RESEARCH ARTICLE

Using the Self-Referencing Task to Produce Durable Change on Food Evaluations Measured via the IAT

Simone Mattavelli, Juliette Richetin and Marco Perugini

In many daily life situations, eating behaviour manifests itself under conditions of automaticity. Associative learning procedures have proven reliable to change food items' evaluations inferred from performances in indirect tasks, such as the Implicit Association Test (IAT). Targeting two alternative food brands, we investigated the impact of the Self-Referencing (SR) task on IAT performances measured immediately after the manipulation and one week later. Capitalizing on the structural features of the SR task, a first study ($N = 145$) demonstrated the durable effect of the manipulation on the IAT. An advantage in automatic responding for the snack brand paired with the self was detected right after the SR task and one week later. Instead, the SR manipulation showed no impact on self-reported evaluations. Moreover, a semantic priming task administered right after the SR task indicated response facilitation for the self-related target brand when self-stimuli were used as primes. Experiment 2 ($N = 268$, pre-registered) targeted two alternative food brands and replicated the lasting effect on the IAT, thus demonstrating the generalizability of the effect. Moreover, we extended the results' validity by showing that SR can generate a durable effect on a behavioural choice task. We discussed the effectiveness of the SR in producing lasting effects on the IAT and other instances of automatic behavior and the potential implications for research in the food domain.

Classic attempts to modify food-related behaviours have long relied on information interventions, based on the idea that deliberation upon our actions' consequences determines what we eat. However, the impact of such interventions can be limited (Webb & Sheeran, 2006). First, whether or not the content of an informative message can influence behaviour might be dependent upon the relevance of such content at the individual level. Second, it is often the case that food-related behaviour occurs within environmental conditions that limit cognitive resources. For instance, a man shopping at the supermarket has to decide whether to choose one food item or the other very quickly. In line with this idea, researchers have started to study food-related behaviours as automatic phenomena. This tendency has produced two main outcomes. First, the development and the rapid growth of indirect methods. Indirect methods assess evaluations of target food by measuring behavioural performances (e.g., binary choice) in tasks characterized by conditions of automaticity (e.g., time constraints). Second, the advent of associative learning procedures as a new class of intervention strategies capable of affecting automatic evaluations

and behaviours. Although associative learning procedures have proven reliable to produce an immediate change in automatic evaluations (e.g., Hollands, Prestwich, & Marteau, 2011), less attention has been devoted to their effects in the long-term. Because food behaviour performed under automaticity conditions is very relevant, scientists need to understand what variables are more likely to produce durable change on evaluations and behaviours performed under automaticity. The present investigation addresses this issue. By focusing on automatic food evaluations indexed by the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), we investigate both the immediate and the lasting effect of an associative procedure, namely the Self-Referencing task.

In the last decades, researchers in health and social psychology have devoted increasing attention to evaluations of target stimuli inferred from indirect measures' performances. Indirect methods are tasks that assess the evaluation of a given stimulus (e.g., a food item) by measuring individual behaviour under conditions of automaticity that vary based on the nature of the task itself. Among such measures, the most known is the IAT. In the IAT, participants rapidly categorize target stimuli (e.g., products of two alternative food brands) and attribute stimuli (e.g., positive and negative words). In one critical block, categories for target and attribute stimuli are combined such that participants need to press a key for one class of target

University of Milano-Bicocca, IT

Corresponding author: Simone Mattavelli
(simone.mattavelli@unimib.it)

stimuli and one class of attribute stimuli (e.g., press left for pictures of Brand 1 and positive words) and another key for the other class of target and attribute stimuli (e.g., press right for pictures of Brand 2 and negative words). In another critical block, the response assignments for the target categories are reversed (e.g., press left for pictures of Brand 2 and positive words; press right for pictures of Brand 1 and negative words). When attribute categories refer to positive and negative valence, differences in performance between the two blocks are interpreted as evidence for differences in the automatic evaluation of the target categories (e.g., better performance in the Brand 1-positive block would reflect a more positive automatic evaluation of Brand 1 than of Brand 2). Evaluations measured through the IAT have shown to be related to self-reported consumption (e.g., Conner et al., 2007), habits (Maison, Greenwald, & Bruin, 2001), choices (e.g., Richetin et al., 2007), and food purchases (Prestwich, Hurling, & Baker, 2011). Moreover, the relation between performances on the IAT and actual behavioural choice (e.g., choosing between fruits and snacks) increases when individuals have limited resources (e.g., high cognitive load; Friese, Hoffman, & Wanke, 2008).

Indirect measures have been originally conceived as capable of revealing those inner preferences or beliefs that people are either unable or unwilling to report (Nosek, Banaji, & Greenwald, 2002). This conceptualization was soon followed by the idea that responses to indirect measures are mediated by mental associations (e.g., Gawronski & Bodenhausen, 2006, 2011). Critically, however, this idea has found scarce empirical evidence. As one example, research repeatedly shows that associative explanations of IAT performance fail to account for empirical findings (Corneille & Stahl, 2019; De Houwer, 2014a). Following more recent elaborations (Van Dessel et al., 2020), we conceive the IAT performance as a behaviour performed under certain procedural conditions. The behaviour captured by the IAT (i.e., categorizing target-attribute pairs under the same response key) is automatic in the sense that it occurs under specific conditions of automaticity (i.e., time constraints). Because many of our decisions in daily life occur under similar conditions, the IAT might constitute a valid test for the effectiveness of interventions aimed at producing changes in food-related behaviour performed under time constraints.

