RUNNING HEAD: TESTING THE ROBUSTNESS OF IR AND OEC EFFECTS

The Influence of Extinction and Counterconditioning on Operant Evaluative Conditioning and Intersecting Regularity Effects

Sean Hughes1, Simone Mattavelli1, 2, Ian Hussey1, and Jan De Houwer1

1*Ghent University*

2 *University of Milan-Bicocca*

**Authors Note**

SH, IH, and JDH, Department of Experimental Clinical and Health Psychology, Ghent University. SM, Dipartimento di Psicologia, Università Milano-Bicocca. This research was conducted with the support of Methusalem Grant BOF16/MET\_V/002 to JDH. Correspondence concerning this article should be sent to sean.hughes@ugent.be.

**Abstract**

One of the most effective methods of influencing what people like and dislike is to expose them to systematic patterns (or ‘regularities’) in the environment, such as the repeated presentation of a single stimulus (mere exposure), two or more stimuli (evaluative conditioning) or to relationships between stimuli and behavior (approach/avoidance). Hughes, De Houwer, and Perugini (2016) found that evaluations also emerge when regularities in the environment *intersect* with one another. In this paper we examined if evaluations established via operant evaluative conditioning and intersecting regularities can be undermined via extinction or revised via counterconditioning. Across seven pre-registered studies (*n* = 1071) participants first completed a learning phase designed to establish novel evaluations followed by one of multiple forms of extinction or counterconditioning procedures designed to undo them. Results indicate that evaluations established in this way were resistant to extinction but could be counterconditioned via stimulus valence reversal. Theoretical and practical implications along with future directions are discussed.

*Keywords*: Intersecting Regularities, Extinction, Counterconditioning, Operant Evaluative Conditioning, Implicit, Attitudes

**The Influence of Extinction and Counterconditioning on Operant Evaluative Conditioning and Intersecting Regularity Effects**

Over the past century research in social and learning psychology has converged on a seemingly simple yet powerful idea: what we like and dislike is exquisitely sensitive to our interactions with the world around us. By exposing people to specific patterns of events in the environment (‘regularities’) we can quickly and easily influence what they like and dislike. [[1]](#footnote-1)

For instance, one can change liking by presenting the same stimulus over and over again: radio broadcasters often play a new song many times shortly after its release, and people repeatedly exposed to that song tend to evaluate it more positively than those who were not (i.e., the mere exposure [ME] effect; Moreland & Topolinski, 2010). Another type of regularity involves pairing stimuli: advertisers often pair a neutral stimulus (e.g., perfume) with a valenced stimulus (e.g., famous actress) to alter evaluations of the former in-line with the latter (i.e., evaluative conditioning [EC] effect; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). A third regularity involves relating certain actions to stimuli. For instance, the act of pushing alcohol away and pulling soft-drinks towards oneself influences evaluations of those stimuli as well as how much they are consumed (i.e., approach/avoidance [AA] effects; Van Dessel, Eder, & Hughes, 2018). Although ME, EC, and AA effects are all instances of evaluative learning, they differ in the type of regularity that leads to changes in liking (i.e., ME: regularity in the presence of one stimulus; EC: regularity in the presentation of two stimuli; AA: regularity in the relationship between stimulus and action).

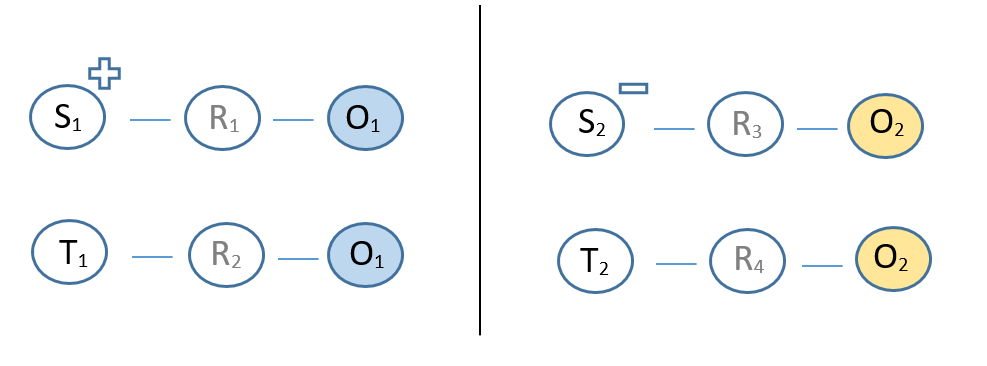
Yet evaluative learning does not stop here. Hughes, De Houwer, and Perugini (2016) recently introduced another way of arranging the environment in order to influence evaluations. They labelled this procedure evaluative learning via *intersecting regularities* (IR). Whereas EC, ME, and AA are relatively simple, insofar as they involve a change in liking due to a single regularity (*see above*), intersecting regularities procedures are more complex: they involve a situation where two or more regularities intersect with one another. By ‘intersect’ we mean that the regularities share one or more elements (e.g., a common stimulus or response), and because of this shared element, a change in liking occurs.

To illustrate this idea more clearly, consider the well-known sensory preconditioning procedure (see De Houwer & Hughes, 2020). Here two neutral stimuli (e.g., Bob and Chris) are initially paired with one another and one of the two is subsequently paired with an aversive stimulus (e.g., Bob is paired with unpleasant images). Research shows that people will come to dislike Bob *and* Chris even though Chris was never directly related with the unpleasant images. Such a procedure establishes *two* regularities between stimuli (i.e., one regularity involving the presentation of Bob and Chris; and another involving the presentation of Bob and unpleasant images). These two regularities also intersect in terms of a common element (Bob), and because of this intersection, a change in liking occurs (Chris is disliked). The dislike of Chris does not stem from a single regularity (e.g., Chris being paired with unpleasant images). Rather it stems from the intersection between one regularity (Bob-Chris) and another (Bob-unpleasant).

Hughes et al. (2016) argued that different regularities can be made to intersect with one another in many different ways, some of which have already been discovered (e.g., sensory preconditioning) and others that have not. To demonstrate their point, they had people complete a simple learning task wherein a certain button had to be pressed whenever a particular stimulus appeared onscreen (see Figure 1). For instance, if they pressed one button when a *positive* source stimulus was displayed then that stimulus disappeared and a neutral outcome stimulus took its place (positive source [S1] 🡪 response 1 🡪 **neutral outcome [O1]**). If a neutral target appeared then pressing a second button caused that stimulus to disappear and the same neutral outcome to appear (neutral target [T1] 🡪 response 2 🡪 **neutral outcome [O1]**). On other trials, pressing a third button whenever a *negative* source stimulus was on screen caused that stimulus to disappear and a second neutral outcome to take its place, while pressing a fourth button when a second neutral target was present caused the same neutral outcome to appear (i.e., negative source [S2] 🡪 response 3 🡪 **neutral outcome [O2]**; and neutral target [T2] 🡪 response 4 🡪 **neutral outcome [O2]**).

**Figure 1**

*Schematic overview of the IR procedure from Hughes et al. (2016) Experiment 2.*



*Note*. S refers to source stimulus, R to a response, O to an outcome stimulus, and T to a target stimulus. The + and – indicate the valence of the source stimulus (either positive or negative).

Put simply, an operant contingency containing a valenced source stimulus ‘intersected’ with a contingency containing a neutral target stimulus (i.e., the two contingencies shared the same outcome stimulus). As a result, people liked target stimulus (T1) and disliked target stimulus (T2), even though neither was directly related with valenced source stimuli during the learning phase. [[2]](#footnote-2) These outcomes were obtained on self-reported, implicit, and behavioral intention measures (see Hughes et al., 2016 or Ebert, Steffens, von Stülpnagel, & Jelenec, 2009, for demonstrations of various IR effects based on different types of operant contingencies; see Mattavelli, Richetin, Gallucci, & Perugini, 2017, for a review and meta-analysis of studies on one type of IR effect; and see Hughes et al., 2016, for a discussion of real world instances of IR effects). [[3]](#footnote-3)

Until now research on learning via intersecting regularities has focused on how such procedures give rise to novel evaluative responses. Yet the robustness of those evaluations still remains to be seen. In other words, can likes and dislikes established in this way be subsequently modified or eliminated using the procedures and methods commonly used to change evaluations using other regularities (such as stimulus pairing)? Given applied and theoretical importance of research on malleability of conditioned changes in liking, we deemed it important to examine the malleability of changes in liking that result from intersecting regularities. In this paper we examined the impact of two intervention procedures that have been highly popular in evaluative learning research: *extinction* and *counterconditioning*.

**Extinction**

Researchon extinction typically relies on a procedure with two phases. Consider, for instance, extinction in the context of evaluative conditioning. In a first phase (acquisition) participants are exposed to a contingency between a neutral conditioned stimulus (CS) and a valenced unconditioned stimulus (US). Thereafter the valence of the CS typically changes in-line with that of the US. During the second phase (extinction) the CS is presented alone in the absence of the US. In this way the extinction phase involves the removal of the (CS-US) contingency that originally gave rise to CS evaluations. Interestingly, many studies reveal no, or only a small, change in EC effects following an extinction procedure (e.g., Baeyens et al., 1988; Blechert, Michael, Williams, Purkis, & Wilhelm, 2008; Baeyens, Díaz, & Ruiz, 2005; Gast & De Houwer, 2013; Hermans, Crombez, Vansteenwegen, Baeyens, & Eelen, 2003; Vansteenwegen, Francken, Vervliet, De Clercq, & Eelen, 2006). That said, other studies have found that EC effects can be reduced following extinction trials (Lipp, Mallan, Libera, & Tan, 2010; Lipp, Oughton, & LeLievre, 2003). A recent meta-analysis confirmed that, across studies, EC effects measured after the extinction procedure are smaller than those measured before an extinction procedure, although the former are still substantial (Hofmann et al., 2010). These findings suggest that EC seems to be driven primarily by CS-US co-occurrences, rather than statistical contingency, and produces lasting changes in liking that persist even when CS and US no longer co-occur.

**Counterconditioning**

The robustness of evaluations can also be examined via counterconditioning. Similar to extinction, counterconditioning also tends to involve a procedure with two phases. For instance, during an initial (acquisition) phase a contingency is established between two stimuli by pairing a neutral CS with a valenced US. In a second (counterconditioning) phase the CS is then paired with a US of the opposite valence (e.g., a CS that was first paired with a positive is now paired with a negative US). People rate the CS in-line with the initial valence of the US after the first phase and then in-line with the subsequent valence of the US after the second phase (e.g., Kerkhof, Vansteenwegen, Baeyens, & Hermans, 2011).

## **The Current Research**

Across a series of studies we examined if evaluations established via intersecting regularities or operant evaluative conditioning (*see below*) can be undone via extinction or modified via counterconditioning. This work was designed to explore environmental moderators of intersecting regularities effects that proved to be vital in the study of other forms of evaluative learning.

***Examining the Robustness of Evaluations Established via IR***

Our main goal was to determine if evaluations established via IR can be modified via extinction procedures (Experiments 1-4) or countercondition procedures (Experiments 5-7).

