about the two symbols (e.g., "*This symbol means that the pictures are the same whereas that symbol means they are opposite*"). Relational training may have resulted in the formation of propositions about a small set of directly trained relations (e.g., "*Pokémon 1 is the same as Pokémon 2*", "*Pokémon 2 is the same as Pokémon 3*" and "*Pokémon 1 is the opposite of positive images*"). Participants may have then generated a set of novel propositions about relationships that were never explicitly trained (e.g., "*Pokémon 3 is bad and Pokémon 6 is good*"). It was these inferred (rather than the directly trained) propositions that may have mediated self-reported responses to the focal and non-focal stimuli. In other words, a series of propositions based on direct stimulus relating may not have been enough – rather these propositions may have to give rise to additional propositions in order to produce the expected generalization effects.

It is often assumed that the formation of propositions depends not only on awareness of the to-be-related stimuli but also the cognitive resources, time, and motivation to relate those stimuli (De Houwer, 2009). Although we did not manipulate these factors it could be argued that all necessary preconditions for the formation of propositions were fulfilled during cue and relational training phases of our experiments. Our findings are also consistent with a growing body of work showing that (once formed) propositions may be activated automatically from memory and guide the evaluation of stimuli on self-report and automatic measures alike (see Hughes, Barnes-Holmes, & De Houwer, 2011; De Houwer, 2014). For example, once the inferred proposition "*Prize 4 is more than Prize 1*" was generated it may have been stored in memory and retrieved automatically during testing.

[10] One could argue that there are ways to implement propositions within an associative model that operates on the basis of interconnected nodes through which activation spreads. Although we do not exclude this possibility, we do not know any current associative model that realizes this possibility.

***Dual-process models*.** Finally, dual-process accounts that allow for rules (Smith & DeCoster, 2000), judgments (Kahneman, 2003), and propositions (Gawronski & Bodenhausen, 2014) to feed into and create novel associations could account for our findings – provided that certain pre-conditions are met. For instance, and similar to the above propositional model, cue training may have given rise to propositions concerning the meaning of the two symbols. These propositions could have been used during relational training to generate additional propositions about the stimulus relations themselves. This set of "directly trained" propositions could subsequently give rise to inference-based propositions about stimuli that were never paired together. These latter propositions could have been subsequently transformed into and stored as associations in memory (explaining the outcomes observed in Experiments 1-6).

To some extent, it is unsurprising that, as a category, dual-process models can account for our results at least as well as single-process propositional models. For any version of a single-process propositional model, one could simply add an associative system to create a dual-process model that explains at least as much as that single-process propositional model. Hence, the real value of dual-process models needs to be evaluated on the basis of what it adds to models that include only one of the processes. Our findings require dual process models to articulate how, when, and why (a) the cue and relational training procedures give rise to propositions, (b) novel propositions will emerge based on a set of directly trained propositions, and (c) how these propositions are subsequently transformed into and stored as associations given that participants were never exposed to any training or instructions to do so. More generally, while the encoding of affirmation or negation propositions seems to fit with traditional notions of mental associations (e.g., "Pokémon-Good") it is not immediately clear how other propositions, such as those involving comparison, distinction, or hierarchy are stored as unqualified links in memory given their inherently relational content. Hence, we

believe that the results of the studies with comparative relations (Experiments 4 and 5) do challenge the idea that propositions can have automatic effects only because they are translated into associations.

The functional level of explanation. Evaluative generalization can also be examined from a functional perspective. Without going into too much detail, functional explanations are not concerned with identifying the mental representations and processes that mediate changes in liking. Rather they aim to identify which aspects of the environment give rise to changes in (evaluative) behavior.

At this level of explanation our findings are in-line with modern (functional) conceptualizations of learning and behavior (e.g., Hayes, Barnes-Holmes, & Roche, 2001). They suggest that generalization may not be a simple perceptual phenomenon as often thought, at least when it comes to organisms with access to the symbolic learning pathway. Rather, generalization effects may represent instances of a more general class of behaviors known as 'relational responses' (i.e., responses that involve responding to one stimulus in terms of another). Whereas many species learn to respond to one stimulus based on its physical similarity or difference to another stimulus, humans gain access to a second, more 'advanced', type of (arbitrarily applicable) relational responding early in their development. The ability to learn and behave in this way effectively unshackles humans from the physical world and allows them to relate stimuli in ways that do not depend on their physical commonalities but that are under the control of contextual cues in the environment. [11] This general argument is also consistent with developments elsewhere in the evaluative learning literature (De Houwer & Hughes, 2016; Hughes, De Houwer, & Barnes-Holmes, 2016) and highlights the possibility that changes in liking to focal stimuli (e.g., in evaluative

