Liking as far as you like yourself: Exploring the Self-Referencing effect across multiple intersecting regularities and its relationship with self-esteem.

Simone Mattavelli, Juliette Richetin, & Marco Perugini

University of Milano-Bicocca

Corresponding author: Simone Mattavelli, Department of Psychology, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo, 1, 20126 Milan, ITALY

mail to: simone.mattavelli@unimib.it

Abstract

Classic instances of evaluative learning require the spatio-temporal contingency between source and target stimuli. However, people can learn to like stimuli in a more indirect fashion. Moreover, many of our preferences are self-referential: we tend to like the objects that are related to ourselves. For instance, it is demonstrated that performing the Self-Referencing task, a categorization task based on intersecting regularities between the self and a target stimulus, results in a transfer of positivity from the self to the target. Unexplored so far is the extent to which intersecting regularities can be exploited to form even less direct connections between the self and the target, which ultimately results in a preference for the latter. In two experiments, we tested the indirect SR effect. To this aim, we set up a two-step learning procedure in which, after completion of the SR task, participants underwent an additional task, where target stimuli categorized in the SR shared a behavioral response with novel targets. Experiment 1 (N=241) showed significant direct and indirect SR effects, both implicitly and explicitly, irrespectively of the type of targeted stimuli (i.e., groups or brands). Experiment 2 (N=174) replicated such findings and showed that explicit, but not implicit, self-esteem moderated the SR effect. Although at a descriptive level the moderating role of explicit self-esteem showed stronger on the indirect SR effect, no statistical evidence was found to support a differential impact on either direct or indirect SR. We discuss the power of IR to transfer self-positivity indirectly across stimuli, the role of self-esteem, and the theoretical implications of our findings.

Keywords: indirect attitude change, intersecting regularities, Self-Referencing, self-esteem

Liking as far as you like yourself: Exploring the Self-Referencing effect across multiple intersecting regularities and its relationship with self-esteem.

1. Introduction

Likes and dislikes are key determinants for many of our behaviors (Allport, 1935). Because most of our preferences are learned, rather than innate (Rozin & Millman, 1987), learning about the pathways through which our preferences form and change is a central aim in psychology. For years, researchers interested in evaluative learning have focused on the impact of stimulus pairings on liking. One prime example of an evaluative learning effect that results from pairing is evaluative conditioning (EC). EC can be described as the change in the valence of a stimulus resulting from the pairing with another positive or negative stimulus (De Houwer, 2007). In its standard operationalization, an EC procedure consists of the repeated presentation of a neutral object (conditioned stimulus, CS) in spatio-temporal contingency with a valenced object (unconditioned stimulus, US, see Hoffman, De Houwer, Perugini, Baeyens, & Crombez, 2010). Therefore, EC is a robust example of the impact of stimulus-pairings on liking. The direct pairing between stimuli represents a powerful environmental regularity that drives our preferences. For instance, a commercial ad might attempt to influence our liking towards a new apparel brand by showing its products along with a famous pop-star.

In many situations, however, likes and dislikes form without direct pairings between valenced and neutral stimuli. Imagine, for instance, that one person, Bob, is watching a football game between two unknown teams, Team X vs. Team Y. Players of Team X wear a black and red jersey, those of Team Y are in blue. Black and red happen to be also the colors of Bob's favorite basketball team. Without any plausible reason to support either side, Bob will probably end up liking Team X over Team Y. Although no direct contact (pairing) between Bob's favorite team and Team X has ever occurred, the mere fact that they share one common element should result in a perceived similarity that enables a transfer of liking. Likewise, research on EC has shown that

conditioned evaluations can transfer to objects that are either pre- or post-associated with the CS involved in a direct pairing with a US (Hammerl & Grabitz, 1996; Walther, 2002).

Based on the idea that our preferences can form in an indirect fashion is learning via Intersecting Regularities (IR, Hughes, De Houwer, & Perugini, 2016). Learning via IR occurs when an initially neutral stimulus is preferred over another neutral stimulus because some functional features of the former intersect with those of a valenced stimulus. For instance, in Hughes et al. (2016, Experiment 2), participants were exposed to a learning task that sought to establish two sets of operant contingencies that intersected in terms of a common outcome. During half of the trials, they learned that pressing one button (R1) in the presence of a positive source stimulus (S1), or a second button (R2) in the presence of a neutral target stimulus (S2), caused an outcome stimulus (O1) to appear. Participants also learned that pressing a third button (R3) in the presence of a negative source (S3), or a fourth button (R4) in the presence of a neutral target (S4) caused the onset of another outcome (O2). In this way, an operant contingency involving a valenced stimulus *intersected* with an operant contingency involving a neutral stimulus. Following such training procedure, participants showed both implicit and explicit preference for ideograph O1 over O2 and brand product S2 over S4, although neither of them was directly related to any of the valenced sources (i.e., S1 and S3).

Intersections between environmental regularities can shape many of our likes and dislikes. Therefore, it is important to understand what features can modulate learning via IR. Moreover, regularities in the environment can intersect at different levels of complexity. For instance, two stimuli can intersect because they both possess one common feature (e.g., two persons who support the same football team) or because they possess two distinct features that happen to be simultaneously possessed by a third common stimulus (e.g., two persons supporting two different teams from two different countries, but both sharing the same home colors). Here we focus on this specific aspect of learning via IR, testing how increasing the level of complexity between regularities affects learning.

Besides originating in an indirect fashion, many of our preferences are self-referential. Previous research on both ownership and EC indeed showed that objects that relate to the self are better liked (Gawronski, Bodenhausen, & Becker, 2007; Gawronski & LeBel, 2008, study 3; Walther & Trasselli, 2003). The Self Referencing task (SR task, Prestwich, Perugini, Hurling, & Richetin, 2010; Perkins & Forehand, 2012) is an associative learning task in which the positivity is carried by the self transfers to neutral objects via IR.[1] Specifically, the SR task requires participants to categorize self-related words and stimuli belonging to a first target category with a common action, while an alternative action is required to categorize other-related words and a second target category. The standard SR effect relies on a link between the self and one target stimulus that has to be inferred from two pairs of operant contingencies that intersect at the level of the performed action. From one first pair of operant contingencies, "*if I see a self-related stimulus, then I press X*" and "*if I see Target1, then I press X*", people learn that target 1 stimuli relate to self. From the second pair of operant contingencies, "*if I see an others-related stimulus, then I press Y*" and "*if I see Target2, then I press Y*", people learn instead that target 2 stimuli relate to others. Therefore, the relations are established via alternative intersecting regularities between two pairs of valenced and neutral stimuli, that is, the common action that is required to categorize them. A meta-analysis (Mattavelli, Richetin, Gallucci & Perugini, 2017) showed that the paradigm leads to genuine changes in evaluation assessed via implicit (e.g., reaction-time tasks) and explicit measures (e.g., self-report) for various target stimuli such as fictitious social groups (Perugini, Richetin, & Zogmaister, 2014) or brands (Perugini, Zogmaister, Richetin, Prestwich, & Hurling, 2013; Richetin, Mattavelli, & Perugini, 2016). However, no studies have investigated whether self-driven preferences can transfer to stimuli that intersect with the self in an indirect way. To exemplify, imagine an additional learning task that establishes two new pairs of operant contingencies, "*if I see*

[1] The term "associative learning" refers to a change in the behavior of the individual that results from relations between stimuli in the environment that are acquired throughout the SR task (see De Houwer, 2009). Stating that the effect is associative is merely to say that it origins from newly acquired relations, whilst it does not refer to the nature of the mental process that mediates the effect.