Experiments 1-3 sought to extinguish evaluations by removing the intersecting element (outcome stimulus) connecting source and target contingencies. We refer to this as an extinction-like procedure because, similar to extinction tasks in EC, it involves the removal of the environmental event that underlies the target evaluation (in this case the common element shared by regularities). [[4]](#footnote-4) Because it proved difficult to consistently extinguish evaluations using such a task, we then decided (in Experiment 4) to use an alternative procedure that has worked in the EC literature (non-contingent stimulus presentations). Once again, evaluations failed to extinguish. In Experiment 5 we turned our attention to counterconditioning and attempted to do so by replacing the valenced source stimulus in one contingency with a stimulus of the opposite valence during the counterconditioning phase. Given the success of this manipulation we then tried to countercondition evaluations, not by changing the valence of the source stimuli, but by ‘reconfiguring the intersection’ itself (i.e., Experiments 6-7). Experiment 7 also tested the idea that there may have been a hidden intersection in our earlier studies that undermined the effectiveness of the extinction and counterconditioning manipulations.[[5]](#footnote-5)

***Examining the Robustness of Operant Evaluative Conditioning Effects***

Although our primary goal was to test the robustness of intersecting regularity effects, our design also allowed us to explore a second issue. As noted earlier, the source contingencies in our studies (i.e., the operant contingencies that contained the valenced source stimulus) also included a neutral outcome. Consequently, the valence of the outcome stimulus could change in-line with the valence of the source stimulus. Whereas changes in liking of the *target* stimulus qualify as instances of IR effects (i.e., effects of intersections between regularities), changes in liking of the neutral *outcome* are instances of operant evaluative conditioning (OEC; i.e., effects of a single stimulus-action-outcome contingency; De Houwer, 2007; Eder, Krishna, & Van Dessel, 2019). Put simply, OEC effects involve a change in liking that is due to the relationship between stimuli and responses in an operant contingency. Our studies offered an opportunity to examine the formation, extinction, and counterconditioning of OEC effects. As far as we know, this is the first time that extinction and counterconditioning of OEC has been examined. [[6]](#footnote-6)

In all of our studies, we assessed liking via self-report ratings, the Implicit Association Test, and a behavioral intention task. We added the IAT because it is assumed to capture more automatic instances of evaluation. The behavioral intention task might reflect a more ecologically valid index of liking. Prior research on evaluative learning via IR has produced effects on each of these measures (Hughes et al., 2016) and we expected similar outcomes here as well.

# ****Experiments 1-4: Extinction of OEC and IR effects****

Our initial goal was to establish new likes and dislikes for outcome stimuli (OEC effect) and target stimuli (IR effects), and once these evaluations were in place, to eliminate them. We did so by removing the outcome stimulus from (a) the contingency containing the valenced source stimulus (Experiment 1), (b) the contingency containing the neutral target stimulus (Experiment 2), or (c) both contingencies (Experiment 3). In Experiment 4, we tried to degrade the intersection even more by using an approach that has proven effective in extinguishing EC effects. Although results on extinction in EC are mixed, it seems that presenting a CS in isolation from a US after acquisition does lead to a small reduction in CS evaluations (i.e., CS-only presentations; Hofmann et al., 2010). In Experiment 4 we implemented a similar task. Doing so not only eliminates intersections between contingencies but also highlights that the elements within those contingencies (stimuli and responses) are no longer related. This may provide yet another signal that the contingencies, and thus intersections, no longer hold. If so, then evaluations may dissipate.

**Method**

***Participants and Design***

146 participants (93 male, *Mage* = 27.9, *SD* = 5.3) [Experiment 1], 108 participants (57 female, *Mage* = 29.7, *SD* = 7.2) [Experiment 2], 111 participants (66 female, *Mage* = 28.8, *SD* = 5.8) [Experiment 3], and 105 participants (54 male, *Mage* = 29.5, *SD* = 6.1) [Experiment 4] completed the study on the Prolific Academic website (https://prolific.ac) in exchange for a monetary reward.

A 2 (*Stimulus*: neutral stimuli related to positive vs. negative source) x 2 (*Training*: Extinction vs. Acquisition-only) mixed design was employed in Experiments 1-4 with the first factor measured within and the second measured between participants. Self-reported ratings, IAT effects, and behavioral intentions were the dependent variables. Three method factors were manipulated between participants: stimulus identity (whether outcome stimulus O1 and target stimulus T1 or outcome stimulus O2 and target stimulus T2 were assigned to positive/negative source stimuli), evaluative task order (self-report or IAT first) and IAT block order (learning consistent vs. inconsistent block first). [[7]](#footnote-7)

**Materials**

***Stimuli***

Two fictitious brand names (Morag and Struan) and two Chinese ideographs served as neutral outcome and target stimuli, respectively, during the acquisition and extinction phases. These stimuli were selected based on a pre-test conducted on a different sample of fifty-one participants (17 women, *Mage* = 26.22, *SD* = 5.15), forty seven of whom provided complete data and whose data was subsequently analyzed. These participants were asked to evaluate two separate sets of ten Chinese symbols and ten fictitious brands by rating them on a scale from -5 to 5. The two selected Chinese ideographs were both neutral in valence: one sample t-tests indicated that their average score did not differ from 0, *t*(47) = .67, *p* = .50 and, *t*(47) = 1.23, *p* = .23. A paired sample t-test indicated no differences in liking between the two, *t*(46) = -.33, *p* = .74. The two brand stimuli selected for use were the most neutral in valence, even though one did differ from 0, *t*(47) = 2.63, *p* = .01, and *t*(47) = 1.42, *p* = .16. Once again the two stimuli did not differ from one another in valence, *t*(46) = 1.19, *p* = .24. A further set of sixteen positive and sixteen negative food images were used as valenced stimuli. In the IAT, two Chinese symbols from the learning phase served as target labels and the words ‘*Good*’ and ‘*Bad*’ as attribute labels. Eight positively valenced and eight negatively valenced adjectives served as attribute stimuli (*delicious*, *tasty*, *nice*, *good*, *gorgeous*, *wonderful*, *yummy* and *pleasant* vs. *rotten*, *disgusting*, *nasty*, *horrid*, *sick*, *vomit*, *horrible*, *unpleasant*) while images of the two Chinese symbols served as target stimuli.

**Procedure**

Participants were provided with a general overview of the experiment, asked for their informed consent, and then told that they would encounter a number of brand products that had purportedly been released into the European marketplace. One group (acquisition-only) completed an acquisition phase and then proceeded directly to the evaluative measures. The other (extinction) completed the acquisition followed by an extinction phase, and only then the evaluative measures. Everyone then answered a series of exploratory questions. The entire session took approximately 30 minutes. See Figure 2 for an overview of the learning tasks used in Experiments 1-7.

***Acquisition Phase***

Prior to the learning task, participants were informed that they would see an image (either food or a Chinese symbol) in the middle of the screen. Their task was to identify the specific key (either ‘D’, ‘C’, ‘J’ or ‘N’) that the item was related to. They were asked to take their time and try to be as accurate as possible. Training consisted of four blocks of twenty trials (80 total). Each trial began with the presentation of a positively or negatively valenced food image (i.e., source stimulus [S1] or [S2]) or one of two Chinese symbols (i.e., target stimulus [T1] or [T2]). Selecting (R1) in the presence of a positive source (S1) or (R2) when presented with neutral target (T1) resulted in the removal of that stimulus from the screen, followed by a 250ms inter-stimulus interval, and the subsequent presentation of a neutral brand name (i.e., outcome stimulus O1). After an inter-trial interval of 1250ms the next trial began. Likewise, selecting (R3) in the presence of a negative source (S2) or (R4) when presented with neutral target (T2) resulted in the removal of that stimulus from the screen, an inter-stimulus interval, and the subsequent presentation of another brand name (outcome stimulus O2) (for an overview see Table 2). Stimulus-key assignments were counterbalanced between participants, such that one group categorized S1/T1 using R1/R2, whereas another group categorized S1/T1 using R3 and R4. If participants emitted an incorrect response then error feedback was displayed for 1500ms. During this time, participants could not emit another response and had to wait until the next trial commenced in order to try again. Following each block, participants were exposed to a feedback screen that displayed their percentage accuracy during the previous section of the task. Instructions emphasized the need for accurate responding especially if past performance was below 90%.

***Extinction Phase***

**Experiment 1***.* The extinction phase was similar to the acquisition phase (i.e., four blocks of 20 trials) with one exception. Once again, each trial began with the presentation of a positive (S1) or negative source (S2) or one of two neutral targets (T1 or T2). Selecting (R1) in the presence of a positive source (S1) resulted in the removal of that stimulus from the screen, but now, there was no subsequent presentation of an outcome. Selecting (R2) when presented with neutral target (T1) resulted in the removal of that stimulus from the screen followed by a 250ms inter-stimulus interval, and the presentation of outcome (O1). After an inter-trial interval of 1250ms the next trial began. Selecting (R3) in the presence of a negative source (S2) resulted in the removal of that stimulus from the screen but no presentation of an outcome. Pressing (R4) when presented with neutral target (T2) resulted in the removal of that stimulus from the screen, an inter-stimulus interval, and the presentation of outcome (O2) (see Figure 2). In case of an incorrect response an error feedback was displayed for 2000ms. During this time participants could not emit another response and had to wait until the next trial in order to try again.

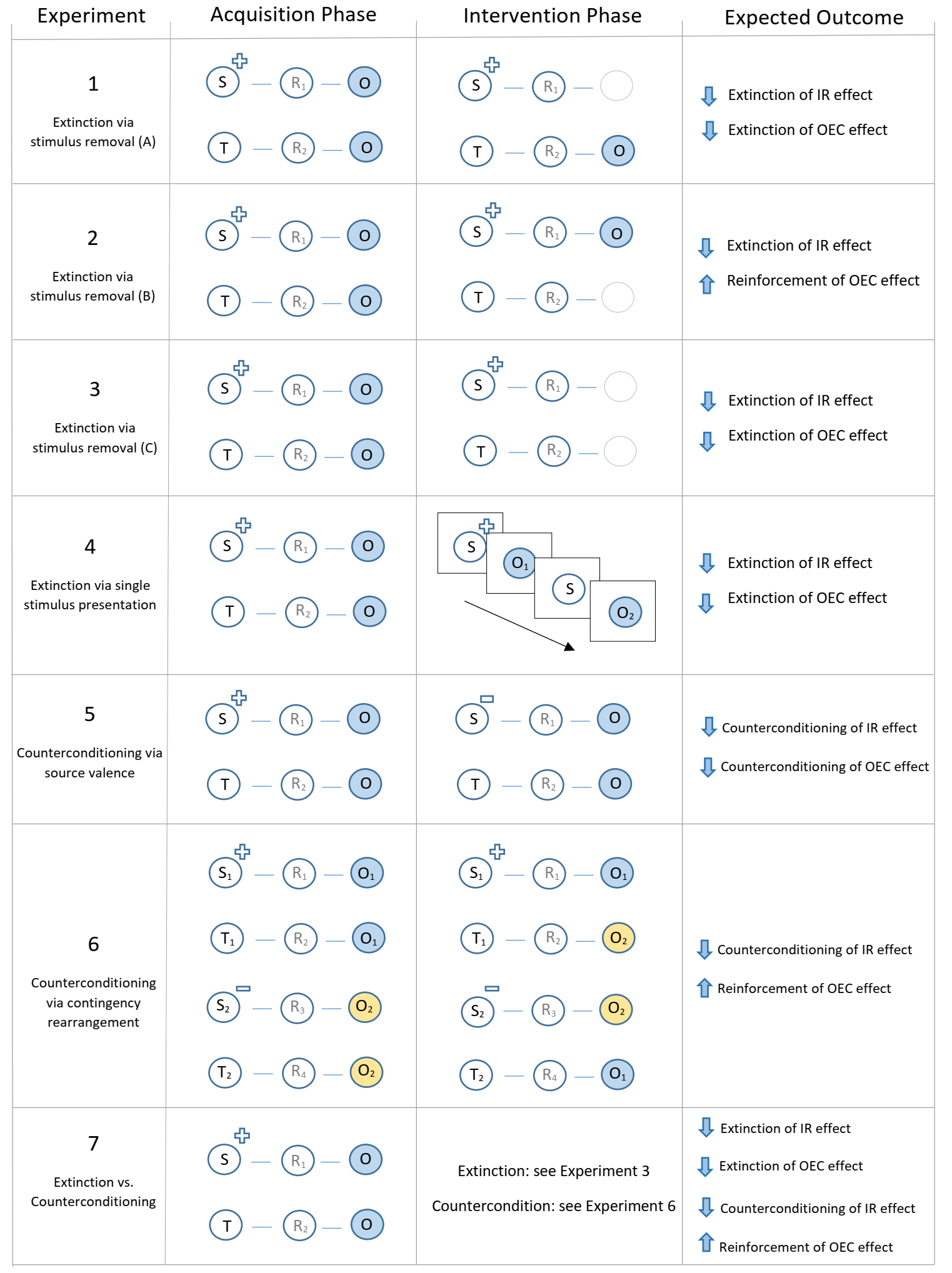
**Experiment 2***.* The extinction phase was similar to that used in Experiment 1 with one notable change. Whereas Experiment 1 attempted to extinguish evaluative responding by removing the outcome from the valenced contingency, Experiment 2 removed the outcome from the non-valenced contingency. Specifically, selecting (R1) in the presence of a positive source (S1) removed that stimulus from the screen, led to a 250ms inter-stimulus interval, and presentation of outcome (O1). Selecting (R2) when presented with neutral target (T1) removed that stimulus and was not followed by an outcome. Selecting (R3) in the presence of a negative source (S2) removed that stimulus from the screen, led to an inter-stimulus interval, and presentation of outcome (O2). Pressing (R4) when presented with neutral target (T2) was not followed by an outcome (see Table 3). [[8]](#footnote-8)