Critically, changes in evaluations reflected by individual responses on indirect tasks often fail to last over time, and the temporal stability of such evaluative changes is lower when compared to self-reported evaluations (Gawronski et al., 2017). The key to detecting any stability of changes in automatic evaluations more robustly is to match the conditions of acquisition with the conditions of measurement of such changes (Gawronski, 2019). Namely, when the environmental conditions within which changes are trained (and acquired) match those within which such changes are measured, automatic changes should show more stability and last longer. Corroborating this idea is a recent research by Chen, Holland, Quandt, Dijksterhuis, and Veling (2019). The authors conditioned responses to food stimuli by asking participants to consistently

respond to certain food items (go items) and not respond to others (no-go items) in a go/no-go training. Next, they measured preferences via a task that somehow mirrored the training phase structure, with participants asked to choose (either to go or not go) repeatedly between the go and the no-go items. The preference for the go items, inferred from the superior number of times in which participants 'went' for the go items, was still detectable after one week. Yet, less is known about the features and the conditions that make intervention strategies more likely to produce lasting changes when performances in the IAT reflect such changes.

Many studies have concentrated on the impact of different intervention strategies thought to influence automatic evaluations. The present work focuses on associative learning procedures. Associative learning procedures produce changes in behaviour that result from regularities in the presence of events (De Houwer, 2014b). One of the most studied and known procedures is evaluative conditioning (EC). EC is defined as the change in a neutral stimulus' liking due to its repeated presentation with another valenced stimulus (De Houwer, 2007). EC procedures represent a powerful tool to change self-reported and automatic evaluations (see Hofmann et al., 2010). For instance, pairing food items with images of health-related consequences (e.g., the picture of obese individuals) can change the evaluation of the targeted food items reflected by the IAT performance (e.g., Hollands, Prestwich, & Marteau, 2011). However, when it comes to producing long-term changes, EC might not be the best means. For instance, in a set of studies on nine interventions aimed to reduce 'implicit bias', Lai and colleagues (2016) found that EC led to an immediate change on the IAT, but had no lasting effect days later. This research demonstrates that producing lasting changes on the IAT via EC is difficult and highlights the need to explore new routes and procedures that can foster such changes.

Hughes, De Houwer, and Perugini (2016) proposed a new associative learning procedure, and its effect on behaviour is defined as learning via Intersecting Regularity (IR). 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. The original procedure proposed by the authors was based on a set of operant contingencies (i.e., pressing one key when either a neutral target or a valenced source appeared on screen) through which participants learned to relate one target stimulus with a positively valenced source and another target stimulus with a negatively valenced source. The Self-Referencing task (SR task; Prestwich et al., 2010; Perkins & Forehand, 2012) is based on such operant contingencies. 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. Because people typically hold a positive view of themselves (Yamaguchi et al., 2007), learning that one stimulus shares something with the self produces a more positive evaluative response to such stimulus. Thus, in the SR task, learning occurs via

training participants to respond to the same response key when either one target stimulus or self-related stimuli appear on the screen. The SR task's effect has proven reliable in affecting self-reported and automatic evaluations, particularly when the latter is measured via the IAT (see Mattavelli et al., 2017 for a meta-analysis). Moreover, some recent studies have attempted to use the SR task to change automatic evaluations towards food items. For instance, repeatedly training participants to press the same key when green vegetables or the self-stimuli appeared on the screen produced more positive IAT scores towards green vegetables and influenced readiness to increase consumption among participants with negative pre-existing explicit attitudes (Mattavelli et al., 2017; see also Demartini et al., 2019 for a similar demonstration on low glycemic index products). In the SR task, new relationships between source and target stimuli are operationalized by a categorization task where the source and target stimuli are assigned to a common response key. Therefore, the SR task is effective for changing automatic evaluations towards targeted food items measured via the IAT. More importantly, because the conditions under which such evaluations are acquired in the SR task match those in which they are measured (i.e., throughout the IAT), the SR task could be ideal for making such change persistent over time.

Two crucial features characterize the SR task. First, it is based on intersecting regularities. Evaluative learning effects result from a categorization task that trains participants to perform the same action in response to source and target stimuli. Second, such changes in the evaluation of the target stimuli are due to one special source, that is, the self. In this perspective, our aim was twofold. We first aimed to increase food items' positivity by training participants to perform the same action in response to stimuli belonging to the target food and stimuli belonging to the self. Second, we examined the durability of the expected effects by investigating whether changes in the food items' evaluations could last at least one week, measuring such changes via an IAT and self-report ratings. Recent research has shown the effectiveness of action-based intervention (i.e., Go/No-go training) on behaviour change towards food items (Chen et al., 2019). However, few published studies have tested the lasting effects of evaluative learning procedures targeting the IAT. In particular, none have done so using a learning procedure that capitalizes on (i) intersecting regularities and (ii) the self as a positive source.

Preliminary study

We conducted a preliminary study to assess the impact of the SR manipulation on IAT scores towards the targeted food items. The results also provided an effect size estimate allowing us to appreciate whether the subsequent main study was sufficiently powered.