Target1, then I press Z" and "*if I see Target3, then I press Z*" versus "*if I see Target2, then I press W*" and "*if I see Target4, then I press W*". From such training, people might learn more than the new relations between target stimuli (i.e., Target1 and Target3 vs. Target2 and Target4). Crucially, they might also learn to relate (i) the self with Target3 and (ii) others with Target4, as both pairs of stimuli intersect through a stimulus to which they both relate (i.e., Target1 and Target2, respectively).

As we said earlier, a key element that characterizes the SR task is its reliance on the self as a source of valence. The idea that the self represents a positive source is supported by ample empirical evidence showing that the majority of individuals hold a positive view of themselves (Yamaguchi et al., 2007). However, such positivity varies across individuals. Importantly, variations at the level of one's self-esteem result in variations in the magnitude of the SR effect. Corroborating this idea, Prestwich and colleagues (2010) showed that the higher the self-esteem, the larger the SR effect on implicit (but not explicit) evaluative change. Also, Mattavelli, Richetin, and Perugini (2019) observed, in a modified form of the SR task in which the self was opposed to an equally valenced attitudinal source, that the more positive the endorsed explicit self-view, the stronger the SR effect was. Thus, both implicit and explicit self-evaluations seem important qualifiers of the SR effect. In this work, we put forward the idea that the moderating role of self-esteem on the SR effect might increase as the relation between the self and a target becomes less direct. To illustrate our reasoning, the reader could think of the attitudinal objects involved in evaluative change (i.e., sources and targets) as vessels fountain. In this fountain, water originates from the first vessel, located at the top, and can fall on the vessel immediately below as soon as it exceeds a threshold level of 5000 ml. Similarly, the same threshold level of 5000 ml must be exceeded for the water to pass from the second to the third vessel, located below. Whether the water will get to the third vessel depends on the amount of water that originates from the first vessel. For instance, 8000 ml of water generated from the first vessel will be enough for the water to reach the second vessel, but not to get to the third one. Going back to the direct and the indirect SR effects,

water represents the valence carried by the self (i.e., the first vessel), which can transfer directly to a proximal target (i.e., the second vessel) and indirectly to a more distal one (the third vessel). Therefore, one might hypothesize that for an indirect SR effect to occur, a higher level of positivity carried by the self (i.e., self-esteem) is needed, compared to that required for the standard SR effect. In essence, the moderating role played by self-esteem should increase as the number of intersecting elements between the self and the final target also increases.

In two experiments, we tested both the direct and the indirect SR effect by relying on two procedurally equivalent learning tasks. After completion of the SR task, in which self- versus other-related stimuli were categorized through the same action as Target1 versus Target2, participants underwent an indirect task in which Target1 versus Target2 were categorized through the same action as Target3 versus Target4. Implicit and explicit evaluation measures were administered for both Target1 versus Target2 (i.e., direct SR effect) and Target3 versus Target4 (indirect SR effect). Moreover, capitalizing on the crucial role of the self as a source of valence to be transferred across target stimuli, in Experiment 2, we focused on whether the hypothesized indirect SR effect could be qualified by one's level of implicit and explicit self-esteem. In essence, assuming that self-esteem reflects the original amount of positivity carried by the self, this amount of positivity could determine the occurrence of the indirect SR effect. In both experiments, direct and indirect SR effects were assessed via explicit ratings and the Implicit Association Test (IAT, Greenwald, McGhee, & Schwartz, 1998). We added the IAT as it is assumed to reflect more automatic instances of evaluation that can influence behavior in unique ways (De Houwer, Teyge-Mocigemba, Spruyt, & Moors, 2009; but see Corneille & Hütter, 2020).

For both experiments, all measures, manipulations, and exclusions in the study are disclosed. The sample sizes were determined to achieve sufficient sensitivity to detect the key hypothesized effects, given logistical constraints. Data collections were terminated before analyses were run. For more details, see sensitivity analyses in the Analysis plan and data preparation paragraphs. Data and

analyses codes for both studies are available at the following link https://osf.io/9s4mk/?view_only=6deedf652b844d69b0673dd4b9bf6e24.

2. Experiment 1

Experiment 1 investigated the indirect SR effect established through multiple intersections between the self and the target stimuli. Moreover, since prior research has shown that the SR effect is robust on both fictitious brands and social groups (see Mattavelli et al., 2017), we also investigated whether the indirect SR effect could be conditional on the type of target stimuli.

2.1 Method

Both the studies received approval by the Ethics Committee of the University of Milan-Bicocca.

2.1.1 Stimuli

Four fictitious social groups and their members (i.e., Lerriani [Target1], Dattiani [Target2], Craviani [Target3], and Zimmiani [Target4]) were used in both Experiment 1 and 2 and taken from previous SR studies (e.g., Perugini et al., 2012; Richetin et al., 2016). Experiment 1 also included four fictitious brand logos with an embedded brand name (i.e., Rama [Target1], Moniso [Target2], Lestea [Target3], and Sabea [Target4]) that we pre-selected out of a list of twelve candidates. From two pretests ($N = 32$ and $N = 33$, respectively), we selected the most neutral pairs in terms of valence, at both implicit (IAT) and explicit (semantic differential on name, logo, products, and brand in general) level. There was no difference (all ps > .184) at the explicit level, whereas, at the implicit level, we observed a significant preference for brand1 over brand2 ($p = .001$)[2] and none between brand3 and brand4 ($p = .155$).

2.1.2 Participants and procedure

[2] Given the between-subjects design of a SR study (one brand is assigned to the self for one half of the participants and to the category others for the other half), this significant difference was not crucial for the main purpose of the study.

Two hundred and forty-one students (168 women, *M*age = 22.91, *SD* = 3.83) took part in the experiment in exchange for course credit. We adopted a 2 (SR task: Target1 - self vs. Target2 - self) x 2 (indirect task: Target3 – Target1 vs. Target4 – Target1) x 2 (Target Type: Groups vs. Brands) between-subjects design. The experimenter introduced participants to one of the four fictitious groups or brands and informed them they would be later asked questions about them. Participants completed an instruction check, followed by the SR task. In the SR task, either Target1 or Target2 exemplars were categorized with the same action as words related to the self, depending on the condition assigned. After that, participants answered an Intersecting Regularities memory question, completed an IAT, and an explicit measure of evaluations towards Target1 and Target2. Participants then completed a second learning task, the indirect task. For half of them, exemplars of the target stimuli priorly related to the self were categorized with the same key as exemplars of Target3 (and exemplars of the target prior related to Others with the same key as Target4). For the other half, the relationships were reversed. Participants then completed a second Intersecting Regularities memory question regarding the second learning task, followed by another IAT and an explicit evaluation to assess the implicit and explicit evaluations towards Target3 and Target4. Participants were thanked and debriefed.

2.1.3 Materials

Instructions check. A task was designed to test whether participants could sort either group or brand stimuli to the correct category label. In a twelve-trials categorization task, three stimuli belonging to the four targets (groups vs. brands, depending on the target condition to which participants were assigned) were presented individually in the middle of the screen. Participants assigned each stimulus to the corresponding category by clicking with the mouse on the category label. The four category labels remained visible aligned on the top part of the screen. The inter-trial interval was 400ms. An 80% correct response rate was considered as a threshold to move on to the SR task. Otherwise, participants read the description of the four categories again.