**Experiment 3***.* We now attempted to extinguish evaluative responding by removing the common intersection (outcome) from both contingencies. Selecting (R1) in the presence of a positive source (S1) or (R2) when presented with neutral target (T1) removed that stimulus from the screen, led to a 250ms inter-stimulus interval, and was not followed by an outcome. Selecting (R3) in the presence of a negative source (S2) or (R4) when presented with neutral target (T2) was also not followed by an outcome.

**Experiment 4***.* The extinction phase consisted of 4 blocks of 20 trials each. Participants were told that they would complete a second task wherein they would only have to observe a stream of stimuli. Each trial involved the presentation of a stimulus (T1, O1, T2, O2) for 1500ms and each stimulus was presented five times per block. After an inter-trial interval of 1500ms the next trial began. No categorization response was required during this phase. Each type of stimulus was presented with equal frequency within each block.

**Figure 2**

*Schematic overview of the procedures and expected effects in Experiments 1-7.*



*Note*. For illustration purposes Figure 2 only display one set of contingencies for most experiments (i.e., the ‘positively valenced’ contingencies). However, each experiment also contained another set of ‘negatively valenced’ contingencies as well.

**Evaluative Measures**

***IAT***

An IAT was administered to measure relative automatic evaluations of the target stimuli. Participants were informed that the two Chinese symbols (T1 and T2) they had encountered during the learning phase (targets) as well as the words ‘Good’ and ‘Bad’ (attributes) would appear on the upper left and right sides of the screen. During each trial a stimulus related to one of those categories would appear in the middle of the screen and they had to assign it to its corresponding category using either the left (‘E’) or right keys (‘I’). If they categorized the image or word correctly the stimulus disappeared from the screen and the next trial began. In contrast, an incorrect response resulted in the presentation of a red ‘X’ which remained on-screen until the correct key was pressed.

Overall, each participant completed seven blocks of trials. The first block of 20 practice trials required them to sort the target stimuli into their respective categories, with one target (T1) assigned to the left (‘E’) key and the other (T2) with the right (‘I’) key. On the second block of 20 practice trials, participants assigned positively valenced stimuli to the ‘Good’ category using the left key and negative stimuli to the ‘Bad’ category using the right key. Blocks 3 (20 trials) and 4 (40 trials) involved a combined assignment of target and attribute stimuli to their respective categories. Specifically, participants categorized the first target (T1) and ‘positive’ words using the left key and the second target (T2) and ‘negative’ words using the right key. The fifth block of 20 trials reversed the key assignments, with target (T1) now assigned to the right key and target (T2) with the left key. The sixth (20 trials) and seventh blocks (40 trials) required participants to categorize target (T1) with ‘negative’ words and target (T2) with ‘positive’ words.

***Self-Report Measures***

Ratings of the two outcome (Brand names: O1 and O2) and target stimuli (Chinese symbols: T1 and T2) were obtained using a series of Likert scales. On each trial, participants were presented with a stimulus and asked to indicate whether they considered it to be ‘*Good/Bad*’, ‘*Pleasant/Unpleasant*’, ‘*Positive/Negative*’ and whether ‘*I like it/I don’t like it*’ using a scale ranging from -5 to +5 with 0 as a neutral point. They also indicated how confident they were in their evaluations from -5 (*not confident*) to +5 (*confident*).

***Behavioral Intention Task***

This task was comprised of two trials: one trial in which the two target stimuli appeared simultaneously onscreen, and another trial where the two outcome stimuli were presented. On the former trial the stimuli appeared as labels on two bottles of ice-tea while on the latter trial they appeared on two bottles of milk). Participants had to indicate, for each pair, which item they would choose if they encountered them in a supermarket. Five answers were possible (i.e., “*I would choose product A*”, “*I would choose product B”*, “*I would choose both of them*”, “*I would choose neither of them*” or “*I don’t know”*).

**Exploratory Questions**

Participants completed a *memory test* which twice assessed their ability to recall the various elements of the intersecting regularities: once after the acquisition phase and then again after the extinction phase. The memory task consisted of eight trials. Four trials probed for the trained Source/Target stimulus 🡪Response relations (e.g., “…*when [Source/Target Stimulus 1/2] was presented which button did you have to press?*”) and provided participants with six response options (i.e., “R1, R2, R3, or R4”, “I don’t know” and “None of the above”). Another four trials probed for the Response 🡪 Outcome relations (e.g., “*…when you pressed R1/R2/R3/R4 what appeared onscreen*”) and provided them with four options (i.e., O1, O2, “I don’t know” and “None of the above”). No feedback was provided for any response emitted during this task. Participants who produced a minimum of 6 out of 8 trials were defined as having passed the memory test while those who failed to do so were defined as having failed the task.

They also completed *believability*, *demand compliance*, and *reactance* measures. These latter questions were asked after the evaluative measures, were included for exploratory purposes, and are therefore not mentioned in subsequent analyses.

**Results**

**Participant Exclusions**

We screened-out participants who (a) failed to complete the entire experimental session and thus provided incomplete data and/or (b) who had IAT error rates above 30% across the entire task, above 40% for any one of the four critical blocks, or who complete more than 10% of trials faster than 400ms (*n* = 49 [Experiment 1], *n* = 14 [Experiment 2], *n* = 16 [Experiment 3], and *n* = 7 [Experiment 4]). This led to a final sample of 97 participants in Experiment 1, 94 in Experiment 2, 95 in Experiment 3, and 98 in Experiment 4.

**Data Preparation**

Self-report ratings were collapsed into four mean scores – one for the target (T1), and another for the outcome (O1) related to positive sources, a third for the target (T2) and a fourth for the outcome (O2) related to negative sources. Two difference scores were then computed – one for the target stimuli (IR effect) and another for the outcome stimuli (OEC effect). Response latency data from the IAT were prepared using the D2 algorithm recommended by Greenwald et al. (2003). IAT scores reflect the difference in mean response latency between the critical blocks divided by the overall variation in those latencies. Scores were calculated so that positive values reflected a preference for the target that was indirectly related to a positive source (T1) relative to that related to a negative source (T2). Negative values indicated the opposite.

**Analytic Plan**

We examined if behavioral intentions, self-reported and automatic stimulus evaluations (*dependent variables*) differed as a function of the type of training received (extinction vs. acquisition-only) (*independent variable*). A series of one-way ANOVAs and *t*-tests were carried out on the rating and IAT data. Behavioralintention data were entered into a logistic regression with positive (based on the acquisition phase) stimulus selection as the reference category. Only results from the T1-T2 comparison are reported (i.e., analyses do not refer to the selections of neither or both targets, or non-responses). Counts of each response for each study and experiment condition were calculated, which were then used to calculate an odds ratio. *p* values were computed via Fischer’s exact test. Haldane-Anscombe corrections were applied to studies where at least one cell contained zero counts (i.e., counts in all cells were increased by 1).

**Hypothesis Testing**

We focused on three questions. First, did participants demonstrate evidence of learning during the acquisition and extinction phases? If so, then they should respond with a high rate of accuracy (we labelled those who responded with greater than 75% accuracy during the final block of training or testing as having “passed” that phase and those who did not as having “failed”). Second did they demonstrate evidence of *evaluative* learning? If so then we would expect to observe an OEC effect (i.e., a preference for the outcome stimulus related to positive over negative sources) and an IR effect (i.e., a preference for the target stimulus related to positive over negative sources) when we examine the data from participants in the acquisition-only group. Third, did the extinction procedures implemented in Experiments 1-4 undermine newly established evaluations? If so, then we would expect to observe a significant decrease in the magnitude of evaluative responding in the extinction relative to acquisition-only group. [[9]](#footnote-9)

***Question 1: How Did Participants Perform During the Acquisition & Extinction Phases?***

As can be seen from Table 1 participants responded with a high degree of accuracy during each phase of the learning task. The vast majority also met the necessary criterion to be labelled as having “passed” a given phase of the learning task (see Table 2). One notable exception was the extinction testing phase in studies where the outcome stimulus was removed from both contingencies (Experiments 3 and 7). This is despite the fact that those same participants had little difficulty passing the extinction training phase in those same experiments.

**Table 1**

*Mean (and standard deviation) accuracy as a function of learning task type (acquisition, extinction, or counterconditioning training or testing) in Experiments 1-7.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Exp** | **Acquisition training** | **Acquisition test** | **Extinction training** | **Extinction testing** | **Counterconditioning training** | **Counterconditioning testing** |
| 1 | 88 (22) | 85 (23) | 96 (12) | 82 (15) | / | / |
| 2 | 93 (15) | 88 (21) | 98 (8) | 89 (12) | / | / |
| 3 | 91 (17) | 83 (20) | 97 (13) | 61 (22) | / | / |
| 4 | 93 (17) | 85 (22) | / | / | / | / |
| 5 | 93 (15) | 87 (23) | / | / | 97 (6) | 87 (23) |
| 6 | 95 (11) | 90 (14) | / | / | 98 (5) | 90 (14) |
| 7 | 89 (21) | 80 (23) | 97 (11) | 75 (25) | 97 (12) | 93 (16) |

**Table 2**

*Percentage of participants who passed each section of the learning task (acquisition, extinction, counterconditioning) in Experiments 1-7.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Exp** | **Acquisition training** | **Acquisition test** | **Extinction training** | **Extinction testing** | **Counterconditioning training** | **Counterconditioning testing** |
| 1 | 83 | 81 | 96 | 91 | / | / |
| 2 | 91 | 84 | 97 | 100 | / | / |
| 3 | 88 | 78 | 98 | 28 | / | / |
| 4 | 90 | 84 | / | / | / | / |
| 5 | 91 | 85 | / | / | 98 | 85 |
| 6 | 93 | 91 | / | / | 100 | 91 |
| 7 | 85 | 75 | 97 | 52 | 96 | 91 |

*Note*. Counterconditioning was not provided in Experiments 1-4 nor was Extinction provided in Experiments 5-6. The type of extinction procedure used in Experiment 4 did not involve collection of training and testing data.