[11] For communication purposes we have greatly simplified the origins, properties, and implications of relational responding for human language and cognition. Those interested in a detailed, accessible treatment of this literature are recommended to read Hughes & Barnes-Holmes (2016) or Hughes, De Houwer, & Barnes-Holmes (2016).

conditioning, persuasion, and mere exposure) and non-focal stimuli (e.g., in generalization procedures such as those used above) may be instances of symbolic relational behavior.

Open Questions and Future Directions

The procedures we introduce here provide a concrete means of asking and answering a whole host of new questions about symbolic evaluative generalization. Empirically speaking, we have only begun to scratch the surface of all the possible ways that symbolic relations may influence evaluations of non-focal stimuli. Future work could determine whether and what properties of evaluative responding emerge when different types, numbers, and combinations of symbolic relations are established. We found that valence was simply transferred in similarity relations, reversed in opposition relations, and changed in relativistic ways when comparative relations were established. It remains to be seen whether other relations also have a signature "fingerprint" when it comes to evaluative change. Note that in Experiments 1-6 we exclusively focused our attention on manipulating the symbolic relationship between focal and valenced stimuli. Future work could also examine whether similar outcomes are obtained when it is the relation between focal and non-focal stimuli that is manipulated. For example, rather than establish *Focal Stimulus-Opposite-Positive* researchers could instead manipulate *Focal stimulus-Opposite-Non-focal stimulus*.

When carrying out this work researchers should remember that symbolic relations can transform many different properties of stimuli – not just their evaluative properties. Although we restricted our analysis to the evaluative domain, symbolic relations may transform functions that are relevant for clinical (e.g., fear, anxiety, disgust), social (accessibility), consumer (brand quality, identification), health (avoidance, escape), and cognitive psychology (attention). We also limited our analysis to the formation of evaluations and remained comparatively silent to their modification. Future work could investigate whether symbolic relations can also be used to change evaluations of non-focal stimuli. We observed

tentative evidence for the idea that certain relations (similarity) give rise to stronger evaluative responses than others (opposition) (Experiment 3). If future work were to replicate this finding then it would support an increasing body of work showing that the type of relationship established between stimuli moderates the strength of evaluative responding (for more on this see Moran, Bar-Anan, & Nosek, 2016). Likewise, although we obtained effects on self-reports, IATs, IRAPs, behavioral choice, and approach-avoidance measures, we failed to find such effects on an evaluative priming task and effects on the IATs were often moderated by IAT block order (consistent vs. inconsistent first). This suggests that implicit non-focal evaluations were perhaps weaker than their self-reported counterparts, and is a topic worth of future investigation. Finally, one reviewer pointed out that our manipulation required participants to first infer the meaning of the contextual cues and then repeatedly relate stimuli in an effortful way. The training phase may have allowed them to 'figure out' the purpose of the task and thus based their evaluative responses on what they thought the researchers wanted to find (demand) and not what they actually thought or felt. Although possible, we found that only a fraction of participants reported that they were demand compliant in the six experiments. Even if one were to dismiss self-reported ratings as evidence of demand we still observed changes in liking on IATs, IRAPs, approach-avoidance, and behavioral choice tasks. Thus the likelihood that the reported findings are purely demand seems rather low. Nevertheless, future work could replicate our findings while controlling for demand in other ways (e.g., employ a cover story or a measure impervious to demand).

Conclusion

In this paper we offer a unique set of procedures that allows for the study of symbolic evaluative generalization in a way that controls for perceptual similarity and direct pairings while also allowing for the manipulation of relation type between stimuli. We introduced the

method, showed that it leads to the generalization of explicit and implicit evaluations, and does so in a way that depends on the type of relation between focal and valenced stimuli. For those operating at the functional level of analysis, such effects seem to represent instances of (arbitrarily applicable) relational responding. For those operating at the mental level, our findings are best explained by single process propositional or dual-process models appealing to some combination of associations and propositions. Regardless of what level of analysis researchers operate at, the distinction between perceptual and symbolic generalization may serve to strengthen the study of evaluative learning by organizing existing and generating new knowledge about the conditions under which likes and dislikes are transmitted and transformed.