SR task. Participants completed two blocks of 40 trials that required them to categorize Target1 versus Target2 stimuli and words related to others (others, them, they, she, he) to one response key (e.g., "Blue") on the keyboard and Target1 versus Target2 stimuli and words related to self (self, me, my, mine, I) to a different response key (e.g., "Yellow"). Next, they repeated the two blocks of 40 trials after switching the keys assigned to each category (i.e., Target1 vs. Target2 stimuli and others assigned to the "Yellow" key, and Target2 vs. Target1 stimuli and self-related words to the "Blue" key). In the case of incorrect classification, a red-X appeared on the screen and remained until correction. The inter-trial interval was 400ms. The order in which participants completed these two blocks was counterbalanced. For each target category, we used five stimuli. The stimuli for all tasks and measures used in both experiments are reported in the Supplementary Materials, Appendix B.

Intersecting Regularities memory question 1. Participants answered the following question: "*The task you've just performed required you to press the same key to categorize stimuli belonging to the self (Me, mine, my) and stimuli belonging to a group. Do you remember which group?*" with 3 options (Target1 vs. Target2 vs. do not remember). Responses were classified as "correct IR memory", "incorrect IR memory", or "no IR memory".

Indirect task. The task mirrored the general structure of the SR task in the number of blocks (and trials), the order of presentation of the stimuli, the number of exemplars, and the response keys (i.e., a blue or yellow button, see figure 1 for a schematic overview of the operant contingencies involved in both the SR and the Indirect task). In one condition, participants pressed the same key to categorize stimuli belonging to exemplars of the group previously related to the self (i.e., Target1 vs. Target2 depending on the SR condition) or Target3, and another key for exemplars belonging to the group previously related to others and Target4. In the other condition, Target4 stimuli were categorized through the same action as the stimuli belonging to the category previously related to the self, while Target3 stimuli were classified with the object priorly related to others.

Intersecting Regularities memory question 2. Depending on the SR condition they were assigned to, participants answered the following question: "*The task you've just performed required you to categorize stimuli belonging to [name of the target categorized through the same action as the self in the SR task: Target1 vs. Target2] and stimuli belonging to another group. Do you remember which group?*" with three options (Target3 vs. Target4 vs. do not remember). Based on their response, participants were classified as "correct IR memory", "incorrect IR memory", or "no IR memory".

Dependent variables.

IAT Target1/Target2. Participants categorized stimuli presented individually and in a random order in the middle of the screen using two keys (i.e., 'Yellow' and 'Blue'). The target concept was Target1, and its contrast was Target2, whereas the attribute categories were Positive and Negative. The order of the two critical blocks was counterbalanced between participants, with half of the participants having the combination Target1 and Positive being presented first and the other half having the combination Target2 and Positive being presented first. All practice blocks consisted of 20 trials, and each critical block consisted of 81 trials (80 trials, including one initial dummy trial). A red X appeared in the middle of the screen for 200ms if the participant did not answer correctly. There was no built-in penalty, and the inter-trial interval was 500ms. For each attribute and target category, we used five stimuli (see Supplementary Materials, Appendix B). We calculated the D6 score following the procedure suggested by Greenwald, Nosek, and Banaji (2003), such as higher scores indicated a preference for Target 1 over Target 2

IAT Target3/Target4. The second IAT was identical to the previous one, except for the targets. Higher scores indicated a preference for Target 3 over Target 4.

Explicit evaluation Target1/Target2. Participants evaluated each target separately on four dimensions (negative/positive, annoying/friendly – worthless/valuable for brand stimuli –, bad/good, unpleasant/pleasant) on 7-point scales. The order of the two targets evaluation was counterbalanced and nested with the order in which the critical blocks were administered in the IAT

(Target1 vs. Target2 first when first presented with positive in the IAT). The average score across the four dimensions was calculated, such that higher scores indicated a preference for Target 1 over Target 2.

Explicit evaluation Target3/Target4. The measure was identical to the previous one, except for the targets. The average score was calculated, such that higher scores indicated a preference for Target 3 over Target 4.

2.2 Analysis plan and data preparation

The data from participants with an error rate higher than 25% in at least one of the learning tasks or the IATs (n = 2) were excluded from the analyses. The meta-analytical findings on the SR highlighted that the recall of the correct IR learned throughout the task is a necessary precondition for the effect to occur (average SR effects of d = 0.68 and d = 0.428 for implicit and explicit evaluations, respectively). In line with these findings, for both Experiments 1 and 2, we present the results based on participants who answered correctly to both the IR memory questions. The results for the unscreened samples are similar, although weaker, and are reported in the Supplementary Materials, Appendix A. Applying these screening criteria led to a final sample of 202 participants[3]. Based on this sample, a sensitivity analysis with power set at .80 and alpha equal to .05 (one-tail, given the presence of a clear directional hypothesis) allows us determining the effect size that can be reliably detected, corresponding to f = 0.1755. For ease of interpretation, transforming this value as Cohen's d yields 0.351 (d = 0.396 for alpha .05, two-tails). In short, this means that the experiment has sufficient power to detect a lower than medium effect size reliably.

The IATs were reliable for both brands (Target1/Target2: α = .86; Target3/Target4: α = .86) and groups (Target1/Target2: α = .86; Target3/Target4: α = .86). Also reliable were the explicit measures for both brands (Target1/Target2: α = .82; Target3/Target4: α = .91) and groups (Target1/Target2: α = .96; Target3/Target4: α = .98). To test the direct SR effect, we conducted two

[3] One participant failed in completing the first explicit measure. This explains the difference in degrees of freedom between implicit and explicit SR effect.

separate factorial ANOVAs with SR condition (Target1 – self vs. Target2 – self) and Target Type (Groups vs. Brands) as fixed factors and both IAT score and explicit liking towards Target 1/2 as dependent variables. We used a similar approach to test the indirect SR effect after we collapsed the SR task and the indirect task conditions into a unique factor – indirect SR condition (Targets indirectly linked to the self vs. other). Doing so led to consider participants who learned to relate Target3 with the target prior categorized with the self (and Target4 with the target prior categorized with others) as belonging to the Target3-self condition. Instead, participants who learned to relate Target4 with the target prior categorized with the self (and Target3 with the target prior categorized with others) were assigned to the Target4-self condition. Indirect SR condition and Target Type were fixed factors in two factorial ANOVAs with IAT scores and explicit liking towards Target3/4 as criteria. Descriptives for both experiments are reported in Tables 1-4.

2.3 Results and discussion

2.3.1 Direct SR effect. On implicit evaluation, we found a significant main effect of SR condition, $F(1, 198) = 20.08$, $p < .001$, $\eta_p^2 = .09$, a main effect of target type, $F(1, 198) = 14.22$, $p < .001$, $\eta_p^2 = .07$ (higher IAT score for brands than for groups) and no significant interaction, $F(1, 198) = 1.83$, $p = .178$, $\eta_p^2 = .01$. On explicit evaluation, there was a significant effect of SR condition, $F(1, 197) = 15.58$, $p < .001$, $\eta_p^2 = .07$, a main effect of target type, $F(1, 197) = 4.66$, $p = .032$, $\eta_p^2 = .02$, and a significant interaction, $F(1, 197) = 6.08$, $p = .015$, $\eta_p^2 = .03$. We further inspected this interaction and observed that the effect of the SR manipulation was significant on explicit evaluation towards groups, $F(1, 98) = 12.25$, $p = .001$, $\eta_p^2 = .11$, and marginal when brands were used as a target, $F(1, 99) = 3.30$, $p = .073$, $\eta_p^2 = .03$. In general, the target paired directly with the self was thus evaluated more positively than the target paired with others at both implicit and explicit level. Implicit and explicit evaluations were positively correlated, $r\ (199) = .23$, $p = .001$.