One hundred nine subjects (66 women, 42 men, and 1 missing information, Mage = 22.42, SD = 3.45) read a description of two lines of products. Ben was presented as a healthy, reduced sugar and fat, but tasty choice and JimJam as a rich, mouth-watering, and tasty option. Participants completed a first IAT and then the SR task. For half of the sample, it consisted of pairing Ben snacks with

the self, whereas for the other half, it consisted of pairing JimJam snacks with the self. Then, participants completed a memory test and, subsequently, a second IAT. For the choice of targets, in a separate pilot study, 31 individuals (17 women, 14 men, Mage = 33.30, SD = 10.59) rated eight fictitious logos and brand names for each category healthy and unhealthy on a 21-point scale from −10 (do not like it at all) to +10 (like it very much). The choice was oriented toward two logos/brand names that were equally neutral, Ben for Healthy bars and JimJam for Unhealthy bars (M = −0.55, SD = 3.38 and M = −0.26, SD = 4.80), $t(30)$ = −0.30, p = 767. We created pictures of different snack bars for the different tasks' stimuli for each of the brands.

An ANCOVA with SR Condition (Ben + Self vs. JimJam + Self) as a fixed factor, IAT score before the manipulation as a covariate, and IAT score after the manipulation as criterion tested the effect of SR on the IAT scores. We found a main effect of the first IAT score, $F(1,107)$ = 57.69, p < 0.001, η_p^2 = 0.35. More central to our concerns, there was a main effect of the SR condition, $F(1,107)$ = 8.54, p = 0.004, η_p^2 = 0.08. Before manipulation, there was no effect of the Experimental condition, $t(108)$ = 0.20, p = 0.844, whereas after the manipulation, there was a significant effect, $t(108)$ = 2.21, p = 0.029, $Cohen's\ d$ = 0.43. In line with our expectations, after completing the SR task, participants in the Ben + Self condition showed a higher IAT score for Ben relative to JimJam than the participants in the JimJam + Self condition.

Experiment 1

The main aim of the study was to assess both the immediate and the lasting effect of the SR task on the IAT. We designed a two-sessions study. In session 1, participants completed an SR task targeting healthy and unhealthy food brands. Three IATs were administered: one at the very beginning of the study (before the SR task) as a baseline measure of automatic evaluations of the two brands, the second right after the SR task, and the third one week later. We also measured changes in self-reported evaluations towards the two brands. Moreover, in session 1, we tested whether relating a brand with the self also produced extra-evaluative effects. In fact, self-referent information gains advantages in terms of memory and attention (e.g., Cunningham et al., 2008) and become more accessible. Because more accessible concepts or stimuli have higher chances of influencing behaviour (Eitam & Higgins, 2010), we tested whether pairing one food brand with the self could increase its response facilitation in a sequential priming task.

Method

We report all manipulations, measures, and exclusions in this study. The data and analysis code for both the preliminary and the main study are available on the Open Science Framework website, osf.io/4fdwk/.

Sample size determination

We did not determine the sample size based on a formal power analysis before the study. However, we chose a sufficiently large sample to detect an effect size in the region

of the one obtained in the preliminary study. With 145 participants, assuming a power of 0.80 (one-tail, a = 0.05), the study could detect an effect size of $d = 0.41$.

Participants and Procedure

One hundred and forty-five individuals (113 women, Mage = 23.50, SD = 3.64) took part in a two-session study after signing an informed consent form. In the first session, after reading the description of the two lines of products, participants carried out a first IAT and then the SR task followed by a memory test. Half of them paired Ben snack bars brands with the self, and the other half paired JimJam snack bars brands with the self. Next, participants underwent a sequential priming task, a second IAT, an explicit evaluation of both brands, and a questionnaire that assessed their self-identification as healthy eaters.[1] After one week, participants reread the brands' description and completed a memory test referring to the SR task completed in the first session, an IAT, an explicit evaluation of both brands. All measures were administered using Inquisit 4.0.5.0. The university ethics committee approved the study. The target stimuli used were identical to those used in the preliminary study.

Materials

Self-referencing task

First, participants completed two blocks of 40 trials in which they categorized Ben (JimJam) pictures and words relating to self (self, me, my, mine, I) to one key on a response box (e.g., 'blue') and JimJam (Ben) pictures and words related to others (they, them, their, his, her) to another response key (e.g., 'yellow'). The brand logo, the name of the brand in white lowercase and uppercase on a black square background, and two different flavors snack bars constituted the five pictures for each brand. Participants then repeated the two blocks of 40 trials switching the keys for the categories, that is, JimJam (Ben) pictures, and other related words were assigned to the 'yellow' key and Ben (JimJam) pictures and self-related words to the 'blue' key. The order in which participants completed these two sets of blocks was counterbalanced. In the case of incorrect classification, a red-X appeared on screen and remained until correction. The inter-trial interval was 400 ms. If participants had a percentage of errors above 15% in the last block of 40 trials, participants completed two additional blocks of 20 trials (one block for each key assignment).

Intersecting Regularities memory test

Participants indicated their recollection of the intersecting regularities between the self and the brand by responding to the following question: 'The task you have just completed consisted of classifying with the same key, words related to the self and pictures related to one brand. Do you remember which brand?' Participants indicated one of two brands or the option 'I don't remember'. Participants had a correct memory (correct response) or not (no recollection or incorrect response).

