**The Impact of Extinction vs. Counter-conditioning on Evaluative Learning via Intersecting Regularities.**

Intersecting Regularities is a new route for changing liking. In a typical IR procedure, individuals perform a task in which they learn that valenced and neutral stimuli are related to each other via one (or more) elements in operant contingencies that intersecting with one another. For instance, a first operant contingency might consist in pressing a red button (R1) in the presence of a positively valenced source stimulus (S1) that leads to the presentation of a neutral outcome (O1). Then in a second contingency, pressing a yellow button (R2) when a neutral target stimulus is present (T1) leads to the exact same outcome (O1). Participants may evaluate the neutral target (T1) more positively than they used to do due to the fact that the two operant contingencies intersect each other in terms of a common outcome (i.e., positive source (S1)🡪 red button (R1) 🡪 neutral stimulus (O1); Neutral target (T1) 🡪 yellow button (R2) 🡪 neutral outcome (O1)). The effectiveness of evaluative learning via IR has been demonstrated on both implicit and explicit attitudes (Hughes, De Houwer & Perugini, 2016). So far IR studies have only focused on attitude formation (i.e., establishing evaluations for novel stimuli). However, an important aspect of (evaluative) learning is how to change evaluations once they’ve been acquired. In Evaluative Conditioning (EC), for instance, one way of altering evaluative responses is via *counter-conditioning*. The aim of the present contribution is to investigate whether counter-conditioning can also be used to change recently acquired evaluative responses via intersecting regularities.

**IR and Extinction**

In EC, extinction refers to an experimental procedure containing two sequential phases. In the first phase (acquisition), the individual is exposed to a contingency between two stimuli - a conditioned stimulus (CS) and an unconditioned stimulus (US). The second phase (extinction) consists of the mere presentation of the CS, without contingent presentation of the US. Previous work indicates that, compared to Pavlovian conditioning, EC is less susceptible to extinction (e.g., Baeyens, Crombez, Van den Bergh, & Eelen, 1988; Díaz, Ruiz, & Baeyens, 2005; Dwyer, Jarrat, & Dick, 2007; Vansteenwegen, Francken, Vervliet, De Clercq, & Eelen, 2006; Gawronski, Gast, & De Houwer, 2015). We explore if changes in liking also decrease in magnitude once the intersection between regularities is put into extinction. In our previous studies we attempted to achieve this outcome with varying levels of success (see <https://osf.io/nkr9g/>).

However, one might argue that the observed IR effect results from regularities that intersect both at the level of the outcomes and at the level of the location of the response required to categorize stimuli. In fact, in the standard IR procedure, participants categorize a positive source and a first target stimuli with key letter located on one side of the keyboard (e.g. D and C), while the negative source and the second target require categorization via key letters located in the opposite side of the keyboard (e.g. J and N). Thus, evaluations can transfer from sources to target object both via the common outcome and via the common key location. In our previous studies we merely focused on extinguishing intersections in term of outcomes, while key location overlap between sources and targets is maintained in the extinction phase. In Study 7 we aim to replicate the procedure adopted in Study 3 by also removing overlap between stimuli in terms of key-response location. Therefore, we slightly modify both the IR and the extinction procedure in such a way that no key pressing is required to categorize stimuli. In each trial either a valenced source or a neutral target appear on screen. Participants are to learn to categorize the stimuli with the correct letters. Crucially, rather than pressing on them on the keyboard, the four letters used in previous studies were displayed horizontally below the relevant stimulus in each trial. By clicking with the mouse on one of the four letters, participants learned what letter was assigned to each stimulus and therefore predicted the appearance of a certain outcome. The location of the letters on the screen varied across the four blocks of trials, such that none of them could be related to another one because of either location on screen (i.e., right vs. left) or closeness. In the extinction condition, this training phase is followed by an equivalent task in which the same food images and Chinese characters are to be categorized with the same letters on screen. However, no outcome will be presented upon correct responses to any of the stimuli (in line with Study 3).