2.3.2 Indirect SR effect. On the implicit evaluation, we found a significant effect of the Indirect SR condition, $F(1, 198) = 11.04$, $p = .001$, $\eta_p^2 = .05$, a main effect of target type, $F(1, 198) = 50.65$,

$p < .001$, $\eta_p^2 = .20$, and no significant interaction, $F(1, 198) = .01$, $p = .935$, $\eta_p^2 = .000$. On explicit evaluation, there was again a significant effect of the Indirect SR condition, $F(1, 198) = 7.39$, $p = .007$, $\eta_p^2 = .04$, no main effect of target type, $F(1, 198) = 1.26$, $p = .263$, $\eta_p^2 = .01$, and no significant interaction, $F(1, 198) = 1.00$, $p = .319$, $\eta_p^2 = .01$. Thus, the Target paired indirectly with the self was evaluated more positively than the Target paired indirectly with others at both implicit and explicit level. Also, implicit and explicit evaluations resulting from the indirect SR manipulation were positively correlated, $r\ (200) = .24$, $p = .001$.

We also looked at correlations between direct and indirect outcomes. The implicit indirect SR effect correlated with both implicit and explicit direct SR effect, $r\ (200) = .17$, $p = .015$ and $r\ (199) = .22$, $p = .002$, respectively. Instead, the explicit indirect SR effect correlated with one's direct explicit SR, $r\ (199) = .28$, $p < .001$, but not with implicit, $r\ (200) = .12$, $p = .092$.

Experiment 1 replicated the direct SR effect on both implicit and explicit evaluative change (Mattavelli et al., 2017). Participants liked better the target stimuli categorized through the same action as the self than those categorized with others. Crucially, we observed a significant indirect SR effect. The target stimuli that in the indirect task were categorized through the same action as those prior related to the self were implicitly and explicitly liked better than their counterparts. We showed that through IR, the positivity of the self could be transferred to stimuli that either directly or indirectly share a feature with it.

3. Experiment 2

In Experiment 2, we sought to replicate the indirect SR effect observed in Experiment 1 on both implicit and explicit evaluative change. Moreover, we focused on one potential moderator of this effect, that is, individuals' self-esteem. The impact of self-esteem on both the direct and indirect SR effects was tested by measuring explicit and implicit self-esteem. Whereas explicit self-esteem reflects participants' self-reported response to Rosenberg's scale (1965), implicit self-esteem was inferred from performances on a self-esteem IAT. Implicit self-esteem has been defined as the

automatic evaluation of the self that influences spontaneous reactions to stimuli that are somehow related to the self (Greenwald & Banaji, 1995). In line with this definition, the self-esteem IAT assesses one's automatic evaluation of the self by comparing how fast participants respond to "Self" (vs. "Others") and "Good" (vs. "Bad") stimuli when such classes of stimuli are categorized either with the same or an opposed response key. Moreover, because Experiment 1 revealed no systematic influence of Target Type, we thus only used fictitious groups as the target category in Experiment 2.

3.1 Method

3.1.1 Participants and procedure.

A hundred and seventy-four participants (77 women, *M*age = 23.11, *SD* = 2.70) participated in the experiment in exchange for course credit. They were randomly allocated to one of the two SR conditions. Participants in each SR group were assigned to two separate indirect task conditions, mirroring what we did in Experiment 1. Target stimuli were the same fictitious social groups used in Experiment 1. The entire procedure was identical to that used in Experiment 1, except for the fact that all participants started the experiment by completing a Self-Esteem IAT, followed by a self-report measure of self-esteem. Moreover, to reduce any impact of the two measures of self-esteem on the associative learning manipulation involving the self (i.e., the SR task), participants completed a filler task (i.e., circling "r" letters in a text) before the SR task.

3.1.2 Materials.

The two learning tasks, the IR memory questions, the IATs, and the semantic differential to measure implicit and explicit evaluations were identical to those used in Experiment 1.

Moderators.

Implicit self-esteem. In this IAT, participants classified words as quickly as possible to different categories when they appeared in the center of the screen. The categories displayed in the upper left and right corners of the screen were positive, negative, self, and others. The order of the two critical blocks was kept fixed for all participants, with the first critical block being Self+Pleasant. Higher scores indicated higher levels of implicit self-esteem. The number of trials in each block was identical

to that of the other IATs administered in the study. A red X appeared in the middle of the screen for 200ms if the participant did not answer correctly. There was no built-in penalty, and the inter-trial interval was of 500ms. For each attribute and target category, we used five stimuli (see Supplementary Materials – Appendix B).

Explicit self-esteem. Explicit self-esteem was assessed using Rosenberg's (1965) scale. Participants indicated their level of (dis)agreement with the ten items through a 6-point scale (1 = *I totally disagree*, 6 = *I totally agree*).[4]

3.2 Analysis plan and data preparation.

Data from participants who made over 25% of errors in at least one of the classification tasks (SR Task, Target Categorization Task, Self-esteem IAT, IAT1, and IAT2) were excluded from the analyses (n = 14). Moreover, we excluded participants who answered incorrectly to at least one of the two IR memory questions (n = 42, see Supplementary Materials – Appendix A for analyses on the full sample). These screening criteria led to a final sample of 118 participants. A sensitivity analysis with power set at .80 and alpha equal to .05 (one-tail) allows us to estimate the effect size that can be reliably detected corresponding to d = 0.46 (d = 0.52 for two-tails). For the moderation effect, considering a reduced sample size of 101 participants (17 participants who did not complete the Rosenberg's self-esteem), fixing alpha and power as above, the effect size that can be reliably detected is f^2 = .0621 (f^2 = .0792 for two-tails), which is in the region of what conventionally is considered a medium effect size (f^2 = .0625).

The two IATs administered to measure implicit evaluations towards groups were reliable (Target1/Target2: α = .82; Target3/Target4: α = .83). The explicit measures were also reliable (Target1/Target2: α = .90; Target3/Target4: α = .92). Both the self-esteem IAT and the Rosenberg's scale showed good reliability (α = .81 and α = .82, respectively). Assessing both direct and the indirect SR effects implied one-way ANOVAs with SR condition (Target1 – self vs. Target2 – self)

[4] Due to an error in programming, 26 participants did not complete the Rosenberg's self-esteem.

on implicit and explicit evaluations towards Target 1/2 and Indirect SR condition (Target3 – self vs. Target4 - self) on implicit and explicit evaluations towards Target 3/4, respectively. Moderation analyses investigated the impact of implicit and explicit self-esteem in qualifying both direct and indirect SR effects. In a final paragraph, we present an additional set of analyses that test whether the effect of self-esteem (either implicit or explicit) significantly changes between the direct SR effect and the indirect SR effect.

3.3 Results and discussion

3.3.1 Direct SR effect. The SR manipulation affected both implicit and explicit evaluations significantly, $F(1, 116) = 31.06$, $p < .001$, $\eta_p^2 = .21$ and $F(1, 116) = 12.92$, $p < .001$, $\eta_p^2 = .10$, respectively. In essence, the group paired with the self was evaluated more positively than the group paired with others, both at the explicit and implicit levels. Implicit and explicit evaluations were correlated, $r\ (116) = .27$, $p = .003$.