***Question 2****:* ***Did Evaluative Learning Take Place?***

**Operant Evaluative Conditioning**. OEC effects emerged in all four studies. Participants self-reported that they liked O1 (the outcome that was part of a contingency with positive sources) and disliked O2 (the outcome that was part of a contingency with negative sources), Experiment 1: *t*(47.93) = 4.91, *p* < .0001, *d* = 1.31, 95% CI [0.68, 1.93], BF10 = 689; Experiment 2: *t*(35.56) = 5.85, *p* < .0001, *d* = 1.67, 95% CI [1.04, 2.3], BF10 > 104; Experiment 3: *t*(43.97) = 6.08, *p* < .0001, *d* = 1.74, 95% CI [1.05, 2.43], BF10 > 104; Experiment 4: *t*(40.05) = 6.24, *p* < .0001, *d* = 1.81, 95% CI [1.14, 2.49], BF10 > 104. Likewise, the odds of selecting O1 were higher than those of selecting O2 in the behavioral intentions task: Experiment 1 (OR = 69.33, 95% CI [6.43, 748.06], *p* < .0001); Experiment 2 (OR = 55.25, 95% CI [5.5, 555.07], *p* < .0001); Experiment 3 (OR = 80, 95% CI [6.39, 1001.41], *p* < .0001); Experiment 4 (OR = 13.22, 95% CI [1.4, 124.91], *p* = .01).

**Intersecting Regularities**. IR effects also emerged across studies. Participants self-reported that they liked T1 (the target that intersected with a contingency containing positive sources) and disliked T2 (the target that intersected with a contingency containing negative sources), Experiment 1: *t*(47.55) = 1.9, *p* = .06, *d* = 0.53, 95% CI [-0.05, 1.11], BF10 = 1.2; Experiment 2: *t*(45.88) = 2.61, *p* = .012, *d* = 0.72, 95% CI [0.16, 1.28], BF10 = 4.7; Experiment 3: *t*(40.34) = 5.68, *p* < .0001, *d* = 1.67, 95% CI [0.99, 2.35], BF10 > 104; Experiment 4: *t*(48.24) = 3.03, *p* = .004, *d* = 0.83, 95% CI [0.23, 1.42], BF10 = 7.8. IAT scores demonstrated evidence for a relative preference for T1 over T2: Experiment 1: *t*(44.43) = 3.78, *p* < .001, *d* = 1.07, 95% CI [0.47, 1.67], BF10 = 75; Experiment 2: *t*(50.19) = 1.91, *p* = .06, d = 0.52, 95% CI [-0.03, 1.07], BF10 = 1.2; Experiment 3: *t*(36.27) = 4.6, *p* < .0001, *d* = 1.39, 95% CI [0.73, 2.04], BF10 = 856; Experiment 4: *t*(39.6) = 3.39, *p* = .002, *d* = 0.99, 95% CI [0.38, 1.59], BF10 = 29. Finally, the odds of selecting T1 were higher than those of selecting T2 in the behavioral intentions task in two of the four studies: Experiment 1 (OR = 2.6, 95% CI [0.6; 11.31], *p* = .28); Experiment 2 (OR = 15.12, 95% CI [2.28; 100.32], *p* = .003); Experiment 3 (OR = 40, 95% CI [3.56; 450], *p* < .001); Experiment 4 (OR = 2.89, 95% CI [0.69; 12.12], *p* = .17).

***Question 3: Was Evaluative Learning Moderated by the Extinction Procedures?***

**Operant Evaluative Conditioning**. Self-reported ratings decreased in magnitude (relative to the acquisition-only group) when the outcome stimulus was removed from both contingencies (Experiment 3), *t*(92.88) = -2.14, *p* = .04, *d* = -0.44, 95% CI [-0.85, -0.03], BF10 = 1.6. There was no difference between extinction and acquisition-only groups when the outcome was only removed from the source contingency (Experiment 1): *t*(93.69) = 0.87, *p* = .39, *d* = 0.18, 95% CI [-0.23, 0.58], BF10 = 0.3, or when stimuli were presented in a non-contingent manner (Experiment 4), *t*(93.74) = -1, *p* = .32, *d* = -0.2, 95% CI [-0.61, 0.2], BF10 = 0.3. Ratings increased in magnitude (as expected) when the outcome stimulus remained in the source contingency and was removed from the target contingency (Experiment 2): *t*(81.01) = 4.48, *p* < .0001, *d* = 0.94, 95% CI [0.5, 1.38], BF10 = 922.

Similarly, behavioral intentions did not differ between the extinction and acquisition-only groups in Experiment 1 (OR = 1.36, 95% CI [0.49, 3.76], *p* = .61), Experiment 2 (OR = 0.55, 95% CI [0.19, 1.54], *p* = .30), Experiment 3 (OR = 0.78, 95% CI [0.26, 2.31], *p* = .78), or Experiment 4 (OR = 0.62, 95% CI [0.19, 2.1], *p* = .55).

**Intersecting Regularities**. No decrease in the magnitude of self-reported ratings (relative to the acquisition-only group) occurred when the outcome was removed from the source contingency (Experiment 1): *t*(93.99) = 1.38, *p* = .17, *d* = 0.28, 95% CI [-0.13, 0.69], BF10 = 0.5, both contingencies (Experiment 3): *t*(92.82) = -1.82, *p* = .07, *d* = -0.37, 95% CI [-0.78, 0.04], BF10 = 0.9, or when stimuli were presented in a non-contingent manner (Experiment 4): *t*(90.04) = 0.59, *p* = .56, *d* = 0.12, 95% CI [-0.28, 0.52], BF10 = 0.2. Ratings increased in magnitude when the outcome was only removed from the target contingency (Experiment 2): *t*(74.64) = 2.59, *p* = .012, *d* = 0.56, 95% CI [0.13, 0.98], BF10 = 4.6.

IAT scores also did not differ between the extinction and acquisition-only groups in Experiment 1: *t*(85.72) = -0.77, *p* = .44, *d* = -0.16, 95% CI [-0.56, 0.25], BF10 = 0.3; Experiment 2: *t*(84.11) = 0.46, *p* = .65, *d* = 0.1, 95% CI [-0.32, 0.51], BF10 = 0.2; Experiment 3: *t*(90.64) = -0.46, *p* = .65, *d* = -0.09, 95% CI [-0.5, 0.31], BF10 = 0.2; or Experiment 4: *t*(91.22) = -0.6, *p* = .5499, *d* = -0.12, 95% CI [-0.52, 0.28], BF10 = 0.3.

Finally, behavioral intentions did not differ between the extinction and acquisition-only groups in Experiment 1: (OR = 0.82, 95% CI [0.28; 2.358], *p* = .79); Experiment 2: (OR = 0.91, 95% CI [0.31; 2.68], *p* = .1); Experiment 3: (OR = 0.52, 95% CI [0.17; 1.61], *p* = .28); Experiment 4: (OR = 1.18, 95% CI [0.41; 3.37], *p* = .79).

**Discussion**

In Experiments 1-4 people encountered an acquisition phase wherein an operant contingency containing a valenced source intersected with a contingency containing a neutral target (i.e., the two contingencies shared a common outcome stimulus). This phase was designed to establish novel evaluations towards outcome (OEC effect) and target stimuli (IR effect). Half of the participants then completed a second phase which removed the intersecting element from one contingency (Experiments 1-2), both contingencies (Experiment 3), or presented the stimuli in a non-contingent manner (Experiment 4), to see if this would reduce or eliminate evaluations.

Results indicated that the acquisition phase gave rise to OEC and IR effects. However, the various ‘extinction’ procedures failed to reduce or eliminate those evaluations (with one exception: removing the outcome from both contingencies reduced OEC effects, although this reduction appears to be a weak one). The absence of extinction is particularly noteworthy given the variety of procedures used, each of which eliminated the intersection present during the acquisition phase. Likewise, the absence of a reduced effect in Experiment 4 is also noteworthy given that this extinction procedure has been found to successfully extinguish EC effects (see Hofmann et al., 2010, for a meta-analysis).

# Experiments 5-6: Counterconditioning

Given the difficulty of undoing evaluations established via operant evaluative conditioning and intersecting regularities, we changed direction in Experiments 5-6, and instead sought to revise likes and dislikes using counterconditioning procedures. Once again participants completed an acquisition phase. Afterwards one group moved directly to the evaluative measures while a second first completed a counterconditioning task. In Experiment 5 this involved replacing the valenced source stimulus in one operant contingency with a stimulus of the opposite valence. In Experiment 6 this involved counterconditioning via ‘contingency rearrangement’ (*see below*).

**Method**

***Participants***

109 participants (69 male, *Mage* = 28.5, *SD* = 5.9) (Experiment 5) and 106 participants (57 female, *Mage* = 29.7, *SD* = 5.6) (Experiment 6) took part via Prolific Academic in exchange for a monetary reward.

**Procedure**

Overall, the study consisted of four phases: acquisition, counterconditioning, evaluative measures, and exploratory questions. These phases were similar to those reported in Experiments 1-4 unless otherwise stated.

***Counterconditioning***

**Experiment 5**. The counterconditioning phase was similar to the acquisition phase with one notable exception:the assignment of valence source stimuli was reversed. Selecting (R1) in the presence of a negative source (S2), or (R2) when presented with neutral target (T1), resulted in the presentation of outcome (O1). Selecting (R3) in the presence of a positive source (S1), or (R4) in the presence of neutral target (T2), resulted in the presentation of outcome (O2) (see Figure 2).

**Experiment 6**. The counterconditioning procedure involved ‘contingency rearrangement’ and consisted of four blocks of 20 trials (80 trials total). Each trial began with the presentation of a positive (S1) or negative (S2) source, or a neutral target (T1 or T2). Selecting (R1) in the presence of a positive source (S1) removed it from the screen, produced a 250ms ITI, and led to the presentation of outcome (O1). Selecting (R2) when presented with target stimulus (T1) resulted its removal, an ITI, and the presentation of outcome (O2). Selecting (R3) in the presence of a negative source (S2) resulted in its removal, an ITI, and the presentation of outcome (O2). Selecting (R4) when presented with neutral target (T2) removed it from the screen and led to outcome (O1).