Sequential priming task

The task was designed to produce response competition, a mechanism that results in fast responses and high accuracy to the target stimulus if the prime has already activated the same response that is also required for the target stimulus (see Wentura & Rothermund, 2014). Participants classified words and pictures presented individually and in a random order in the middle of the screen using two keys of a response box (i.e., 'yellow' and 'blue'). They had to classify the words and pictures of the Ben versus JimJam snack. Before each target, a prime was presented. Like in a typical sequential priming procedure, a trial started with the presentation of a fixation cross for 500 ms, then a prime for 250 ms, followed by a blank screen for 50 ms, and finally, the target that stayed on the screen until the participant's response. The inter-trial interval was of 1500 ms. There were three categories of primes (Self vs. Others vs. Neutral) with five stimuli for each. The targets were the names of brands (Ben vs. JimJam) with ten stimuli for each. Each target was presented three times and each prime four times, leading to 60 trials by block with two consecutive blocks. Half participants used the left key to indicate Ben, and the other half used the right key. Before the two blocks, participants completed a block of five practice trials without primes and using targets not shown in the test blocks. Error feedback was given with a red X in the middle of the screen for 200 ms in the case of an incorrect response. Each test block included two dummy trials with neutral primes and targets presented in practice.

IAT

The three IATs (baseline, right after the SR task, and one week after the SR task) had the same structure. Participants classified words and pictures individually presented in a random order in the middle of the screen, using two keys (i.e., 'blue' and 'yellow'). The target concept was Ben and its contrast JimJam, whereas the attribute categories were Positive and Negative. We used five words (positive, joy, happy, paradise, and nice; negative, hell, ugly, sad, and pain) and five pictures (the uppercase and lowercase names of the brands, and three stimuli different from the ones used for the SR task) for each attribute and target category, respectively. The order of the two critical blocks was counterbalanced between participants, with half of the participants having the combination Ben and Positive being presented first and the other half having first the combination JimJam and Positive. All practice blocks consisted of 20 trials, and each critical block consisted of 81 trials (80 + 1 initial dummy trial). A red X appeared in the middle of the screen for 200 ms in case of an incorrect response, without requiring correction (no built-in penalty), the inter-trial interval was 500 ms, and the category labels stayed on the upper part of the screen throughout the task.

Self-reported ratings of the logos/brands

There were two forms of self-report: One rating for each type of bar and one relative rating. First, participants rated the type of snack bars separately (e.g., 'For me the snack

bars Ben are:') with four pairs of adjectives assessing more affective and more utilitarian aspects (useless/useful, disgusting/appetizing, repelling/attractive, unhealthy/healthy) on 7-point scales. The order of presentation of the two brands in the rating phase was matched with the order in which each brand was paired with positive stimuli in the first critical block of the IAT (i.e., 'Ben' first when first paired with positive words in the IAT vs. 'JimJam' first when first paired with positive in the IAT). Then, participants rated one snack bar brand relative to the other on four dimensions (interesting, beneficial, inviting, and appealing) on 7-point scales.

Results

Preliminary analyses and data preparation

The main exclusion criterion was the percentage of errors above 25% in any classification task (Sequential Priming and IATs). None of the participants fell into this category. Fifteen participants did not attend the second session, and their data were, therefore, excluded from the analyses (N = 130). One hundred and twenty-four participants (95.4%) remembered the intersecting regularities correctly, and seven (4.6%) did not.

The mean error in the sequential priming task was very low (4.1%), and therefore, we did not analyze it further. Reaction times with two standard deviations above the mean were discarded before computing the average reaction times of correct responses and subsequently log-transformed for relevant combinations of prime and target (though for ease of interpretation, we reported raw reaction times difference scores in **Table 1**). We calculated three non-redundant indexes based on these combinations such that a positive value indicated faster responses for the brand Ben compared to the brand JimJam. The score 'self-prime facilitation' indicated whether the targeted brand (i.e., paired with the self) was categorized

faster (i.e., leads to shorter reaction times) compared to the other brand when primed with the self. The score 'other-prime facilitation' indicated whether the targeted brand produced faster responses than the other brand when primed with the others. The score 'neutral-prime facilitation' indicated whether the targeted brand was categorized faster than the other brand when primed with neutral words. The 'neutral-prime facilitation' score does not consider the two focal primes (i.e., self and others). However, it is still an indicator of conditional (prime-based) response facilitation because it reflects the ease of categorizing the two brands when the prime is neutral. To report potential unconditional effects of the SR, we also computed a difference score across primes for each of the two experimental conditions (**Table 1**, 'unconditional facilitation').

For all three IATs, we calculated the D6 score (Greenwald, Nosek, & Banaji, 2003; α = 0.86, α = 0.86, and α = 0.83, respectively). For the explicit attitude, we performed a Principal Component Analysis for Time 1 and Time 2. We considered eight attitude scores. Four were derived from the difference in the scores obtained by each brand on the four semantic differentials, whilst the other four were the relative evaluations. For Time 1, we extracted two uncorrelated factors (r = 0.18) explaining 64.98% of variance. After a Varimax rotation, the first factor explained 44.01% of the variance and included the five items related to appetizing/hedonic aspects. The second factor explained 20.96% of the variance and included the three items related to utilitarian aspects. For Time 2, we also extracted two uncorrelated factors (r = 0.15) explaining 70.19% of variance. After a Varimax rotation, the first factor explained 45.68% of the variance and included the five items related to appetizing/hedonic aspects. The second factor explained 24.51% of the variance and included the three items related to utilitarian aspects. We saved the

Table 1: Descriptive statistics – Experiment 1.