The first set of moderation analyses considered implicit self-esteem as a potential moderator of the direct SR effect. Implicit self-esteem did not moderate the impact of the SR task on either implicit or explicit evaluative change, $\beta = -.05$, $p = .529$, $f^2 = .003$ and $\beta = .04$, $p = .639$, $f^2 = .002$, respectively. We conducted the same analysis with explicit self-esteem as a potential moderator. Explicit self-esteem did not moderate the SR effect on implicit evaluative change, $\beta = -.05$, $p = .595$. However, on explicit evaluation, besides the main effect of the SR manipulation, $\beta = .35$, $p < .001$, $f^2 = .12$, and the non-significant effect of explicit self-esteem, $\beta = .10$, $p = .283$, $f^2 = .01$, we did find a marginal interaction, $\beta = .17$, $p = .074$, $f^2 = .03$. The interaction did not reach the standard level of significance. However, if one considers that we have an unequivocally unidirectional hypothesis, the corresponding one-tail probability is .037. According to this, we explored further the moderation effect using the Johnson-Neyman technique, which probes interactions with continuous moderators (Johnson & Neyman, 1936; Johnson & Fay, 1950). This method identifies the values of the moderator at which the independent variable effect on the criterion transitions from significant to non-significant or vice versa (i.e., regions of significance). Considering that the sample's mean

on the self-esteem scale was 4.52 (*SD* = 0.77), a direct SR effect emerged for individuals with scores higher than 3.96.

3.3.2 Indirect SR effect. Results showed an effect on both IAT score, $F(1, 116) = 10.15$, $p = .002$, $\eta_p^2 = .08$ as well as on explicit, $F(1, 116) = 4.28$, $p = .041$, $\eta_p^2 = .04$, demonstrating an indirect SR effect. Implicit and explicit evaluations were not correlated, $r\ (116) = .08$, $p = .374$. When we looked at the correlation between standard and indirect SR effect, we only found a significant correlation between the two explicit scores, $r\ (116) = .34$, $p < .001$ (all other *r*s < .15, *p*s > .105).

Similar to what we observed with the direct SR effect, implicit self-esteem did not qualify the indirect SR effect on either implicit or explicit evaluative change, β = .02, $p = .826$, $f^2 = .000$ and β = .003, $p = .974$, $f^2 = .000$, respectively. Instead, moderation analyses with explicit self-esteem revealed a different pattern. On implicit evaluation there was a significant main effect of the indirect SR condition, β = .28, $p = .007$, $f^2 = .07$, no effect of explicit self-esteem, β = .01, $p = .889$, $f^2 = .001$, and the interaction term showing a tendency towards significance, β = .17, $p = .098$, $f^2 = .03$. As we did previously, we inspected the interaction with the Johnson-Neyman technique. The indirect SR effect was significant for scores higher than 4.24 on explicit self-esteem. The same analysis on explicit evaluation revealed no main effect of indirect SR, β = .14, $p = .128$, $f^2 = .02$, no effect of explicit self-esteem, β = .07, $p = .459$, $f^2 = .01$, but a significant interaction, β = .20, $p = .034$, $f^2 = .04$. The Johnson-Neyman technique showed that the impact of the indirect SR manipulation on explicit evaluation change was significant at scores higher than 4.69 on explicit self-esteem**.**

3.3.3 Comparing the role of self-esteem on direct versus indirect SR effects.

We computed four new indicators of implicit/explicit evaluative change (i.e., Implicit-Direct SR, Implicit-Indirect SR, Explicit-Direct SR, Explicit-Indirect SR) such that positive scores always reflected a preference for the target that was related with the self in the SR task and the Indirect task. Thus, such scores did not take into account the identity of the target that was either directly

(i.e., Target1 or Target2) or indirectly related (i.e., Target3 or Target4) to the self. These new scores allowed us to compare the direct and indirect effects (see meta-analytical findings below) and to test whether self-esteem did have a differential impact on them. To this aim, we conducted a series of mixed ANCOVAs, with direct and indirect scores (implicit or explicit) as within-subjects outcomes and self-esteem (implicit or explicit) as a covariate. The critical effect of interest here is a significant interaction between the within-subjects variable and the self-esteem measure. This interaction was not significant for implicit self-esteem with the difference between direct and indirect SR effect referred to implicit, $F(1, 116) = .48$, $p = .493$, $\eta_p^2 = .004$, and explicit evaluations, $F(1, 116) = .28$, $p = .597$, $\eta_p^2 = .002$. A similar pattern of findings was found with explicit self-esteem was entered into the model as a covariate. Although at a descriptive level, explicit self-esteem showed a stronger impact on the indirect effect, the interaction with the difference between direct and indirect effect did not reach significance on either implicit or explicit evaluative change, $F(1, 99) = 2.43$, $p = .123$, $\eta_p^2 = .02$ and $F(1, 99) = .43$, $p = .515$, $\eta_p^2 = .004$, respectively. Note, however, that explicit self-esteem had a significant main effect on the explicit SR, $F(1, 99) = 5.75$, $p = .018$, $\eta_p^2 = .055$, meaning that the SR effect was stronger with increasing explicit self-esteem.

Experiment 2 replicated the indirect SR effect on both implicit and explicit evaluative changes. Moreover, we wanted to investigate whether and how implicit and explicit self-esteem moderated the impact of both direct and indirect SR manipulations. Whilst implicit self-esteem showed no impact in qualifying either the direct or the indirect SR effect, the explicit self-esteem played a more relevant role. First, a tendency towards an interaction effect revealed that the direct SR effect on explicit evaluations required a moderately high level of explicit self-esteem. Considering the indirect SR effect, we found explicit self-esteem to be even more important in qualifying the observed change on explicit liking as well as demonstrating a tendency to qualify the effect of the manipulation on implicit evaluation. However, we did not found evidence for a significant difference in the role of self-esteem (either implicit or explicit) to qualify the direct and indirect SR effects. On the other side, recall that the Johnson-Neyman technique identified

differential thresholds in the score of explicit self-esteem for the significant effect of the SR manipulations on explicit evaluations: 3.96 for the direct SR and 4.65 for the indirect SR effect. This difference in the thresholds is not trivial and almost as large as one standard deviation (0.90 SD). Putting together the relevant results, a noticeably higher score in explicit self-esteem is needed for a significant indirect SR effect relative to a direct one, explicit self-esteem tends to have a more important role for indirect compared to direct SE effects, yet this difference was not statistically significant.

4. Meta-analytical findings: Does evaluative change decrease as the number of intersections grows?

With the aim of proving the robustness of learning via multiple intersecting regularities, data from the two studies were collapsed into a single dataset ($N = 320$). The focal test involved a direct comparison between the direct and the indirect SR effect. We used the same scores computed to test the differential impact of self-esteem on the direct versus indirect SR effect. First, direct and indirect evaluations significantly correlated when measured both implicitly and explicitly, $r\ (320) = .16$, $p = .003$ and $r\ (319) = .30$, $p < .001$, respectively. Results from a paired-sample *t*-test on implicit evaluations was only marginally significant, $t(319) = 1.97$, $p = .050$, $d = 0.14$. The same analysis showed evidence for no reduction in the effect on explicit evaluations, $t(318) = 1.23$, $p = .221$, $d = 0.08$ (see Table 5 for descriptives and effect sizes). In essence, these analyses offer further evidence to the robustness of intersecting regularities to produce evaluative learning effects. When the number of intersections between the source and a target gets more complex and articulated, evaluative change remains significant and shows little statistical evidence of a decrease in magnitude.