References

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record, 60*(3), 527-542.

Bennett, M. P., Meulders, A., Baeyens, F., & Vlaeyen, J. W. (2015). Words putting pain in motion: The generalization of pain-related fear within an artificial stimulus category. *Frontiers in Psychology, 6*. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4415322/

Cicero, S. D., & Tryon, W. W. (1989). Classical conditioning of meaning—II. A replication and triplet associative extension. *Journal of behavior therapy and experimental psychiatry, 20(3),* 197-202.

Custers, R., & Aarts, H. (2011). Learning of predictive relations between events depends on attention, not on awareness. *Consciousness and Cognition, 20,* 368-378.

Deacon, T. W. (1997) The symbolic species. New York: Norton

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. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science*, *6*(2), 202–209.

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

De Houwer, J., Crombez, G., Baeyens, F., & Hermans, D. (2001). On the generality of the affective Simon effect. Cognition & Emotion, 15, 189–206. doi:10.1080/0269993004200051

De Houwer, J., Hughes, S., & Barnes-Holmes, D. (2016). Associative learning as higher-order cognition: Learning in human and nonhuman animals from the perspective of propositional theories and Relational Frame Theory. *Journal of Comparative Psychology, 130,* 215-225.

Elsner, B., & Hommel, B. (2004). Contiguity and contingency in action effect learning. *Psychological Research, 68,* 138-154.

Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: valence asymmetries. *Journal of Personality and Social Psychology*, *87*(3), 293-311.

Gawronski, B., & Bodenhausen, G. V. (2014). The associative-propositional evaluation model: Operating principles and operating conditions of evaluation. *Dual-Process Theories of the Social Mind*, 188–203.

Gawronski, B., & Quinn, K. A. (2013). Guilty by mere similarity: Assimilative effects of facial resemblance on automatic evaluation. *Journal of Experimental Social Psychology*, *49*(1), 120–125.

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.

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.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational Frame Theory: A Post-Skinnerian account of human language and cognition.* New York: Plenum Press.

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.

Hughes, S., & Barnes-Holmes, D. (2016). Relational frame theory: The basic account. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds), *The Wiley handbook of contextual behavioral science* (pp. 129-178), West Sussex, UK: Wiley-Blackwell.

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorizing in implicit attitude research: Propositional and behavioral alternatives. *The Psychological Record, 61*(3), 465-496.

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., De Houwer, J., & Perugini, M. (2016a). Expanding the boundaries of evaluative learning research: How intersecting regularities shape our likes and dislikes. *Journal of Experimental Psychology: General, 145(6)*, 731-754.

Hughes, S., De Houwer, J., & Perugini, M. (2016b). The functional-cognitive framework for psychological research: Controversies and resolutions. *International Journal of Psychology*, *51*(1), 4–14.

Hütter, M., Kutzner, F., & Fiedler, K. (2014). What is learned from repeated pairings? On the scope and generalizability of evaluative conditioning. *Journal of Experimental Psychology: General*, *143*(2), 631-643.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist, 58*(9), 697-720.

Krieglmeyer, R., Deutsch, R., De Houwer, J., & De Raedt, R. (2010). Being Moved: Valence Activates Approach-Avoidance Behavior Independently of Evaluation and Approach-Avoidance Intentions. *Psychological Science, 21*, 607–613.

Krieglmeyer, R., & Deutsch, R. (2010). Comparing measures of approach-avoidance behaviour: The manikin task vs. two versions of the joystick task. *Cognition & Emotion, 24* , 810-828.

Lissek, S., Kaczkurkin, A. N., Rabin, S., Geraci, M., Pine, D. S., & Grillon, C. (2014). Generalized anxiety disorder is associated with overgeneralization of classically conditioned fear. *Biological Psychiatry, 75*(11), 909–915.

Maio, G. R., & Haddock, G. (2007). Attitude change. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (2nd ed. pp. 565–586). New York, NY: Guilford.

Meulders, A., & Vlaeyen, J. W. (2013). The acquisition and generalization of cued and contextual pain-related fear: An experimental study using a voluntary movement paradigm. *Pain, 154*(2), 272–282.