	JimJam + Self (n = 67)		Ben + Self (n = 63)	
	M (SD)	95% CI	*M (SD)*	95% CI
IAT score before manipulation	0.27 (0.48)	[0.15; 0.39]	0.27 (0.50)	[0.15; 0.40]
IAT score after manipulation	0.23 (0.56)	[0.10; 0.36]	0.38 (0.52)	[0.25; 0.52]
IAT score – Time 2	0.23 (0.56)	[0.10; 0.36]	0.43 (0.49)	[0.30; 0.57]
Explicit Affective Attitude – Time 1	–0.03 (0.84)	[–0.27; 0.22]	0.03 (1.15)	[–0.22; 0.28]
Explicit Utilitarian Attitude – Time 1	–0.02 (0.99)	[–0.27; 0.22]	0.02 (1.02)	[–0.23; 0.28]
Explicit Affective Attitude – Time 2	–0.02 (0.84)	[–0.26; 0.23]	0.02 (1.15)	[–0.23; 0.27]
Explicit Utilitarian Attitude – Time 2	–0.06 (0.99)	[–0.31; 0.18]	0.07 (1.01)	[–0.18; 0.32]
Self-prime Facilitation	–2.71 (44.68)	[–14.39; 8.97]	18.98 (51.90)	[6.94; 31.03]
Other-prime Facilitation	–0.62 (41.44)	[–10.57; 9.33]	8.02 (40.89)	[–18.28; 2.25]
Neutral-prime Facilitation	1.81 (35.99)	[–6.97; 10.59]	3.04 (36.67)	[–6.02; 12.09]
Unconditional Facilitation	–0.51 (31.51)	[–7.92; 6.90]	4.67 (29.69)	[–2.97; 12.31]

Note: Sequential priming task scores are indicated in milliseconds.
For all criteria, all positive scores indicate a preference for Ben compared to JimJam.

factor scores for the analyses. For all scores, positive values indicated an advantage for the snack brand Ben.

Besides inspecting the SR task's impact on each variable separately, we tested the relationships between the manipulation and the outcome variables using Structural Equation Modeling (SEM) with the lavaan package (Rosseel, 2012). The added value of using SEM is twofold. First, as this work's main focus is on the effect of the SR on the IAT measured immediately after the manipulation and a week, we could estimate the effect's temporal stability by testing the variability in IAT performances over time. Second, SEM offers a broader and more comprehensive view of the effects of the SR manipulation on all outcomes by taking into account the relationships between them.

Effects on Response Facilitation
We first investigated the omnibus interaction effect between the type of prime (Self vs. Other vs. Neutral) and the SR manipulation (Ben+Self vs. JimJam+Self) with a mixed ANOVA. The effect of the type of prime was significant, $F(1.94, 248.62) = 3.99$, $p = 0.021$, $\eta_p^2 = 0.03$.[2] Post-hoc analyses indicated the Ben stimuli's advantage in terms of facilitation of response was stronger when the self was primed than when Others was used as prime ($p = 0.018$). The effect of SR manipulation was not significant, $F(1, 128) = 0.93$, $p = 0.338$. Moreover, there was a significant interaction effect between the type of prime and the SR condition, $F(1.94, 248.62) = 5.73$, $p = 0.004$, $\eta_p^2 = 0.04$. We further examined the effect of SR on the three indexes separately as they reflected distinct questions. The effect of SR was significant for the Self-prime Facilitation index, $F(1, 128) = 6.55$, $p = 0.012$, $\eta_p^2 = 0.05$, but was not significant for the Other-prime Facilitation index, $F(1, 128) = 1.05$, $p = 0.308$, nor for the Neutral-prime Facilitation index, $F(1, 128) = 0.04$, $p = 0.847$. The participants in the Ben+Self condition were faster to classify Ben stimuli when primed with the self, compared to the participants in the JimJam+Self condition.

Effects on IAT scores
We conducted a 2 (Time: Time 1 after manipulation vs. Time 2) × 2 (Experimental Condition: Ben+Self vs. JimJam+Self) mixed ANCOVA with the IAT taken before the SR manipulation as a covariate to control for baseline differences. There was no main effect of Time, $F(1, 127) = 0.81$, $p = 0.369$. The IAT administered before the manipulation was a significant predictor, $F(1, 127) = 81.86$, $p < 0.001$, $\eta_p^2 = 0.39$. More central to our concerns, the effect of Experimental Condition was significant, $F(1, 127) = 7.02$, $p = 0.009$, $\eta_p^2 = 0.05$, indicating the lasting effect of the SR manipulation. Participants showed higher IAT scores for the brand paired with the self in the SR task, demonstrating the temporal stability of the SR effect. Moreover, the interaction terms Time x before IAT and Time x Experimental Condition were not significant, $F(1, 127) = 0.43$, $p = 0.512$ and $F(1, 127) = 0.35$, $p = 0.554$, respectively. The latter finding is of particular interest here, as it shows that the impact of the SR task did not differ when comparing immediate and lasting effects on the IAT.

Effects on self-reported ratings
We conducted a 2 (Time: Rating 1 vs. Rating 2) × 2 (Experimental Condition: Ben+Self vs. JimJam+Self) mixed ANOVA to test the temporal effect of SR on the explicit affective and utilitarian attitude scores separately. There was no effect of Time, no effect of Experimental Condition, and no significant interaction for either of the two attitude scores (all p's > 0.579). Although no significant results emerged from the analyses, the means at Time 2 for both the affective and utilitarian components of the explicit attitude were in the expected direction with a preference for the brand paired with the self (see **Table 1**).