5. General discussion

Across two experiments, we investigated whether the positivity carried by the self can transfer across multiple stimuli through complex chains of intersecting regularities. We replicated the standard (direct) SR effect, showing that objects classified through the same action as the self in

a categorization task are better liked than those classified through the same action as the contrast category ‘others’. Crucially, we found the first evidence of an indirect SR effect. After being exposed to an additional learning task in which stimuli previously related to the self were categorized with a third class of stimuli – while stimuli related to others went with a fourth class of stimuli – the third class received greater liking than its counterpart.

In Experiment 1, an indirect SR effect was observed when targeting both fictitious brands and groups (although the effect on explicit evaluation was marginal on brands). Indirect evaluative change established via associative learning procedure has been demonstrated previously in EC literature (Hammerl & Grabitz, 1996; Walther, 2002; Gast & De Houwer, 2012). Only one single empirical study tested the same indirect effect in associative learning via IR (Hughes et al., 2016, Experiment 4). Thus, we provided new evidence that IR is a powerful way to generate indirect links between valenced source and neutral stimuli. Moreover, all previous studies used ‘general’ valenced stimuli as attitudinal sources. For instance, Walther (2002) showed that both positive and negative valence carried by a source could spread across a two-step EC procedure. We expanded previous literature by showing that an indirect transfer of liking is possible when using a more complex attitudinal source, the self. Experiment 1 showed that, by increasing the sequence of intersecting elements, the transfer of self-positivity induced by the SR task does not confine to stimuli that share the first-order link with it.

In Experiment 2, beside replicating both the direct and the indirect SR effect, we measured individuals’ self-esteem at both implicit and explicit level and tested whether this dispositional characteristic qualified both the standard and the indirect SR effect. First, implicit self-esteem did not moderate either the direct or the indirect SR effect, regardless of whether the criterion was an implicit or explicit evaluation change. Second, the direct SR effect on explicit evaluation change was tendentially moderated by participants’ explicit self-esteem scores. Finally, the moderating role of explicit self-esteem showed stronger as the relationship between the self and the target became less direct (i.e., indirect SR effect). Such a pattern was evident in both explicit and implicit indirect

SR effects. However, a comparison of the impact of self-esteem on the direct and the indirect SR effects showed that the difference between the direct and the indirect SR did not interact significantly with either implicit or explicit self-esteem. This result does not support the idea of an enhanced impact of self-esteem when the intersections between the self and the target increases in complexity. However, we believe that the observed tendency (i.e., a higher impact of explicit self-esteem on the indirect SR effect) would benefit from further dedicated studies.

Concerning the role of self-esteem, our findings diverge from past research on the relations between SR and self-esteem. Prestwich and colleagues (2010, Experiment 2) found that the SR effect on implicit evaluation was moderated by implicit, but not explicit, self-esteem. There are, however, some methodological aspects that might account for the observed discrepancy. First, in Prestwich et al. (2010), there was no measure of explicit evaluation change following the SR manipulation. Second, our SR manipulation involved a matched control condition (i.e., the experimental and the control group solely differed regarding the target object paired with the self). In contrast, in Prestwich et al. (2010), participants in the control condition did not perform any task. Third, we used fictitious social groups in the study, while Prestwich et al. (2010) targeted existing drink products. Having said that, we do not rule out the possibility that the lack of moderation of implicit self-esteem might be attributable to measurement validity. In particular, past research has shown that implicit self-esteem is poorly correlated with its explicit counterpart (Bosson, Swann, & Pennebaker, 2000), which is what we found in Experiment 2, $r(101) = .06$, $p = .571$. Moreover, implicit self-esteem suffers from low convergent validity (Falk, Heine, Takemura, Zhang, & Hsu, 2013). And it has low predictive validity of different criteria that are meant to be related to one's self-view, such as well-being, depression, and negative self-thoughts (Buhrmester, Blanton, & Swann, 2011). One major problem with measuring implicit self-esteem has to do with the fact that the self is a multifaceted concept (e.g., Marcus & Wurf, 1987). Thus, the automatic associations between evaluative concepts and the self, inferred from performances in the IAT, can vary depending on what facet of the self is momentarily salient and accessible for the individual

(Gawronski & Bodenhausen, 2007). Taken together, all these limitations of implicit self-esteem might account for the inconsistency observed across studies.

What is indisputable from the current findings is the joint power of the self and intersecting regularities to shape both implicit and explicit evaluations. From a functional perspective, learning can be conceived as any change in the behavior of an organism that results from regularities in the environment of the organism (De Houwer, Barnes-Holmes, & Moors, 2013). Thus, both direct and indirect SR effects are instances of learning originating from intersections between regularities in the environment that involve the self and a target object. The direct SR effect shows that the positivity of a source stimulus (i.e., the self) can transfer to a target stimulus provided that the two share an action response. The indirect SR effect pushes further on this idea, demonstrating that once a direct relationship is established, the target stimulus can function itself as the intersecting element that connects the self with a second class of target stimuli. Because of this chain of multiple intersections, individuals learn that source and target stimuli are somehow related. According to the *shared features principle* (De Houwer, Richetin, Hughes, & Perugini, 2019; Hughes, De Houwer, Mattavelli, & Hussey, in press), when people learn that two stimuli share one thing, they infer that other features between the two are also shared. Applied to both direct and indirect SR effects, this principle implies that (i) the same behavior (i.e., pressing the same key) for direct SR and (ii) the same common stimulus (i.e., a target category) for the indirect SR serve as basic commonalities between the self and target stimuli. With no other information available, people use such commonalities to infer that the self and the target stimuli are similar regarding other features or properties (i.e., liking).

Our findings also have theoretical implications concerning the mental processes responsible for the direct and indirect SR effects. Associative theories of learning assume that the detection of environmental regularities results in mental links between the representations in memory of those stimuli involved in regularities (e.g., Baeyens, Eelen, Crombez, & Van den Bergh, 1992; Baeyens, Eelen, & Crombez, 1995). It would, therefore, be plausible to predict that the strength of the

evaluative effect would reflect the strength of the relationship between source and target stimuli. One such index of strength might consist of the number of intersections required to establish a relation between stimuli: The higher the number of the intersections between stimuli, the weaker their associative strength. As our meta-analytical findings showed, we did not find clear-cut statistical evidence in favor of a reduction in the SR effect when comparing the direct and the indirect effect. On both implicit and explicit evaluations, introducing one additional set of intersecting features between the self and the final target did not lead to a significant reduction in evaluative change. Thus, at least concerning this prediction, a purely associative interpretation will have difficulties to account for the current findings.