Miller, R. R., & Escobar, M. (2001). Contrasting acquisition-focused and performance -focused models of acquired behavior. *Current Directions in Psychological Science, 10*(4), 141-145.

Ranganath, K. A., & Nosek, B. A. (2008). Implicit attitude generalization occurs immediately; explicit attitude generalization takes time. *Psychological Science, 19*(3), 249–254.

Ratliff, K. A., Swinkels, B. A., Klerx, K., & Nosek, B. A. (2012). Does one bad apple (juice) spoil the bunch? Implicit attitudes toward one product transfer to other products by the same brand. *Psychology & Marketing, 29*(8), 531–540.

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

Skowronski, J. J., Carlston, D. E., Mae, L., & Crawford, M. T. (1998). Spontaneous trait transference: Communicators take on the qualities they describe in others. *Journal of personality and social psychology, 74(4)*, 837-848.

Tryon, W. W., & Cicero, S. D. (1989). Classical conditioning of meaning—I. A replication and higher-order extension. *Journal of behavior therapy and experimental psychiatry, 20(2),* 137-142.

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

Verosky, S. C., & Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science, 21(6),* 779-785.

Völckner, F., & Sattler, H. (2006). Drivers of brand extension success. *Journal of Marketing, 70*(2), 18–34.

Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology*, *3*(1), 1-14.

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.

Zebrowitz, L. A., White, B., & Wieneke, K. (2008). Mere exposure and racial prejudice: Exposure to other-race faces increases liking for strangers of that race. *Social Cognition*, *26*(3), 259-275.

Appendix A

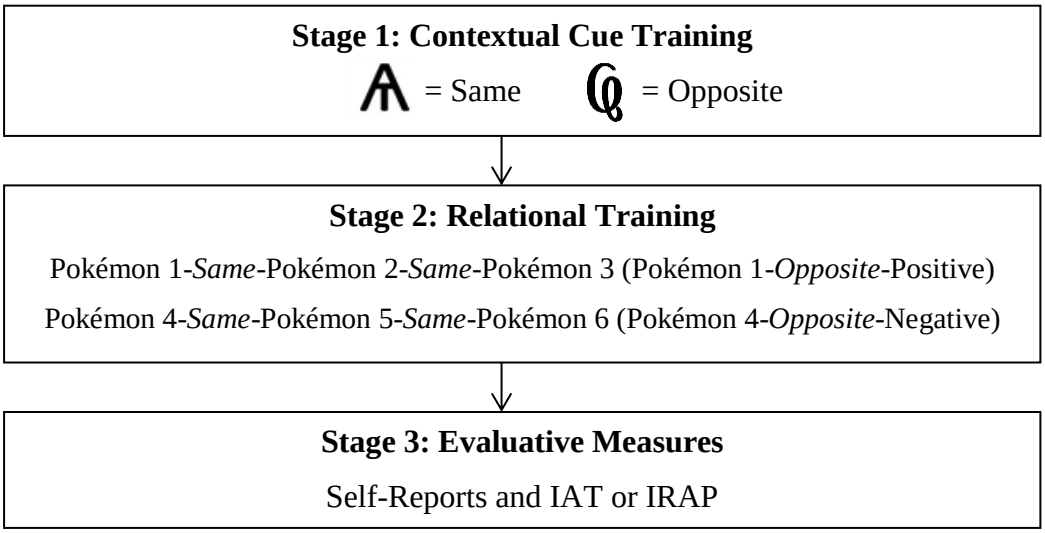

Figure 1. Overview of the experimental sequence in Experiments 1-2.