In short, we sought to rearrange the contingencies so that a ‘neutral’ contingency (Neutral Target 1 🡪 R2 🡪 **Neutral Outcome 2**) which previously intersected with a ‘positively valenced’ contingency (Positive Source [S1] 🡪 R1 🡪 Neutral Outcome 1) now intersected with a ‘negatively valenced’ contingency (Negative Source [S2] 🡪 R3 🡪 **Neutral Outcome 2**). We did the same with the other two contingencies (i.e., made a ‘neutral contingency’ that originally intersected with a ‘positively valenced contingency’ during acquisition now intersect with a ‘negatively valenced contingency’ during counterconditioning) (see Figure 2). [[10]](#footnote-10)

**Results**

**Participant Exclusions**

Participants with incomplete data or who had excessive error or speed rates were excluded (*n* = 14 in Experiment 5 and *n* = 16 in Experiment 6). This resulted in a final *n* = 95 in Experiment 5 and *n* = 90 in Experiment 6.

**Hypothesis Testing**

We once again asked three questions. First, did participants demonstrate evidence of learning during the acquisition and counterconditioning phases? Second, did they demonstrate evidence of *evaluative* learning? Third, did the counterconditioning procedures undermine newly established evaluations? If so, we would expect a significant decrease in the magnitude of evaluative responding in the counterconditioning relative to acquisition-only group.

***Question 1: How Did Participants Perform During the Acquisition & Counterconditioning Phases?***

As can be seen from Table 1 participants responded with a high degree of accuracy during each phase of the learning task. The vast majority also met the necessary criterion to be labelled as having “passed” a given phase of the learning task (see Table 2).

***Question 2****:* ***Did Evaluative Learning Take Place?***

**Operant Evaluative Conditioning**. OEC effects emerged in both studies. Participants self-reported that they liked O1 (the outcome that was part of a contingency with positive sources) and disliked O2 (the outcome that was part of a contingency with negative sources), Experiment 5: *t*(42.86) = 5.54, *p* < .0001, *d* = 1.64, 95% CI [0.95, 2.33], BF10 = 8148; Experiment 6: *t*(35.56) = 3.27, *p* = .002, *d* = 1.03, 95% CI [0.35, 1.71], BF10 = 15. The odds of selecting O1 were also higher than those of selecting O2 in the behavioral intentions task in Experiment 5 (OR = 10.5, 95% CI [2.15, 51.28], *p* = .005), and Experiment 6 (OR = 42, 95% CI [3.2, 551.57], *p* = .002).

**Intersecting Regularities**. IR effects emerged in both studies. Participants self-reported that they liked T1 (the target that intersected with a contingency containing positive sources) and disliked T2 (the target that intersected with a contingency containing negative sources) in Experiment 5: *t*(41.2) = 3.15, *p* = .003, *d* = 0.94, 95% CI [0.31, 1.57], BF10 = 14; and Experiment 6: *t*(39) = 3.17, *p* = .003, *d* = 0.95, 95% CI [0.27, 1.62], BF10 = 9. IAT scores also demonstrated evidence for a relative preference for T1 over T2: Experiment 5: *t*(37.19) = 4.45, *p* < .0001, *d* = 1.35, 95% CI [0.69, 2.01], BF10 = 474; Experiment 6: *t*(32.15) = 3.61, *p* = .001, *d* = 1.17, 95% CI [0.47, 1.86], BF10 = 42. Finally, the odds of selecting T1 were higher than those of selecting T2 in the behavioral intentions task in Experiment 5 (OR = 26.67, 95% CI [4.64; 153.22], *p* < .001), but not in Experiment 6 (OR = 5.6, 95% CI [0.81; 38.51], *p* = .09).

***Question 3: Was Evaluative Learning Moderated by the Counterconditioning Procedures?***

**Operant Evaluative Conditioning**. Self-reported ratings decreased in magnitude (relative to the acquisition-only group) when counterconditioning involved reversing the valence of the source stimulus (Experiment 5), *t*(85.69) = -5.17, *p* < .0001, *d* = -1.07, 95% CI [-1.51, -0.63], BF10 = 12450. They increased in magnitude, as expected, in the contingency rearrangement condition, which involved additional exposure to operant evaluative conditioning, *t*(86.47) = 2.18, *p* = .032, *d* = 0.46, 95% CI [0.03, 0.89], BF10 = 1.7. Behavioral intentions did not differ between the counterconditioning and acquisition-only groups in Experiment 5, OR = 1.6, 95% CI [0.57, 4.46], *p* = .44, or Experiment 6, OR = 2.77, 95% CI [0.92, 8.32], *p* = .1.

**Intersecting Regularities**. Self-reported ratings decreased in magnitude (relative to the acquisition-only group) when counterconditioning involved the reversal of source stimulus valence (Experiment 5), *t*(64.79) = -3.26, *p* = .002, *d* = -0.68, 95% CI [-1.1, -0.26], BF10 = 24, but not when contingency rearrangement took place (Experiment 6), *t*(81.53) = -0.84, *p* = .41, *d* = -0.18, 95% CI [-0.6, 0.24], BF10 = 0.3. IAT scores did not differ between the counterconditioning and acquisition-only groups in Experiment 5 *t*(91.05) = -1.87, *p* = .06, *d* = -0.39, 95% CI [-0.8, 0.03], BF10 = 1, orExperiment 6: *t*(76.52) = -0.85, *p* = .39, *d* = -0.18, 95% CI [-0.61, 0.24], BF10 = 0.3. Finally, behavioral intentions did not differ between the counterconditioning and acquisition-only groups in Experiment 5: (OR = 1.26, 95% CI [0.43; 3.71], *p* = .79), or Experiment 6: (OR = 0.78, 95% CI [0.26; 2.34], *p* = .78).

**Discussion**

Experiments 5-6 exposed participants to an acquisition phase designed to establish novel evaluations towards outcome (OEC effect) and target stimuli (IR effect). Half of the participants then completed a second phase that sought to countercondition those evaluations via stimulus valence reversal (Experiment 5) or contingency rearrangement (Experiment 6). Results indicated that the acquisition phase gave rise to OEC and IR effects. Interestingly, while counterconditioning via stimulus reversal significantly decreased the magnitude of self-reported evaluations (Experiment 5) counterconditioning via contingency rearrangement did not do so for IR effects. In short, it appears that OEC and IR effects can be counterconditioned via stimulus valence reversal.

**Experiment 7: Extinction vs. Counterconditioning**

In attempting to explain the resistance of OEC and IR effects to extinction and counterconditioning we identified one possibility: many of the studies reported here involved not only a ‘visible’ intersection (the outcome) but also a ‘hidden’ intersection (response location). Specifically, during training participants categorized one of the valenced sources and a neutral target using keys located on ‘left’ side of the keyboard (e.g. D or C). They also categorized the other valenced source and neutral target using keys on the ‘right’ side of the keyboard (e.g. J or N). Thus, stimuli not only intersected in terms of a common outcome but also in terms of a common response feature (use of left or right hand). This second intersection was still present during certain extinction phases (e.g., in Experiments 1-3 but not 4) and partially during the counterconditioning phase of Experiment 6 (but not Experiment 5 where we observed counterconditioning). Thus, even when certain outcome stimuli were no longer presented, and the intersection changed, participants still used the same hands to respond to S1 and T1 (left hand) and S2 and T2 (right hand). It may be that stronger extinction and counterconditioning effects emerge when both intersections are eliminated. We examined this possibility in Experiment 7. We were also interested in comparing the relative effectiveness of extinction or countercondition in changing IR effects. We therefore recruited three groups of participants and exposed them to either (a) only the acquisition phase, (b) acquisition and then extinction, or (c) acquisition and then counterconditioning.

**Method**

**Participants and Design**

Three hundred and eighty six participants (222 women, *M age* = 29.1, *SD* = 5.8) took part in an online experiment via Prolific Academic in exchange for a monetary reward.

**Procedure**

Participants completed an acquisition phase, and either proceeded to the evaluative measures (acquisition-only) or first completed an extinction or counterconditioning task.

***Acquisition Phase***

The structure of the acquisition phase was similar to that administered in Experiments 1-6 with two exceptions: participants now emitted a response using a mouse rather than keyboard and the location of the responses varied randomly across trials (thereby ensuring no common response location could emerge). The four response options (D, C, J, and N) were printed onscreen below the stimulus on each trial. Clicking on one of the four letters with the mouse led to the removal of the stimulus, a short (250ms) intra-trial interval, and finally the outcome stimulus. Pilot testing indicated that participants found this version of the task to be difficult. We therefore provided a fifth block of trials in situations where they emitted less than 80% correct responses during the fourth block.

***Extinction Phase***

A similar extinction phase was used as in Experiment 3 with three exceptions: we changed the nature of responding (mouse instead of key-press), randomized the location of response options across trials, and provided a fifth block of trials in situations where they emitted less than 80% correct responses during the fourth block.

***Counterconditioning Phase***

A similar counterconditioning phase was used as in Experiment 6 with two exceptions: we changed the nature of responding (mouse instead of key-press), randomized the location of response options across trials, and provided a fifth block of trials in situations where they emitted less than 80% correct responses during the fourth block. Once again this counterconditioning phase was expected to countercondition IR effects and boost OEC effects.

***Exploratory Questions***

Along with the other questions we also included a matching to sample (MTS) procedure. This task was included for exploratory purposes, delivered at the very end of the experiment, and will not be discussed further.

**Results**

**Participant Exclusions**

Participants with incomplete data or who had excessive IAT error or speed rates were excluded (*n* = 73). This led to a final sample of 313 participants.

**Hypothesis Testing**

We were interested in four questions. First, did participants demonstrate evidence of learning during the acquisition and intervention phases? Second, did they demonstrate evidence of *evaluative* learning? Third, did the extinction and/or counterconditioning procedures undermine newly established evaluations? Fourth, was counterconditioning or extinction more effective in doing so?

***Question 1: How Did Participants Perform During the Acquisition & Intervention Phases?***

As can be seen from Table 1 participants responded with a high degree of accuracy during each phase of the learning task. Most also met the criterion needed to be labelled as having “passed” a given phase of the learning task (see Table 2).

***Question 2****:* ***Did Evaluative Learning Take Place?***

**Operant Evaluative Conditioning**. OEC effects emerged such that participants self-reported liking O1 and disliking O2: *t*(98.77) = 5.43, *p* < .0001, *d* = 1.06, 95% CI [0.65, 1.48], BF10 = 33252. Behavioral intentions also favored O1 over O2, OR = 7.56, 95% CI [2.26, 25.22], *p* < .001.

**Intersecting Regularities**. IR effects emerged such that participants self-reported liking T1 and disliking T2: *t*(98.37) = 2.24, *p* = .028, *d* = 0.44, 95% CI [0.05, 0.83], BF10 = 1.9. IAT scores also demonstrated evidence of a relative preference for T1 over T2, *t*(101.79) = 3.92, *p* < .001, *d* = 0.77, 95% CI [0.37, 1.17], BF10 = 146. Behavioral intention did not favor T1 over T2, OR = 2, 95% CI [0.69, 5.76], *p* = .29.

***Question 3: Was Evaluative Learning Moderated by Extinction or Counterconditioning?***

**Operant Evaluative Conditioning**. Self-reported ratings did not decrease in magnitude when the outcome was removed from both contingencies (extinction), *t*(187.92) = 0.69, *p* = .49, *d* = 0.1, 95% CI [-0.18, 0.37], BF10 = 0.2. They increased in magnitude, as expected, following counterconditioning, which involved additional exposure to operant evaluative conditioning, *t*(204.38) = 2.23, *p* = .03, *d* = 0.31, 95% CI [0.03, 0.58], BF10 = 1.5. Behavioral intentions were not moderated by either the extinction, OR = 0.9, 95% CI [0.43, 1.88], *p* = .85, or counterconditioning procedures, OR = 0.61, 95% CI [0.3, 1.23], *p* = .21.