Based on the propositional account of human associative learning (De Houwer, 2007, 2009, 2014; 2019; Mitchell, De Houwer, & Lovibond, 2009), learning effects are mediated by propositions that the organism forms about newly acquired relations between stimuli in the environment. For instance, people learn to like Nespresso when they see it paired with George Clooney because the pairing might result in a proposition that qualifies the relation between the two stimuli (e.g., “George Clooney likes Nespresso”). In a similar fashion, people might like the target stimulus related to the self in the SR task because they can elaborate some sort of proposition upon the fact that the self and the target share a common response key (e.g., “The group Lerriani and I have some features in common”). Similar reasoning should then apply for the indirect SR effect, through which participants learn that one target stimulus is related to the target priorly related to the self (e.g., “The group Craviani and I have one group in common”). When it comes to explaining evaluative learning effects, many propositions can potentially account for the observed outcome. In the present research, an alternative explanation envisages that no relation between the self and the target stimulus is elaborated propositionally. The direct SR effect would be mediated by a proposition that relates the target with valence (e.g., “The group Lerriani is positive”) and the indirect SR effect by another proposition that relates the new target with the target known to be

positive (e.g., "The group Craviani is related to the positive group").[5] However, in assuming a marginal, if not absent, role played by the self on both the direct and the indirect SR effect, this alternative interpretation does not explain why a target stimulus should become positive after the SR task. Based on the current findings and prior SR studies (Prestwich et al., 2010; Mattavelli et al., 2019, Experiment 3), self-esteem represents an important variable in qualifying the extent to which stimuli learned to be related to the self acquire positivity. Therefore, at least when it comes to the direct manifestation of the SR effect, the proposition from the alternative interpretation described earlier (e.g., "The group Lerriani is positive") should be conditional upon two premises that account for (i) the relationship between the self and the target and (ii) the positivity of the self (e.g., "*If* I am positive *and* the group Lerriani is related to myself, *then* the group Lerriani is positive"). However, we acknowledge that deciding what propositions are responsible for the direct and indirect SR effects is rather speculative at this stage. Future studies should investigate better this point to clarify the nature of the propositions that are responsible for the change in evaluations of target stimuli that relate to the self, either directly or indirectly.

This research is not without limitations. First, we investigated the indirect SR effect in one single direction. Namely, we established a first set of relationships between the self and one intermediate target (and other-stimuli with a second intermediate target), and then a second set between the intermediate targets and two classes of final targets. This procedure is quite similar to the one used in second-order conditioning (i.e., once a stimulus has been conditioned, it can serve as a source for new conditioning, e.g., Hammerl & Grabits, 1996; Walther, 2000). We did not investigate whether the SR effect occurs when the order of established relationships is reversed, with two conditional stimuli related first, and then one of them put in relation with the self (e.g., sensory preconditioning, e.g., Ecker & Bar Anan, 2019). Future studies should investigate whether indirect SR effects are qualified by the order with which stimuli relationships are established.

[5] We thank an anonymous reviewer for suggesting this alternative explanation.

Second, as we argued in a previous paragraph, this research tested self-esteem as a unique moderator of the observed effects of the self on liking. However, some might say that the complexity of self-concept goes beyond what is reflected by one's self-view. In fact, the self also has other specific structural characteristics that are fundamental to its psychological function (e.g., self-concept clarity, Campbell et al., 1996). Although research focusing on the role of the self in shaping evaluations has primarily focused on self-esteem as a general indicator of self-concept valence, future investigations could take additional properties of the self into account when looking at both direct and indirect SR effects.

In sum, across two experiments, we demonstrated that the positivity of the self could transfer towards objects that are indirectly related to the self. This indirect transfer of liking occurred via intersecting regularities. Taken together, the two experiments proved (i) that the self represents a powerful source of positivity, (ii) that such positivity can easily transfer towards other objects provided that some, even minimal, elements of intersection are shared with the self, and (iii) that the SR effect increases with increased explicit self-esteem. The effect of explicit self-esteem however, was not statistically different for direct versus indirect SR effect, despite some evidence at a descriptive level and different thresholds for significant direct vs. indirect SR effects in the Johnson-Neyman decomposition analyses. Besides additional focused studies on this issue, there is the need to identify other features that can modulate the SR effect when transitioning from direct to indirect. Moreover, as we know from previous studies, the self is not only characterized by its positivity (e.g., Rogers, Kuiper, & Kirker, 1977) and some of the properties of the self can be acquired by new objects related to the self via IR (Mattavelli et al., 2017). Future research can test whether these cognitive properties (e.g., memory, accessibility, identification) might also transfer to novel stimuli through indirect intersections of regularities.

References

Allport, G. W. (1935). Attitudes: a handbook of social psychology. *Clark University Press, Worcester.*

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

Baeyens, F., Eelen, P., Crombez, G., & Van den Bergh, O. (1992). Human evaluative conditioning: Acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behaviour research and therapy*, *30*(2), 133-142. http://dx.doi.org/10.1016/0005-7967(92)90136-5

Bosson, J. K., Swann Jr, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited?. *Journal of Personality and Social Psychology, 79*(4), 631-643. https://doi.org/10.1037/0022-3514.79.4.631

Buhrmester, M. D., Blanton, H., & Swann Jr, W. B. (2011). Implicit self-esteem: Nature, measurement, and a new way forward. *Journal of Personality and Social Psychology, 100*(2), 365-385. https://doi.org/10.1037/a0021341

Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavallee, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of Personality and Social Psychology, 70(1*), 141-156. http://dx.doi.org/10.1037/0022-3514.70.1.141

Corneille, O., & Hütter, M. (2020). Implicit? What Do You Mean? A Comprehensive Review of the Delusive Implicitness Construct in Attitude Research. *Personality and Social Psychology Review.* http://dx.doi.org/10.1177/1088868320911325

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, *10*(2), 230-241. https://doi.org/10.1017/S1138741600006491

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior*, *37*(1), 1-20. https://doi.org/10.3758/LB.37.1.1

De Houwer, J. (2014). Why a propositional single-process model of associative learning deserves to be defended. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 530-541). New York: Guilford Press.

De Houwer, J. (2019). Moving beyond System 1 and System 2: Conditioning, implicit evaluation, and habitual responding might be mediated by relational knowledge. *Experimental Psychology, 66*(4), 257-265. https://doi.org/10.1027/1618-3169/a000450

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. https://doi.org/10.3758/s13423-013-0386-3

De Houwer, J., Richetin, J., Hughes, S., & Perugini, M. (2019). On the assumptions that we make about the world around us: A conceptual framework for feature transformation effects. *Collabra: Psychology, 5(1*). 10.1525/collabra.229

De Houwer J, Teige-Mocigemba S, Spruyt A, & Moors A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*. *135(3*), 347–368. https://doi.org/10.1037/a0014211

Ecker, Y., & Bar-Anan, Y. (2019). Sensory preconditioning of evaluation requires accurate memory of the co-occurrence between the neutral stimuli. *Journal of Experimental Social Psychology, 85*, 103886. https://doi.org/10.1016/j.jesp.2019.103886

Falk, C. F., Heine, S. J., Takemura, K., Zhang, C. X., & Hsu, C. W. (2015). Are implicit self-esteem measures valid for assessing individual and cultural differences?. *Journal of Personality, 83*(1), 56-68. https://doi.org/10.1111/jopy.12082

Gast, A., & De Houwer, J. (2012). Evaluative conditioning without directly experienced pairings of the conditioned and the unconditioned stimuli. *The Quarterly Journal of Experimental Psychology*, *65*(9), 1657-1674. https://doi.org/10.1080/17470218.2012.665061

Gawronski, B., & Bodenhausen, G. V. (2007). Unraveling the processes underlying evaluation: Attitudes from the perspective of the APE model. *Social Cognition, 25*(5), 687-717. https://doi.org/10.1521/soco.2007.25.5.687