Network of relationships between variables
Correlations among all variables are reported in **Table 2**. The self-reported affective ratings and IAT scores at Time 1 and self-prime facilitation correlated significantly, although the SR manipulation did not significantly affect the self-reported ratings. For the sake of parsimony, SEM focused only on the IAT, self-reported affective ratings, and self-prime facilitation variables, because they showed a pattern of relations with the SR manipulation or among each other (see **Figure 1**). The overall fit was acceptable, with a slightly significant chi-square index, $\chi^2 (42) = 62.56$, $p = 0.021$, but a high CFI (0.96) and a not-significant RMSEA (0.061, $p = 0.264$). The three latent variables significantly correlated at time 1, whereas the SR manipulation had a significant effect on IAT scores and the self-prime facilitation. The most remarkable result is the very high stability of the IAT after the SR manipulation (0.91). Therefore, the SR effect was demonstrated to be stable both at the mean and at the correlational level.
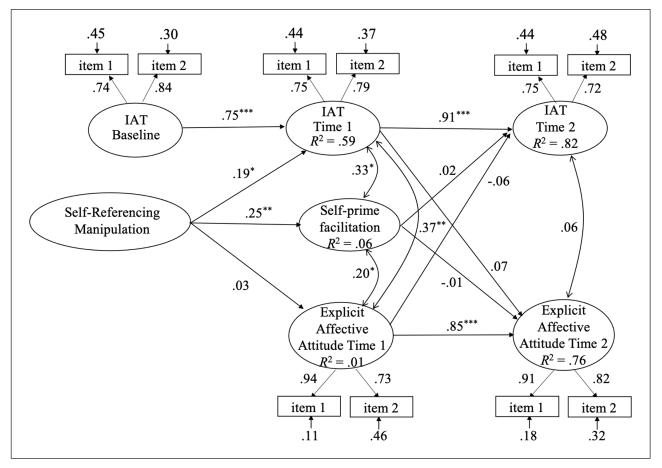
Discussion
In this study, we replicated previous findings on the SR task and showed that pairing the self with a neutral food brand led to more positive IAT scores of this brand than the one paired with others. More importantly, such an effect lasted one week after manipulation, proving that the SR represents a powerful tool to influence evaluations measured via behavioural responses produced under specific automaticity conditions. Conversely, there was no effect on self-report evaluations, neither when measured right after the SR manipulation nor after one week. Moreover, results from a sequential priming task showed when the self was used as prime, the brand prior paired with the self was processed quicker than the brand paired with others. Last but not least, the SEM model showed that the effect of the SR task on the IAT was stable at the mean and the correlational levels.

Based on these findings, we conducted a second pre-registered experiment. The primary aim of Experiment 2 was to provide formal replication for the lasting effect observed on the IAT. Moreover, this study tested whether the effect generalizes over two alternative classes of food stimuli. Last, we added ecological and construct validity by testing whether the SR effect, measured after training and one week later, was reflected by performances in a food choice task.

Table 2: Correlations – Experiment 1.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Experimental Condition (SR)	1													
2. IAT score before manipulation	0.01	1												
3. IAT score after manipulation	0.15	0.58**	1											
4. IAT score Time 2	0.19*	0.54**	0.66**	1										
5. IAT2_T1-IAT1_T1	0.16	−0.36**	0.55**	0.20*	1									
6. IAT1_T2-IAT1_T1	0.20*	−0.40**	0.14	0.56**	0.57**	1								
7. Explicit Affective Attitude Time 1	0.03	0.20*	0.30**	0.22*	0.14	0.04	1							
8. Explicit Utilitarian Attitude Time 1	0.02	0.13	0.11	0.15	0.00	0.04	0.00	1						
9. Explicit Affective Attitude Time 2	0.02	0.15	0.31**	0.23**	0.20*	0.11	0.74**	0.04	1					
10. Explicit Utilitarian Attitude Time 2	0.06	0.11	0.15	0.06	0.06	−0.04	0.05	0.64**	0.00	1				
11. Self-prime Facilitation	0.22*	0.11	0.24**	0.20*	0.16	0.11	0.19*	0.09	0.17	0.11	1			
12. Other- prime Facilitation	−0.09	0.08	−0.07	0.03	−0.15	−0.04	0.00	0.04	−0.02	0.01	0.22*	1		
13. Neutral- prime Facilitation	0.02	0.05	0.16	0.14	0.12	0.10	0.06	0.01	−0.06	0.08	0.34**	0.31**	1	
14. Unconditional accessibility	0.08	0.11	0.16	0.18*	0.07	0.08	0.13	0.07	0.06	0.10	0.77**	0.69**	0.72**	1

Note: ** $p < 0.01$. * $p < 0.05$. + $p < 0.10$.

IAT2T1_IAT1T1 = difference between the first IAT at time 1 and the IAT at time 2.

IAT1T2_IAT1T1 = difference between IAT before and after the manipulation at time 1, IAT1T2_IAT1T1 = difference between the manipulation at time 1.

Figure 1: The estimated model of relations between self-referencing manipulation, self-prime facilitation, IAT and affective ratings at Time 1 and Time 2.