**Intersecting Regularities**. Neither self-reported ratings, *t*(202.2) = 1.56, *p* = .12, *d* = 0.22, 95% CI [-0.06, 0.49], BF10 = 0.5, nor IAT scores, *t*(205.09) = 1.11, *p* = .27, *d* = 0.15, 95% CI [-0.12, 0.43], BF10 = 0.3, nor behavioral intentions, OR = 1.1, 95% CI [0.53, 2.29], *p* = .85, differed in the extinction relative to acquisition-only group. Although self-reported ratings decreased in the counterconditioning (relative to acquisition-only) group, *t*(207) = -2.5, *p* = .01, *d* = -0.35, 95% CI [-0.62, -0.07], BF10 = 2.7, this was not the case for IAT scores, *t*(206.02) = 0.84, *p* = .40, *d* = 0.12, 95% CI [-0.16, 0.39], BF10 = 0.2, nor behavioral intentions, OR = 1.24, 95% CI [0.61, 2.53], *p* = .59.

***Question 4: Which was More Effective in Moderating Evaluations: Extinction or Counterconditioning?***

A series of paired t-tests revealed that the magnitude of IAT scores, *t*(203.3) = -0.34, *p* = .73, *d* = -0.05, 95% CI [-0.32, 0.23], BF10 = 0.2, OEC effects, *t*(198.46) = 1.24, *p* = .22, *d* = 0.17, 95% CI [-0.1, 0.45], BF10 = 0.3, behavioral intentions (IR), OR = 1.13, 95% CI [0.56, 2.29], *p* = .86, and behavioral intentions (OEC), OR = 0.67, 95% CI [0.34, 1.33], *p* = .29, did not differ in the extinction relative to counterconditioning groups. The only difference was found in self-reported IR effects, which were smaller in the counterconditioning relative to extinction group, *t*(203.14) = -3.91, *p* < .001, *d* = -0.54, 95% CI [-0.82, -0.26], BF10 = 167.

**Discussion**

Once again, OEC and IR effects emerged. An extinction procedure which removed the outcome stimulus from both contingencies did not influence the magnitude of these newly established evaluations. Likewise, a counterconditioning procedure which involved contingency rearrangement was only partially successful in that it reduced self-report, but not IAT scores or behavioral intentions. Directly comparing the impact of the extinction and counterconditioning procedures revealed that the latter decreased self-reported evaluations (but not IAT scores or behavioral intentions) to a greater extent than the former.

**Meta-Analyses**

We carried out a series of multilevel meta-analyses to ask three general questions about our findings that individual studies lacked the power to address or to make general conclusions from: (a) do OEC and IR procedures give rise to evaluations *in general*, (b) are evaluations moderated by extinction or counterconditioning *in general*, and (c) do those effects differ when we exclude participants who failed the learning task? Analyses were conducted using the metafor R package (Viechtbauer, 2010). All models employed a Restricted Maximum Likelihood estimator function. In each case, study was entered as a random intercept in order to acknowledge the non-independence of each study’s outcome variables, and outcome variable type (i.e., IAT, self-reported evaluations, behavioural intentions) was entered as a random slope in order to acknowledge that changes of different magnitudes may be observed between them. Prior to meta-analysis, behavioral intention data were converted from Odds Ratios to Cohen’s *d* scores using the method specified by Hasselblad and Hedges (1995; see also Sánchez-Meca, Marín-Martínez & Chacón-Moscoso, 2003) which has been shown to balance ease of use, bias, and coverage. Meta-analyses were not pre-registered, although the hypotheses assessed within them are similar to the those pre-registered in the individual experiments.

## **Question 1: Do OEC and IR Procedures Give Rise to Novel Evaluations in General?** Each of our studies employed multiple evaluative measures (self-reports, IATs, behavioral intentions). These measures were not included for theoretical reasons (e.g., to examine dissociations between automatic and non-automatic evaluations) but instead to provide convergent evidence for evaluative learning. We therefore wanted to know if operant evaluative conditioning and intersecting regularities gave rise to novel evaluations *in general* (i.e., regardless of the specific measure used). To answer this question we carried out multilevel meta-analyses of both the IR and OEC effects within the acquisition-only group (see Figure 3).

***Operant Evaluative Conditioning***

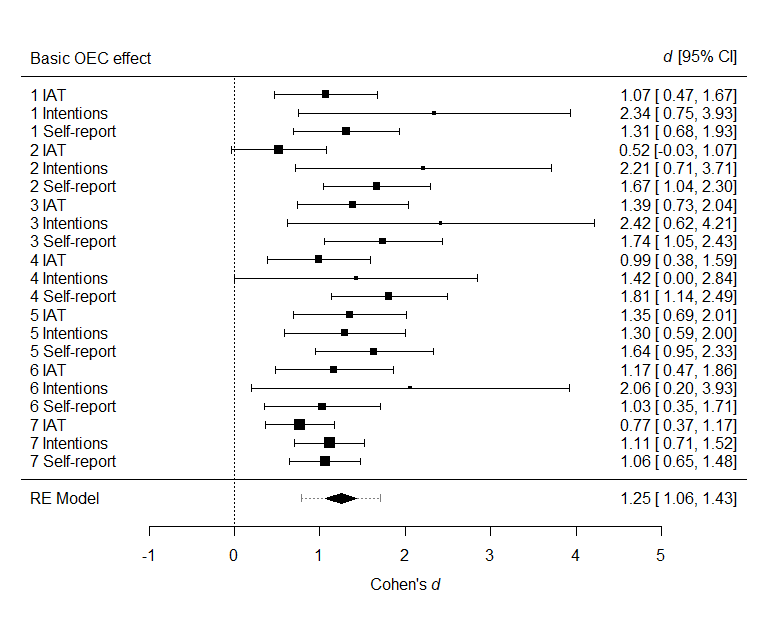
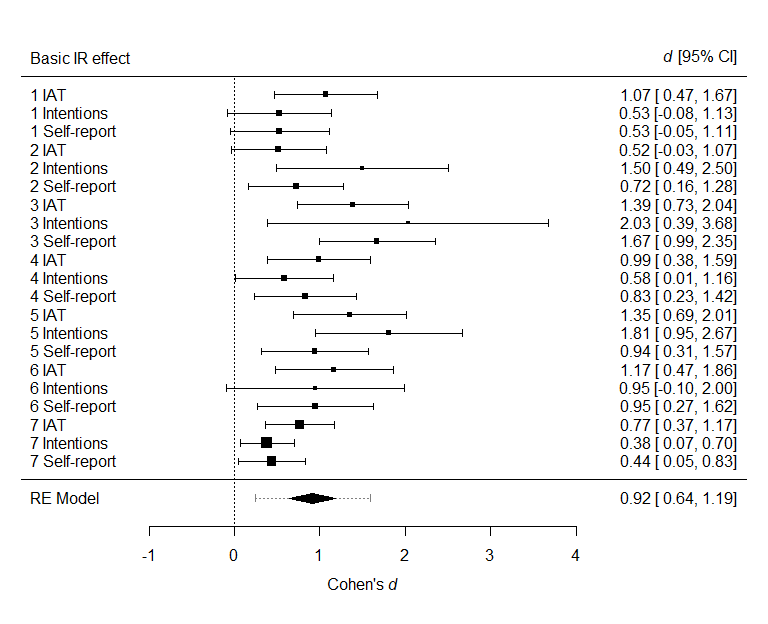
The meta-analytic model indicated that a change in liking takes place after OEC, *d* = 1.25, 95% CI [1.06, 1.43], *p* < .001.

***Intersecting Regularities***

The meta-analytic model indicated that a change in liking takes place after IR training, *d* = 0.92, 95% CI [0.64, 1.19], *p* < .001.

**Figure 3**

*Meta-analytic models outlining the IR and OEC effects. In each forest plot, squares represent observed Cohen’s d effect sizes, size of square represents weighting in the model, and error bars represent 95% Confidence Intervals (CIs) around the effect size.*



## **Question 2: Are Evaluations Moderated by Extinction or Counterconditioning?**

Four variants of extinction procedure and two counterconditioning procedures were implemented in Experiments 1-7. These interventions moderated evaluations in certain studies and failed to do so in others. The question remains: to what extent do “extinction” and “counterconditioning” moderate evaluations that were established via intersecting regularities *in general*? A moderator multilevel meta-analysis was conducted on the OEC and IR effects to answer this question.

***Extinction***

The meta-analytic model indicated that, in general, OEC effects, *d* = -0.01, 95% CI [-0.20, 0.19], *p* = XXX, and IR effects, *d* = .01, 95% CI [-0.13, 0.15], *p* = XXX, were not moderated by the extinction procedures used in this paper.

***Counterconditioning***

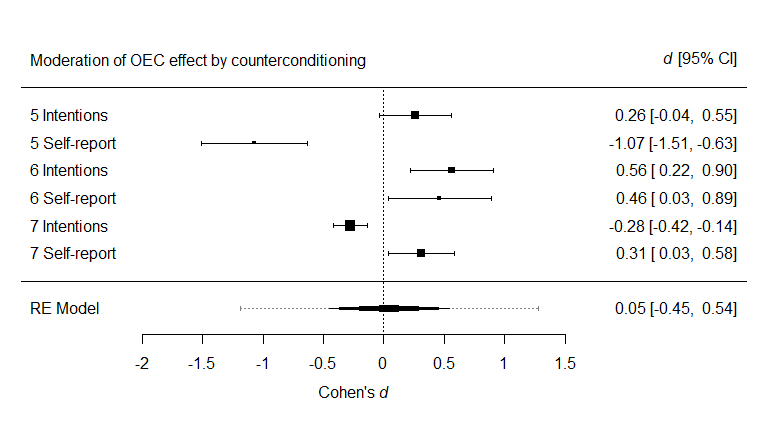
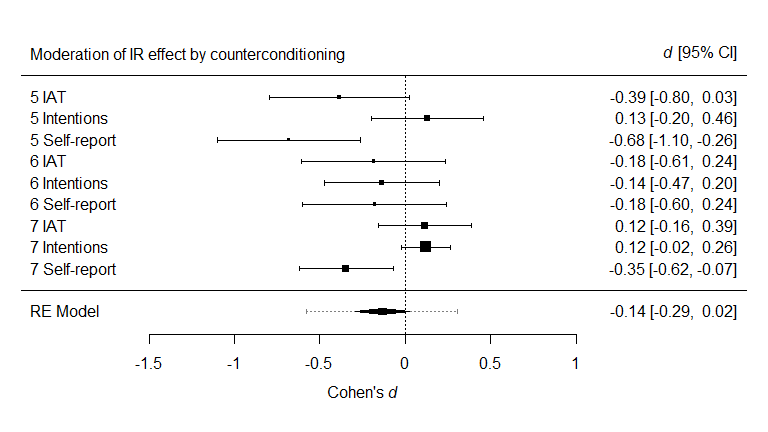
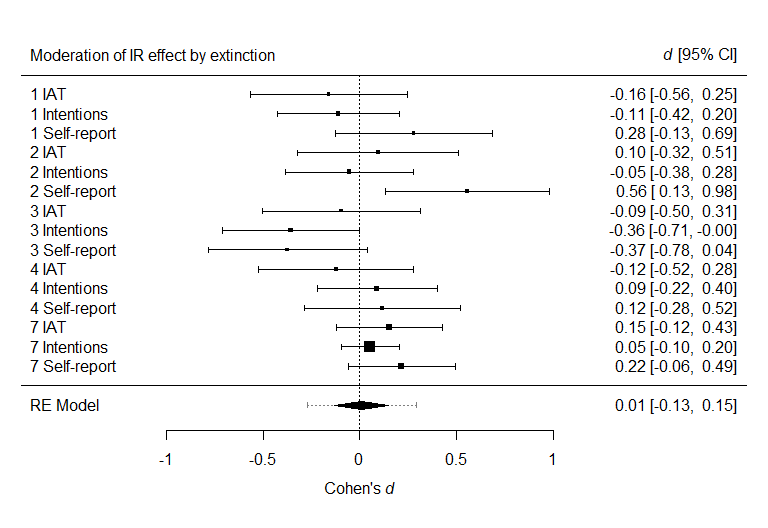
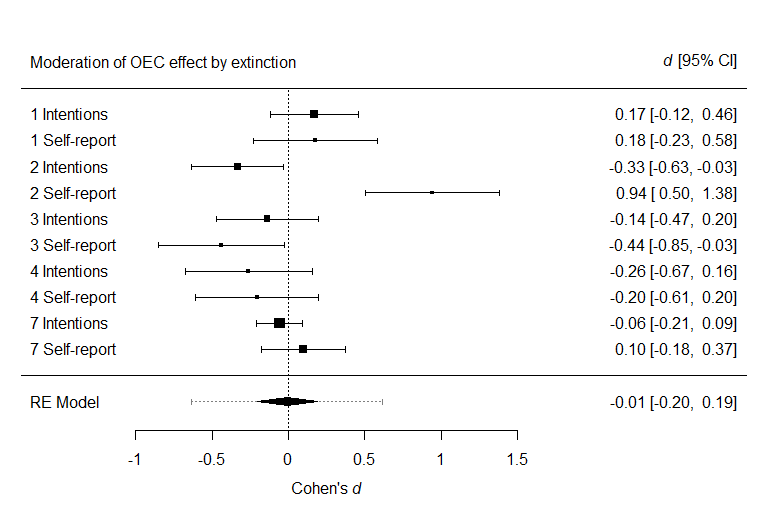
The meta-analytic model indicated that, in general, IR effects, *d* = -0.14, 95% CI [-0.29, 0.02], *p* = XXX, were not moderated by the extinction procedures used in this paper. [[11]](#footnote-11)