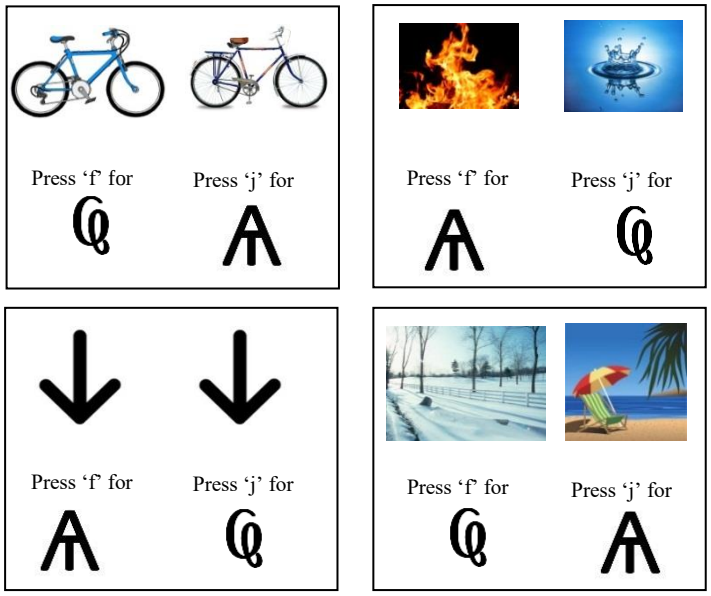

Figure 2. Examples of cue training and testing trials. Each trial consisted of two pictures at the top and two symbols at the bottom of the screen. Selecting the contextual cue deemed correct on any given training trial resulted in the presentation of "Correct" while selecting the cue deemed incorrect caused "Incorrect" to appear (no feedback was presented following responding on test trials).