Gawronski, B., Bodenhausen, G. V., & Becker, A. P. (2007). I like it, because I like myself: Associative self-anchoring and post-decisional change of implicit evaluations. *Journal of Experimental Social Psychology*, *43*(2), 221-232. https://doi.org/10.1016/j.jesp.2006.04.001

Gawronski, B., & LeBel, E. P. (2008). Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology*, *44*(5), 1355-1361. https://doi.org/10.1016/j.jesp.2008.04.005

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4-27. https://doi.org/10.1037/0033-295X.102.1.4

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109(1)*, 3-25. https://doi.org/10.1037/0033-295X.109.1.3

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

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

Hammerl, M., & Grabitz, H. J. (1996). Human evaluative conditioning without experiencing a valued event. *Learning and Motivation*, *27*(3), 278-293. https://doi.org/10.1006/lmot.1996.0015

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: a meta-analysis. *Psychological Bulletin*, *136*(3), 390-421. https://doi.org/10.1037/a0018916

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. http://dx.doi.org/10.1037/xge0000100

Hughes, S., De Houwer, J., Mattavelli, S., & Hussey, I. (in press). The shared feature principle: If two objects share a feature, people assume those objects also share other features. *Journal of Experimental Psychology: General.* http://dx.doi.org/10.1037/xge0000777

Johnson, P. O., & Fay, L. F. (1950). The Johnson-Neyman technique, its theory and application. *Psychometrika, 15*(4), 349–367. https://doi.org/10.1007/BF02288864

Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs, 1*, 57–93.

Marcus, H., & Wurf, E. (1987). The dynamic self-concept: a sociopsychological research. *Annual Review of Psychology, 6,* 1-14. https://doi.org/10.1146/annurev.ps.38.020187.001503

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. http://dx.doi.org/10.1016/j.jesp.2017.02.006

Mattavelli, S., Richetin, J., & Perugini, M. (2019). Not all positive categories are alike: Exploring the superiority of the self as a positive source for associative attitude change via intersecting regularities. *European Journal of Social Psychology*, *49(3)*, 574-588. http://dx.doi.org/10.1002/ejsp.2518

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

Perkins, A. W., & Forehand, M. R. (2012). Implicit self-referencing: The effect of nonvolitional self-association on brand and product attitude. *Journal of Consumer Research*, *39*(1), 142–156. http://dx.doi.org/10.1086/662069

Perugini, M., Richetin, J., & Zogmaister, C. (2014). Indirect measures as a signal for evaluative change. *Cognition & Emotion*, *28*(2), 208-229. http://dx.doi.org/10.1080/02699931.2013.810145

Perugini, M., Zogmaister, C., Richetin, J., Prestwich, A., & Hurling, R. (2013). Changing implicit attitudes by contrasting the self with others. *Social Cognition*, *31*(4), 443-464. http://dx.doi.org/10.1521/soco_2012_1003

Prestwich, A., Perugini, M., Hurling, R., & Richetin, J. (2010). Using the self to change implicit attitudes. *European Journal of Social Psychology*, *40*(1), 61-71. http://dx.doi.org/10.1002/ejsp.610

Richetin, J., Mattavelli, S., & Perugini, M. (2016). Increasing implicit and explicit attitudes toward an organic food brand by referencing to oneself. *Journal of Economic Psychology*, *55*, 96-108. http://dx.doi.org/10.1016/j.joep.2016.01.006

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, *35*(9), 677-688. https://doi.org/10.1037/0022-3514.35.9.677

Rosenberg, M. (1965). Rosenberg self-esteem scale (RSE). *Acceptance and Commitment Therapy. Measures package*, *61*, 52.

Rozin, P., & Millman, L. (1987). Family environment, not heredity, accounts for family resemblances in food preferences and attitudes: A twin study. *Appetite, 8(2),* 125-134. http://dx.doi.org/10.1016/S0195-6663(87)80005-3

Walther, E. (2002). Guilty by mere association: evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology*, *82*(6), 919-934. http://dx.doi.org/10.1037/0022-3514.82.6.919

Walther, E., & Trasselli, C. (2003). I like her, because I like myself: self-evaluation as a source of interpersonal attitudes. *Experimental Psychology*, *50*(4), 239-246. https://doi.org/10.1026//1618-3169.50.4.239

Yamaguchi, S., Greenwald, A. G., Banaji, M. R., Murakami, F., Chen, D., Shiomura, K., ... & Krendl, A. (2007). Apparent universality of positive implicit self-esteem. *Psychological Science*, *18*(6), 498-500. https://doi.org/10.1111/j.1467-9280.2007.01928.x

Table 1. *Means and Standard Deviations for the direct SR effect in Experiment 1.*

	Groups		Brands	
	Target1 - Self *n* = 50	Target2 - Self *n* = 51	Target1 - Self *n* = 53	Target2 - Self *n* = 48
IAT score	.04 (.47)	-.17 (.52)	.39 (.48)	-.01 (.47)
Explicit evaluation score	.55 (2.28)	-1.30 (2.94)	.46 (1.08)	.03 (1.28)

Table 2. *Means and Standard Deviations for the indirect SR effect in Experiment 1.*

	Groups		Brands	
	Target3 - Self *n* = 50	Target4 - Self *n* = 51	Target3 - Self *n* = 56	Target4 - Self *n* = 45
IAT score	.29 (.44)	.09 (.44)	-.16 (.47)	-.37 (.47)
Explicit evaluation score	.40 (2.81)	-.80 (2.74)	-.29 (1.55)	-.84 (1.74)

Table 3. *Means and Standard Deviations for the direct SR effect in Experiment 2.*

	Target1 - Self *n* = 59	Target2 - Self *n* = 59
IAT score	.16 (.47)	-.33 (.48)
Explicit evaluation score	.38 (2.21)	-1.03 (2.04)

Table 4. *Means and Standard Deviations for the indirect SR effect in Experiment 2.*

	Target3 - Self *n* = 98	Target4 - Self *n* = 94
IAT score	.19 (.41)	-.06 (.42)
Explicit evaluation score	.08 (2.45)	-.75 (1.83)

Table 5. *Estimates, effect sizes and p-values for direct and indirect SR effects in Experiment 1 and 2 combined.*

	M (*SD*)	*Cohen's d* [95% CI]	*p*
Implicit – direct SR	.19 (.50)	.38 [.26; .49]	< .001
Implicit – indirect SR	.12 (.47)	.26 [.14; .37]	< .001
Explicit – direct SR	.62 (2.10)	.30 [.18; .41]	< .001
Explicit – indirect SR	.45 (2.26)	.20 [.09; .31]	< .001

Note. All the effects reported here are calculated such as more positive scores reflect a preference for the target (directly or indirectly) related with the self.

SR TASK			
BLOCK 1		BLOCK 2	
STIMULUS	RESPONSE	STIMULUS	RESPONSE
Self	yellow	Others	blue
Target 1	yellow	Target 2	blue
Others	blue	Self	yellow
Target 2	blue	Target 1	yellow
INDIRECT TASK			
BLOCK 1		BLOCK 2	
STIMULUS	RESPONSE	STIMULUS	RESPONSE
Target 1	yellow	Target 2	blue
Target 3	yellow	Target 4	blue
Target 2	blue	Target 1	yellow
Target 4	blue	Target 3	yellow

Figure 1. *Schematic overview of the stimuli and response involved in the SR and Indirect task*