Experiment 2

The results of the first experiment were promising. Therefore, we conducted a pre-registered conceptual replication, introducing some relevant variations in the experimental procedure that allowed us to extend the significance of our findings. In particular, Experiment 2 differed from Experiment 1 in three main aspects. First, two new food brands were used to test whether previous findings generalized over alternative classes of food items. In fact, whilst in Experiment 1, we limited our investigation to healthy versus unhealthy snacks, in Experiment 2 we focused on food items in general. Second, after the IAT and the self-report ratings in both sessions, we included a food choice task adapted from Schakel and colleagues (2018). Such a measure added ecological validity to the previous findings. It tested whether evaluative changes observed on behavioural responses to the IAT resulted in changes in behavioural choices made under different environmental conditions. Moreover, with the inclusion of a behaviour measure, we tested the construct validity of the SR effect on the IAT. To do so, we designed a behavioural measure that required participants to make a food choice under the same conditions of automaticity (i.e., time pressure) that characterize behavioural responses performed in the IAT. Third, the experiment's design was streamlined to focus on the key effects and allow for its deployment online. Therefore, neither the baseline IAT nor the sequential priming task were administered in Experiment 2. In line

with the results from Experiment 1, we expected the SR manipulation to affect IAT performances after the SR task and one week later. Two alternative hypotheses were plausible for the effect of the SR task on self-report ratings. On the one hand, results might be in line with Experiment 1, with no SR effect at Time 1 or Time 2. On the other hand, the change in the type of target (general food instead of healthy food) could increase the chance to observe a significant SR effect, at least at time 1. Finally, we expected the SR effect to affect food choice behaviour. However, because we predicted the latter effect to be lower in magnitude than those on IAT performances (both at Time 1 and Time 2) and because the study was powered on the IAT effect, testing this hypothesis will be mainly exploratory. The design as well as the analyses and sampling plan were preregistered via OSF (https://osf.io/4duq8) and reviewed by the International Review of Social Psychology before conducting the study. There were no deviations to report. All data and analysis code are available in the OSF repository at https://osf.io/2ue8n.

Method
Sample size determination
This study aimed to replicate the lasting impact of the SR task on the IAT score (main effect of the SR manipulation in a 2 × 2 mixed ANOVA). We thus estimated the required sample size for our study based on the effect size of the SR manipulation and the correlation between IAT measures

observed in Experiment 1. Using GPower3, given effect size $d = 0.381$, a correlation between measures of $r = 0.66$, and assuming a power of .90 (two-tailed, a = .05), the estimated sample size was $N = 244$. Considering an estimation of approximately 10% of participants to be excluded based on our exclusion criteria (errors above 25% in any of the IATs and failure in completing both sessions), we planned to stop data collection when reaching a sample of 270 participants.

Participants and Procedure
A total of 300 participants took part in both sessions. Thirty-two participants who failed to complete either session were excluded from the final sample. This left us with 268 (118 females, Mage = 27.48, SDage = 8.27) participants. Unlike Experiment 1, the study was conducted online via Prolific Academic. In the first session, participants first provided their informed consent. Then they were presented with two fictitious food brands, Lestea and Sabea, taken from previous SR studies (Mattavelli, Richetin, Perugini, unpublished). Participants then completed the SR task and the memory test. Next, participants underwent an IAT, a food choice task, and an explicit evaluation of both brands. After one week, participants were presented with the same two brands, completed the memory test referring to the SR task completed in the first session, an IAT, the food choice task, and the explicit evaluation measure. All measures were administered using Inquisit 6. The University ethics committee approved the study.

Materials
The IATs and the SR task were identical to those administered in Experiment 1, except for the type of target stimuli. We replaced the healthy and unhealthy food brands used in Experiment 1 (i.e., Ben vs. JimJam) with two fictitious generic brands (i.e., Lestea and Sabea). Also identical to that administered in Experiment 1 was the memory question, with the name of the new brands presented as response options. However, we modified the self-report ratings in two ways. First, only two semantic differentials targeting each brand were administered, with no relative measure. Second, we used only affective adjectives (bad/good, unlikable/likable, repelling/attractive, unpleasant/pleasant). Finally, we added a food choice task.

Food choice task
A computerized food choice task was administered at the end of both sessions. The task is adapted from Schakel and colleagues (2018). Participants were presented with seven food product pairs, each containing one item of the brand priorly related to the self versus one item of the brand related to others. For each pair, participants were asked to indicate which of the two food products they would have chosen at that moment. Because we wanted this behavioural choice to be made under specific conditions of automaticity (i.e., time pressure), participants were asked to go with their gut feelings and select their preferred option from each pair as fast as they could. The final food choice scores were determined by summing the food choices for the brand Lestea, with scores ranging from 0 to 7.

Results
Data preparation
We excluded the data from participants with an error percentage above 25% in any of the IATs ($N = 21$) and from those who failed in completing both sessions ($N = 4$). Applying these screening criteria led to a final sample of 243 participants.[3] At time 1, 188 participants (77%) showed correct memory of the intersecting regularities acquired in the SR task, and 158 participants (65%) did so at time 2.

For the two IATs, we calculated the D6 score (Greenwald et al., 2003). Both the IATs showed high reliability (IAT1: α = 0.87; IAT 2: α = 0.90). For the self-report ratings, for both Time 1 (α = 0.82) and Time 2 (α = 0.88), we computed four differential scores, one for each item of the semantic differentials used for either brand. Then, we averaged the four scores into a unique score. For all scores, positive values indicated an advantage for the snack brand Lestea.

SR effect on each outcome variable
We conducted a series of 2 (Time: Time 1 vs. Time 2) × 2 (SR manipulation: Lestea+Self vs. Sabea+Self) mixed ANOVAs on IAT scores, food choices, and self-report evaluations (see **Table 3** for descriptives).

Effects on IAT scores (pre-registered)
There was no main effect of Time, $F(1, 241) = 3.46$, $p = 0.064$. The effect of SR manipulation was significant, $F(1, 241) = 27.00$, $p < 0.001$, $\eta_p^2 = 0.10$. Higher IAT scores emerged for

Table 3: Descriptive statistics – Experiment 2.