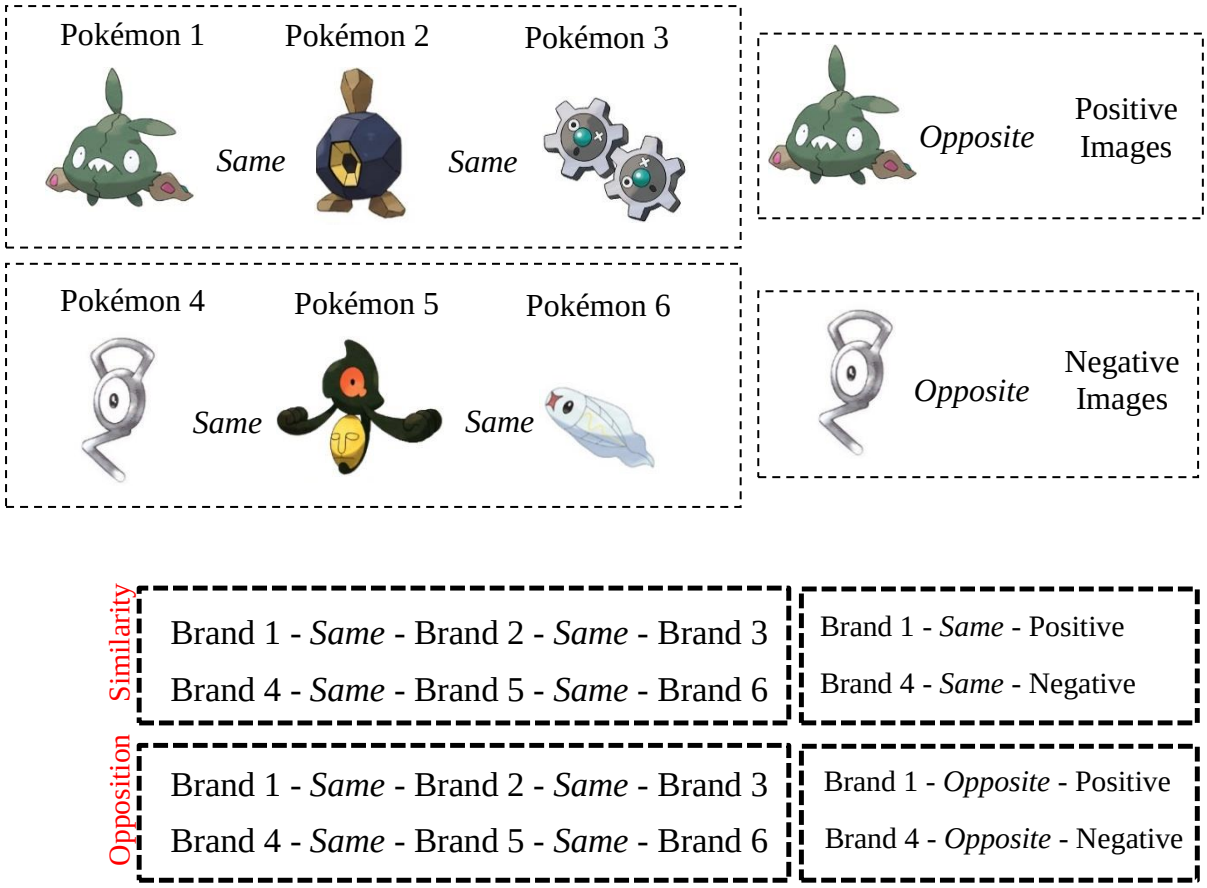

Figure 3. Schematic of the similarity relations (*above*) and opposition relations trained in Experiments 1-4 (*below*). Two symbolic similarity relations were established using a cue meaning 'Same' (*Pokémon/Brand 1-2-3* vs. *Pokémon/Brand 4-5-6*). An opposition relation was then established between the focal stimulus in each relation and valenced stimuli (e.g., *Pokémon/Brand1-Opposite-Positive* and *Pokémon/Brand 4-Opposite-Negative*). In Experiments 3 (for half the participants) and 4 a similarity relation was established between the focal stimuli and valenced stimuli.

Figure 4. Examples of the four IRAP trial-types used in Experiment 2. A Pokémon appeared at the top of the screen along with a valenced adjective in the middle of the screen (e.g., '*Pleasant*' or '*Disgusting*') and two relational response options ('*True*' and '*False*') at the bottom of the screen.

Prize 1 - *Less than* - **Prize 2** - *Less than* - **Prize 3** - *Less than* - **Prize 4** - *Less than* - **Prize 5**

Prize 1 = € 0.01 per trial

Prize 2 = € 0.25 or €1 per trial

Figure 5. Overview of the comparative relation established during relational training in Experiments 5-6. Five nonsense words served as the names of prizes that were symbolically related as being less than one another (Prize 1 < 2 < 3 < 4 < 5). Valence was then established for Prize 1 by pairing it with repeated access to €.01 per trial and Prize 2 by pairing it with access to €.25 (Experiment 5) or €1 per trial (Experiment 6).

Supplementary Materials

Experiments 1-4. Preparation of IAT, IRAP, Priming, and Approach-Avoidance data.

IAT. Following the recommendations of Greenwald, Nosek, and Banaji (2003), response latency data for the IAT were prepared using the *D* scoring algorithm. These *D* scores reflect the difference in mean response latency between the critical blocks divided by the overall variation in those latencies. In Experiments 1-3 positive scores reflected a response bias for Pokémon/Brand 6 over Pokémon/Brand 3. In Experiment 4 positive values indicate a response bias for Brand 3 over Brand 6 (see Figure 3). Negative scores indicated an opposite response pattern.

IRAP. The primary data obtained from the IRAP was response latency, defined as the time in milliseconds (ms) that elapsed from the onset of each trial to the first correct response emitted. Response latencies were transformed into *D*-IRAP scores using an adaptation of Greenwald et al.'s (2003) *D* algorithm (see Barnes-Holmes et al., 2010). We calculated four *D*-IRAP scores, one for each of the stimulus relations assessed by the task (*Pokémon 3-Negative*, *Pokémon 3-Positive*, *Pokémon 6-Negative* and *Pokémon 6-Positive*). Positive values indicate a response bias towards affirming a relationship between a Pokémon and positive items (or rejecting the relation between a Pokémon and negative items). Negative scores indicated an opposite pattern of responding. The four trial-type scores were also collapsed into a single (overall) *D*-IRAP score with positive values indicating a relative response bias favoring Pokémon 6 over 3 and negative values the opposite.

Priming. Evaluative Priming data was first screened for outliers. Trials on which the target was incorrectly classified (3.8%) or with reaction times below 300ms or greater than 1000ms (3.4%) were excluded. Two priming scores were calculated, one for the brands directly related to a focal stimulus (Brand 2 and Brand 4) and another for the brands indirectly related to the focal stimulus (Brand 3 and Brand 6). This was achieved by

subtracting the mean response latencies for trials in which a brand name appeared before positive targets from those in which it appeared before negative targets. A positive score indicates quicker responding when a prime appeared before positive target stimuli. Negative scores indicate quicker responding when a prime appeared before negative target stimuli.

Approach-avoidance. In accordance with Krieglmeyer et al. (2010) we excluded the first trial of each block as well as (a) trials in which participants made an error (6.0% of trials) or (b) trials with latencies below 150ms or above 1500ms (2.6% of trials). For each participant, we calculated the mean response latencies for the block in which they approached Brand 3 and avoided Brand 6 (compatible block) and the block in which they approached Brand 6 and avoided Brand 3 (incompatible block).