**Extinction through the removal of stimuli**

**Study 7**

During Phase 1 participants first learn (Positive Source (S1) 🡪 R1 🡪 Neutral Outcome (**O1)**; Neutral Target (T1) 🡪 R2 🡪 Neutral Outcome (**O1**). Similarly, they learn that (Negative Source (S2) 🡪 R3 🡪 Neutral Outcome (**O2)**; Neutral Target (T2) 🡪 R4 🡪 Neutral Outcome (**O2**). We would expect O1 and T1 to be positively valenced and O2 and T2 to be negatively valenced after this phase. Rather than pressing on computer keys, the categorization occurred via mouse clicking on letters appearing on screen on an horizontal line. The distance between each letter was constant, while the location of the letters on the screen varied randomly across trials.

During phase 2, participants performed the same action in response to the same stimuli. However, the presentation of any outcome after stimuli responses was eliminated (i.e., S1🡪R1🡪***Nothing***; T1🡪R2🡪***Nothing*** and S2🡪R3🡪***Nothing***; T2🡪R4🡪***Nothing***).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ACQUISITION** | | | **EXINCTION** | | |
| **STIMULUS** | **RESPONSE** | **OUTCOME** | **STIMULUS** | **RESPONSE** | **OUTCOME** |
| Positive source (S1) | Click D (R1) | **Neutral outcome (O1)** | Positive source (S1) | Click D (R1) | **/** |
| Neutral target (T1) | Click C (R2) | **Neutral outcome (O1)** | Neutral target (T1) | Click C (R2) | **/** |
| Negative source (S2) | Click J (R3) | **Neutral outcome (O2)** | Negative source (S2) | Click J (R3) | **/** |
| Neutral target (T2) | Click N (R4) | **Neutral outcome (O2)** | Neutral target (T2) | Click N (R4) | **/** |

**IR and Counter-conditioning**

In EC, counter-conditioning refers to an experimental procedure containing two sequential phases. In the first phase (acquisition), the individual is exposed to a contingency between two stimuli - a conditioned stimulus (CS) and an unconditioned stimulus (US). The second phase (extinction) consists of the presentation of the CS with an US of opposite valence. Counter-conditioning seems effective in changing the valence of a stimulus (e.g., Kerkhof, Vansteenwegen, Baeyens, & Hermans, 2011; Baeyens, Eelen, Van den Bergh, & Crombez, 1989). In Studies 5-6 we tested the impact of different ways of operationalizing counterconditioning on evaluative responses that had been formed via IR. Unlike EC, counter-conditioning in IR is not based on stimuli pairings, rather, on novel regularities that intersect and give rise to new relations between source and target stimuli. We explored different ways of manipulating intersecting regularities between stimuli in a counter-conditioning phase and tested their impact on evaluative responding.

In a first experiment (Study 5) we tested a counterconditioning procedure in which the original valenced stimuli presented in the acquisition phase are replaced by stimuli of opposite valence. Specifically in the first acquisition phase participant learn Positive source (S1) 🡪 R1 🡪 **O1;** Neutral Target (T1) 🡪 R2🡪 **O1**; and Negative Source (S2) 🡪 R3🡪 **O2**; Neutral Target (T2) 🡪 R4🡪 **O2**). In the counter-conditioning phase, the positive source (S1) was replaced by a new negative source (S2) while the negative source (S2) was replaced by a new positive source (S1). Thus, in the counterconditioning phase participants learn that the same outcome (O1) appears when the correct key is pressed to respond to both T1 and S2 and another outcome (O2) appears upon correct response to both T2 and S1. We showed that this procedure influences the evaluation of the target stimuli.

An alternative way to induce counter-conditioning via IR was tested in Study 6 by using the same stimuli presented in the acquisition phase, but switching the outcomes (S1🡪R1🡪O1; T1🡪R2🡪O2; S2🡪R3🡪O2; T2🡪R4🡪O1) (i.e., by ‘re-wiring the connections’ rather than replacing the valenced stimuli). In this way a positive stimulus (S1) that in the acquisition phase used to intersect with a neutral one (T1) because of the same outcome O1, in the counterconditioning phase would intersect with a second target (T2). This way to implement a counterconditioning procedure based on re-writing the connections between sources and target did not impact participants’ evaluations. In Study 7 we aim at re-testing the counterconditioning procedure designed in Study 6 by removing any form of intersections learned by participants in the acquisition phase.