## **Question 3: Do Our Conclusions Change When Only Considering Those Who Passed the Learning Phase?**

Thus far we have analyzed the data of all participants regardless of their performance on the learning task. However, upon reflection, people who performed poorly during that task may be responsible for the absence of extinction and counterconditioning (i.e., if they did not discriminate the contingencies during the acquisition and intervention phases then it seems unlikely that evaluative effects will emerge or be later modified). With this in mind, we carried out a the same set of analyses as reported throughout this paper after excluding participants who did not pass the learning phase (i.e., who did not demonstrate accuracy >70% on the final block of the training and testing within the learning phase; see Supplementary Materials). A broadly comparable same set of conclusions emerged as reported above. Thus the absence of extinction and counterconditioning effects cannot be attributed to a failure of participants to ‘learn’ during the acquisition and intervention phases.

**Figure 4**

*Meta-analytic models outlining moderation of the IR and OEC effects by intervention type (extinction [top panels] or counterconditioning [bottom panels]). In each forest plot, squares represent observed Cohen’s d effect sizes, size of square represents weighting in the model, and error bars represent 95% Confidence Intervals (CIs) around the effect size.*



**General Discussion**

Across seven studies we sought to gain a deeper understanding of the conditions under which evaluations established via intersecting regularities or operant evaluative conditioning can either be undone (via extinction) or modified (via counterconditioning). During an acquisition phase, participants learned that a contingency containing a valenced source ‘intersected’ with a contingency containing a neutral target (i.e., that they both contained a common outcome stimulus). An extinction procedure was then administered which eliminated the intersection by removing the common outcome from the valenced (Experiment 1), target (Experiment 2), or both contingencies (Experiment 3). Experiment 4 examined if a different extinction procedure (CS-only presentations) would eliminate evaluations. In Experiments 5-7 we sought to countercondition evaluations, by either replacing the valenced source with a stimulus of the opposite valence (Experiment 5) or by contingency rearrangement (Experiments 6-7). Participants in the acquisition-only group never encountered an extinction or counterconditioning phase and proceeded directly to the evaluative measures.

**Summary of Findings**

***Intersecting Regularities***

A multilevel meta-analysis of Experiments 1-7 shows that that evaluative learning via intersecting regularities gives rise to strong changes in likes and dislikes, replicating prior work in this area (Hughes et al., 2016). A moderator meta-analysis indicated that such evaluations are relatively ‘sticky’ insofar as eliminating the ‘intersection’ between regularities failed to reduce or eliminate evaluations. Taken together, our findings suggest that learning via intersecting regularities may produce lasting changes in liking that persist even when those intersections are no longer present. A second moderator meta-analysis further reinforced the ‘stickiness’ of these evaluations: counterconditioning procedures in general failed to reverse or eliminate IR effects. However, closer inspection of the data reveals that self-reported evaluations (and IAT scores to a far weaker extent) were reversed in Experiment 5, suggest that reversing the valence of the source stimulus may be a more effective way of counterconditioning (Experiment 5) than contingency rearrangement (Experiments 6-7).

***Operant Evaluative Conditioning***

A multilevel meta-analysis of Experiments 1-7 also showed that operant evaluative conditioning gave rise to strong changes in likes and dislikes whereas a moderator meta-analysis indicated that extinction procedures, in general, failed to reduce or eliminate those evaluations. Inspection of the data revealed one notable exception: removing the outcome from both contingencies did decrease self-reported ratings. This was also the case in Experiment 5, the only study designed to countercondition OEC effects. Here too self-reported ratings were reduced when source stimulus valence was reversed from acquisition to counterconditioning. Taken together, the individual studies suggest that extinction and counterconditioning of OEC effects may be possible under specific conditions, but that the evaluative learning effects were not undermined in general.

**Empirical Implications**

***Extinction of Evaluations***

On the one hand, our findings are broadly consistent with past work suggesting that evaluations established via regularities (e.g., EC) can be difficult to extinguish (Hoffman et al., 2010; but see Lipp et al., 2003; 2010). It seems that once a relationship between source and target stimuli has been established, and the valence of the former has transferred to the latter, removing the intersection that initially gave rise to those evaluations may be “too little, too late” (i.e., post-acquisition changes to the intersection does not decrease liking).

On the other hand, the absence of extinction effects was likely due to the specific parameters used in our studies and extinction may occur if other conditions are met. For instance, it may be that participants viewed the contingencies during the acquisition phase as being a-contextual and the altered contingencies they encountered during the extinction phase in a highly contextual manner (i.e., what was initially learned [acquisition] applies across contexts whereas what is later learned [extinction] only applies to one specific context). Likewise, extinction of evaluations could be facilitated by using a single instead of multiple valenced sources (as we used), presenting stimuli simultaneously instead of sequentially, or even asking participants to rate the targets and outcomes multiple times. Future work should better study the boundary conditions of extinction in the context of IR and OEC. ***Counterconditioning of Evaluations***

Our findings on counterconditioning are also consistent with those elsewhere in the attitudes literature. For instance, when it comes to EC, preferences can be reversed or be eliminated following experience (Hu, Gawronski, & Balas, 2017) or instruction-based counterconditioning via stimulus valence reversal (Gast & De Houwer, 2013), and the former is often more effective than the latter (Hu et al., 2017). In the impression formation literature, evaluations can be formed when people are told that certain positive behaviors are characteristic of a fictional person and then later reversed when they are given contradictory information (e.g., Mann & Ferguson, 2015). At the same time, counterconditioning seems to be a more powerful technique for changing evaluations than other procedures such as extinction. This is true not only for likes and dislikes (Gast & De Houwer, 2013), but also fear (Raes, & De Raedt, 2012), disgust (Engelhard, Leer, Lange, & Olatunji, 2014), and eating behaviors (Van Gucht et al., 2013). Thus establishing a novel relation between a stimulus and valence (counterconditioning) may be more effective in changing liking than trying to eliminate the relation between stimulus and valence altogether (extinction). Future work could explore whether (a) experience-based counterconditioning is also more effective that instruction-based counterconditioning in IR and (b) if counterconditioning is more or less effective than other change procedures (e.g., US revaluation).

***Behavior Change***

It also remains to be seen whether changes in self-reported and automatic evaluations via extinction or counterconditioning correlate with changes in other classes of (real-world) behavior. For instance, Hollands, Prestwich, and Marteau (2011) counterconditioned unhealthy snacks (typically considered positive) with negative images. This intervention not only increased the negativity of implicit attitudes toward unhealthy snacks but also the subsequent consumption of healthy foods instead of those snacks. One possibility would be to countercondition evaluations via IR and determine whether these changes impact meaningful outcomes (e.g., purchasing and consumption behavior of brands). Note, that research on IR has mostly focused on establishing or changing evaluations towards *novel* stimuli. However, Mattavelli, Avishai, Perugini, Richetin, and Sheeran (2017) used the Self-Referencing task, an IR-based paradigm in which stimuli are related with the (generally positive) concept of self, to countercondition green vegetables in a population of participants who did not like green vegetables. This intervention led to more positive implicit attitudes towards green vegetables and to an increased intention to consume them in future. That said, it still remains to be seen if IR-based procedures are also effective when it comes to actual behavioral change (e.g., increased green vegetable consumption).

Still other work suggests that the impact of counterconditioning and extinction may be context dependent (e.g., Van Gucht et al., 2013). For instance, counterconditioned effects are relatively stable across time *within* a given context (i.e., little evidence for spontaneous recovery; see Kerkhof et al., 2011), and yet such effects are often confined to those contexts in which they are counterconditioned, returning once more when the organism finds itself in a novel environment (Bouton, 2000). As mentioned earlier, it would be worth examining the long-term effects and context dependency of counterconditioning of IR and OEC effects in order to determine if these changes in evaluation are stable across time and context.

**Practical Implications**

The ultimate goal when changing evaluations is to demonstrate that doing so leads to a corresponding change in behavior. For instance, an advertisement sets out to increase consumer liking of a brand product with the hope that this change in liking will lead people to actually purchase the product itself. Therefore, it seems useful to identify learning pathways that produce changes in liking that persist across time and in the face of extinction. Our data suggest that this is true for evaluative learning via IR and OEC, where changes in liking were still detectable even when the intersection or contingencies was subsequently disrupted. If anything, IR and OEC effects persisted in the face of extinction procedures. Thus if a consumer product acquires a positive valence via IR or OEC, people may continue to like that item even when they later encounter it by itself in the supermarket. That said, such evaluations are open to change: if one’s product has acquired an unwanted negative valence one could alter those evaluations via a counterconditioning procedure in order to bias the item in a more positive direction. Future work could take this idea one step further and compare IR and OEC to other known evaluative learning pathways (e.g., ME, EC, AA) to determine which pathway influences evaluations and behavior to the greatest extent.