In fact, learning via IR established in the acquisition phase can be seen as the result of regularities that intersect at the level of (i) the outcomes (i.e., O1 and O2) and (ii) the location of the response required to categorize stimuli. The latter intersection reflects the fact that participants categorize a positive source and a first class of target stimuli with key letters located on one common side of the keyboard (e.g. D and C), while the negative source and the second target require categorization via key letters both located in the opposite side of the keyboard (e.g. J and N). Thus, evaluations can transfer from sources to target object both via the common outcome and via the common key location. Crucially, Study 6 the counterconditioning phase differed from the acquisition just in terms of the valence of the stimuli related to both brands and Chinese symbols. In this procedure, pressing two keys located on the same side should have reinforced the newly learned link between valenced source and target stimuli. On the other hand, in Study 6, both valenced stimuli and targets remained related with the same response. What switched was the type of outcome presented upon correct response to one or the other target object. Therefore, Study 6 suffered from a potential conflict between a previously established and reinforced intersection in terms of key location and the newly acquired intersection based on outcome stimuli.

To address this issue, we slightly modify both the IR and the counterconditioning procedure in such a way that no key pressing is required to categorize stimuli. In each trial either a valenced source or a neutral target appear on screen. Participants learned to categorize the stimuli with the correct letters (responses). Crucially, rather than pressing on them on the keyboard, the four letters used in previous studies were displayed horizontally below the relevant stimulus in each trial. By clicking with the mouse on one of the four letters, participants learned what letter was assigned to each stimulus and therefore predicted the appearance of a certain outcome. The location of the letters on the screen varied across the four blocks of trials, such that none of them could be related to another one because of either location on screen (i.e., right vs. left) or closeness. In the counterconditioning condition, this training phase was followed by an equivalent task in which the same food images and Chinese characters were to be categorized with the same letters presented on screen. However, like we did in Study 6, the type of outcome (O1 and O2) presented to participants upon correct responses to the target stimuli (T1 and T2 ) were reversed compare to the acquisition phase.

**Counterconditioning via re-writing stimuli connections with no intersection in key location**

**Study 7**

During Phase 1 participants first learn (Positive source (S1) 🡪 R1 🡪 Neutral Outcome (**O1)**; Neutral Target (T1) 🡪 R2 🡪 Neutral Outcome (**O1**). Similarly, they learn that (Negative Source (S2) 🡪 R3 🡪 Neutral Outcome (**O2)**; Neutral Target (T2) 🡪 R4 🡪 Neutral Outcome (**O2**). We would expect O1 and T1 to be positively valenced and O2 and T2 to be negatively valenced after this phase. Rather than pressing on key letters, the categorization occurred via mouse clicking on letters appearing on screen on a horizontal line. The distance between each letter was constant, while the location of the letters on the screen varied randomly across trials.

During phase 2 (i.e., counterconditioning), participants performed the same actions in response to the same stimuli. However, while the outcomes presented upon correct responses to source stimuli (S1 and S2) remained constant, those following the target stimuli (T1 and T2) were reversed (i.e., S1🡪R1🡪**O1**; T1🡪R2🡪***O2*** and S2🡪R3🡪**O2**; T2🡪R4🡪***O1***).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ACQUISITION** | | | **COUNTERCONDITIONING** | | |
| **STIMULUS** | **RESPONSE** | **OUTCOME** | **STIMULUS** | **RESPONSE** | **OUTCOME** |
| Positive source (S1) | Click D (R1) | **Neutral outcome (O1)** | Positive source (S1) | Click D (R1) | Neutral outcome (O1) |
| Neutral target (T1) | Click C (R2) | **Neutral outcome (O1)** | Neutral target (T1) | Click C (R2) | **Neutral outcome (O2)** |
| Negative source (S2) | Click J (R3) | **Neutral outcome (O2)** | Negative source (S2) | Click J (R3) | Neutral outcome (O2) |
| Neutral target (T2) | Click N (R4) | **Neutral outcome (O2)** | Neutral target (T2) | Click N (R4) | **Neutral outcome (O1)** |