**Limitations**

One limitation was the difficulty we observed in creating an extinction procedure which effectively undermined evaluations of the target stimulus (IR effects). It may be that the extinction procedure used in Experiments 1-3 still retained some valenced elements (e.g., the responses emitted in the presence of the source stimuli) which may have hampered our efforts to extinguish target evaluations. Experiment 4 sought to control for this possibility by presenting stimuli without the need to emit responses – but even this task is not without its own issues (e.g., presenting stimuli in a non-contingent way might be perceived as being unrelated to the acquisition phase; see our previous point about contextual versus a-contextual learning). Another possibility would be to simply omit the valenced contingencies entirely and just expose participants to the target contingencies during extinction. Or to replace the valenced source with a neutral source (although this may come close to the counterconditioning procedure used in Experiment 5). In either case, future work could seek to build and refine on our efforts here.

Another limitation was the presence of both a ‘visible intersection’ (e.g., common outcome) and a ‘hidden intersection’ (i.e., common response locations) connecting the contingencies in many studies. This latter type of intersection may have augmented the IR and OEC effects during the acquisition phase and undermined attempts to reduce them during extinction and counterconditioning. That said, when this hidden intersection was absent (Experiment 4) or controlled for (Experiment 7) we still failed to observe extinction or counterconditioning. Nevertheless, we recognise that this factor likely played a role in the findings reported here. Future work should therefore control for and examine this issue more systematically seeking to establish and change IR effects.

**Conclusion**

We examined the robustness of evaluations established via intersecting regularities and operant evaluative conditioning. Although we could generate novel evaluations via both learning pathways, we could not easily extinguish those evaluations using variants of commonly used procedures. This supports the idea that, once formed, IR effects may be difficult to eliminate. Some evidence was found that evaluations could be counterconditioned under certain conditions (by reversing source stimulus valence) but not others (contingency rearrangement). However, this work represent the first time that these recently discovered learning pathways have been examined in this way. We encourage others to further explore promising strategies for altering what people like and dislike.

**Ethics Statement.** The Ethics Committee of the Faculty of Psychology and Educational Sciences at Ghent University granted ethical approval for the study procedures. All participants were assured that no harm would come to them in the process of experiment, and were told that this experiment involved a learning task, a speeded computer task, and self-reported questions. The results of all tests were kept confidential. Participants were informed that they had the right to stop the experiment at any time during the experiment. Written consent was obtained before the experiment began.

**Data accessibility.** Our data, materials, and code can be found at https://osf.io/u6vtz/

**Competing Interests**. The authors have no financial or non-financial competing interests to disclose.

**Authors Contributions**. SH conceptualized the studies, carried out data collection, data analysis, drafted and revised the manuscript; SM conceptualized the studies, carried out data collection, drafted and revised the manuscript; JDH conceptualized the studies and revised the manuscript. IH analyzed the data and revised the manuscript. All authors gave final approval for publication.

**References**

Baeyens, F., Crombez, G., Van den Bergh, O., & Eelen, P. (1988). Once in contact always in contact: Evaluative conditioning is resistant to extinction. *Advances in Behaviour Research and Therapy*, *10*(4), 179-199.

Blechert, J., Michael, T., Williams, S. L., Purkis, H. M., & Wilhelm, F. H. (2008). When two paradigms meet: Does evaluative learning extinguish in differential fear conditioning?. *Learning and Motivation*, *39*(1), 58-70.

Bouton, M. E. (2000). A learning theory perspective on lapse, relapse, and the maintenance of behavior change. *Health Psychology*, *19*(1S), 57-63.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory*, *11*(5), 485-494.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology, 10,* 230–241.

De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin & Review, 20(4)*, 631-642.

De Houwer, J., & Hughes, S. (2020). The psychology of learning: An introduction from a functional-cognitive perspective. *The MIT Press*.

Ebert, I. D., Steffens, M. C., von Stülpnagel, R., & Jelenec, P. (2009). How to like yourself better, or chocolate less: Changing implicit attitudes with one IAT task. *Journal of Experimental Social Psychology, 45,* 1098–1104.

Eder, A. B., Krishna, A., & Van Dessel, P. (2019). Operant evaluative conditioning. *Journal of Experimental Psychology: Animal Learning and Cognition, 45(1)*, 102-110.

Engelhard, I. M., Leer, A., Lange, E., & Olatunji, B. O. (2014). Shaking that icky feeling: effects of extinction and counterconditioning on disgust-related evaluative learning. *Behavior Therapy, 45(5),* 708-719.

Gast, A., & De Houwer, J. (2013). The influence of extinction and counterconditioning instructions on evaluative conditioning effects. *Learning and Motivation*, *44*(4), 312-325.

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,* 197-216.

Hasselblad, V., & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin, 117(1)*, 167-178.

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

Hollands, G. J., Prestwich, A., & Marteau, T. M. (2011). Using aversive images to enhance healthy food choices and implicit attitudes: An experimental test of evaluative conditioning. *Health Psychology*, *30*(2), 195-203.

Hu, X., Gawronski, B., & Balas, R. (2017). Propositional Versus Dual-Process Accounts of Evaluative Conditioning: II. The Effectiveness of Counter-Conditioning and Counter -Instructions in Changing Implicit and Explicit Evaluations. *Social Psychological and Personality Science*, DOI: https://doi.org/10.1177/1948550617691094

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

Kerkhof, I., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2011). Counterconditioning. An Effective Technique for Changing Conditioned Preferences. Experimental Psychology, 58, 31-38.

Lipp, O. V., Mallan, K. M., Libera, M., & Tan, M. (2010). The effects of verbal instruction on affective and expectancy learning. *Behaviour Research and Therapy*, *48*(3), 203-209.

Lipp, O. V., Oughton, N., & LeLievre, J. (2003). Evaluative learning in human Pavlovian conditioning: Extinct, but still there?. *Learning and Motivation*, *34*(3), 219-239.

Mann, T. C., & Ferguson, M. J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *Journal of Personality and Social Psychology, 108(6),* 823-849.

Mattavelli, S., Avishai, A., Perugini, M., Richetin, J., & Sheeran, P. (2017). How Can Implicit and Explicit Attitudes Both Be Changed? Testing Two Interventions to Promote Consumption of Green Vegetables. *Annals of Behavioral Medicine*, 1-8.

Mattavelli, S., Richetin, J., Gallucci, M., & Perugini, M. (2017). The Self-Referencing task: Theoretical overview and empirical evidence. *Journal of Experimental Social Psychology*, *71*, 68-82.

Moreland, R. L., & Topolinski, S. (2010). The Mere Exposure Phenomenon: A LingeringMelody by Robert Zajonc. *Emotion Review, 2(4)*, 329-339.

Raes, A. K., & De Raedt, R. (2012). The effect of counterconditioning on evaluative responses and harm expectancy in a fear conditioning paradigm. *Behavior Therapy, 43(4),* 757-767.

Sánchez-Meca, J., Marín-Martínez, F., & Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods, 8(4)*, 448-467.

Van Dessel, P., Eder, A. B., & Hughes, S. (2018). Mechanisms underlying effects of approach-avoidance training on stimulus evaluation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44(8)*, 1224-1241.

Van Gucht, D., Baeyens, F., Hermans, D., & Beckers, T. (2013). The inertia of conditioned craving. Does context modulate the effect of counterconditioning?. *Appetite, 65,* 51 -57.

Vansteenwegen, D., Francken, G., Vervliet, B., De Clercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology-Animal Behavior Processes*, 32, 71-79.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36(3)*, 1-48.

1. The concept of a ‘regularity’ is simply a term denoting any state “in the environment…that entails more than the presence of a single stimulus or behavior at a single moment in time.” (De Houwer, Barnes-Holmes, & Moors, 2013, p. 634; for more on this topic see De Houwer & Hughes, 2020). [↑](#footnote-ref-1)
2. Unlike EC, where there are only two (conditioned [CS] and unconditioned [US]) stimuli involved, the intersecting regularities procedures reported here involve three key stimuli: the valenced *source* stimulus, neutral *outcome* stimulus, and the neutral *target* stimulus. We will adopt the latter terminology from this point onwards. [↑](#footnote-ref-2)
3. Note: learning via intersecting regularities does not refer to a single procedure but rather a class of procedures that each set out to (a) create multiple regularities that (b) intersect with one another in terms of a common element (e.g., stimulus or response). The work reported in this paper focuses on just one such procedure. [↑](#footnote-ref-3)
4. Extinction procedures in the context of classical and operant conditioning not only remove the regularity that originally gave rise to the change in behavior but also (typically) involve the removal of the valenced stimulus as well. Although many of the extinction procedures used in this paper also did so (Experiments 1, 3, 4, 7), others only removed the regularity and contingued to present the valenced stimulus (Experiment 2). [↑](#footnote-ref-4)
5. The procedures described in Experiments 1-7 are - strictly speaking - not extinction or counterconditioning tasks given that extinction and counterconditioning typically refer to procedures used in the classical and/or operant conditioning literatures and not to situations involving intersections between regularities. Rather than open a conceptual debate surround the meaning of these two terms, we were simply interested in testing the *robustness* of IR effects in the face of manipulations that attempt to undo (which is often the goal of extinction tasks) or modify (which is often the goal in counterconditioning tasks) the intersections that gave rise to the original IR effects. We will continue to refer to extinction- and counterconditioning-like tasks for communication sake. [↑](#footnote-ref-5)
6. Different types of OEC can be distinguished depending on what is the valenced event and what is the initially neutral event that acquires a new valence within a single operant contingency. In the present set of experiments, the valenced event is a stimulus that signals the nature of the correct response (i.e., the source) and the neutral event is the outcome of the response. In other types of OEC such as Approach-Avoidance learning, the valenced event is the response (i.e., approaching or avoiding) whereas the neutral event is the stimulus that signals the correct response. In still other types of OEC, the outcome is the valenced event and the response or the stimulus signaling the response are the initially neutral event. [↑](#footnote-ref-6)
7. Note that the study designs and data-analysis plans for all experiments are available on the Open Science Framework website (https://osf.io/u6vtz/). We report all manipulations and measures used in our experiments. All data were collected without intermittent data analysis. The data analytic plan, experimental scripts, and data are available at the above link. Deviations from pre-registration can also be found at the above link. [↑](#footnote-ref-7)
8. This procedure should extinguish IR effects but leave OEC effects intact. Indeed, if anything, the procedure may further strengthen OEC effects given that it provides double the exposure to the contingencies underpinning OEC effects relative to the acquisition-only group. [↑](#footnote-ref-8)
9. Our main interest was in evaluative learning via IR. Thus extinction of IR effects was tested on ratings, IAT scores, and behavioral intentions whereas extinction of OEC effects was only tested on ratings and intentions. [↑](#footnote-ref-9)
10. The counterconditioning procedure in Experiment 6 should impact outcome and target stimuli in different ways. It could potentially reverse evaluations of target stimuli while leaving intact (or strengthening) previously acquired outcome evaluations (i.e., countercondition IR effects while boosting OEC effects given that it involves additional exposure to the same operant evaluative conditioning contingencies as in the acquisition phase). [↑](#footnote-ref-10)
11. As we previously mentioned, the extinction and counterconditioning procedures were designed to modify IR effects. In certain cases (Experiments 2, 6, 7) these procedures boosted rather than undermined OEC effects. As such, the meta-analytic effect for the OEC effects should be treated with caution, and the forest plot is only provided as a visual overview of effects across studies. [↑](#footnote-ref-11)