	Lestea + Self (n = 119)		Sabea + Self (n = 124)	
	M (SD)	95% CI	M (SD)	95% CI
IAT score – Time 1	–0.08 (0.46)	[–0.17; 0.00]	–0.40 (0.46)	[–0.48; –0.32]
IAT score – Time 2	–0.16 (0.58)	[–0.26; –0.06]	–0.43 (0.49)	[–0.53; –0.34]
Explicit Attitude – Time 1	–0.23 (0.99)	[–0.41; –0.05]	–0.29 (1.03)	[–0.46; –0.11]
Explicit Attitude – Time 1	–0.19 (1.07)	[–0.40; 0.00]	–0.31 (1.09)	[–0.50; –0.11]
Food choice – Time 1	3.44 (1.48)	[3.17; 3.71]	3.11 (1.51)	[2.85; 3.38]
Food choice – Time 2	3.45 (1.52)	[3.18; 3.73]	3.06 (1.50)	[2.79; 3.32]

Note: For all criteria, all positive scores indicate a preference for Lestea compared to Sabea.

the brand paired with the self in the SR task. Also, the interaction term Time × SR manipulation was not significant, $F(1, 241) = 0.50$, $p = 0.482$. This non-significant interaction was consistent with the results observed in Experiment 1. It showed that the impact of the SR task did not differ when comparing immediate and lasting effects on the IAT.

Effects on self-reported ratings (pre-registered)
There was no main effect of Time, $F(1, 239) = 0.03$, $p = 0.859$. Neither the effect of SR manipulation nor the Time × SR manipulation interaction term were significant, $F(1, 239) = 0.54$, $p = 0.465$ and $F(1, 239) = 0.16$, $p = 0.692$, respectively. Thus, we replicated the null effect of the SR manipulation on self-reported evaluation of the two brands (even when generic food brands were used as target stimuli).

Effects on food choice (pre-registered)
We found no main effect of Time, $F(1, 241) = 0.04$, $p = 0.837$. The effect of SR manipulation was significant, $F(1, 241) = 4.68$, $p = 0.032$, $\eta_p^2 = 0.02$. This result indicates that completing the SR task produced automatic food choices in favor of the brand categorized through the same action as the self in the SR task. The interaction term Time x SR manipulation was not significant, $F(1, 241) = 0.15$, $p = 0.704$. Thus, similar to what we found on the IAT scores, the impact of the SR task on automatic food choices showed persistence after one week.

Pattern of relationships between variables (pre-registered)
The correlations between the outcome variables are reported in **Table 4**. We examined the associations between the variables with SEM (**Figure 2**). We tested a full model with cross-lagged paths included in it (dashed arrows). The overall fit was good, with a non-significant chi-square index, $\chi^2(31) = 20.26$, $p = 0.930$, a high CFI (1.00) and a non-significant RMSEA (0.00, $p = 0.999$).[4] The SR manipulation significantly affected IAT scores at time 1, but not the other two outcome variables. At time 1, the IAT score predicted the food choice task, and the same pattern replicated at time 2.

The role of IR memory on the lasting SR effect (not pre-registered)
In an exploratory way, we conducted two separate sets of moderation analyses. We tested whether the participants' memory of the intersecting regularities (i.e., correct memory vs. incorrect/no memory) moderated the immediate and lasting SR task effects on the IAT and the food choice task.

For the immediate effects, SR condition was the main predictor, IR memory at time 1 was the moderator and the IAT and the food choice task at time 1 were the dependent variables. When the IAT was considered as criterion ($R^2 = 0.11$), we found a significant SR effect, $b = 0.32$, $p < 0.001$, 95% CI [0.21; 0.45], a non significant main effect of IR memory $b = 0.02$, $p = 0.692$, 95% CI [−0.10; 0.14], and a non-significant interaction term, $b = 0.07$, $p = 0.287$, 95% CI [−0.06; 0.19]. The same analysis on food-choice ($R^2 = 0.04$), showed a marginal SR effect, $b = 0.11$, $p = 0.082$, 95% CI [−0.01; 0.24], a non significant main effect of IR memory, $b = 0.06$, $p = 0.331$, 95% CI [−0.06; 0.19], but a significant interaction term, $b = 0.15$, $p = 0.021$, 95% CI [0.02; 0.26]. Decomposing this interaction showed that the SR effect was significant for participants with correct memory, $b = 0.19$, $p = 0.009$, 95% CI [0.05; 0.33], and not significant and descriptively in the opposite direction for those with either incorrect memory or no memory of the intersecting regularities, $b = −0.16$, $p = 0.227$, 95% CI [−0.42; 0.10].

For the lasting effects, SR condition was the main predictor, IR memory at time 2 was the moderator and the IAT and the food choice task at time 2 were the dependent variables. On the IAT ($R^2 = 0.09$), we found a significant SR effect, $b = 0.25$, $p < 0.001$, 95% CI [0.12; 0.38], a non significant main effect of IR memory $b = 0.05$, $p = 0.403$, 95% CI [−0.07; 0.18], and a significant interaction term, $b = 0.17$, $p = 0.007$, 95% CI [0.05; 0.29]. Decomposing this interaction showed that the SR effect was significant for participants with correct memory, $b = 0.38$, $p < 0.001$, 95% CI [0.23; 0.53], and null for those with either incorrect memory or no memory of the intersecting regularities, $b = 0.02$, $p = 0.844$, 95% CI [−0.19; 0.23]. Then we repeated the same analysis with food-choice as criterion ($R^2 = 0.07$). We found a significant SR effect, $b = 0.14$, $p = 0.033$, 95% CI [0.01; 0.27], a non significant main effect of IR memory $b = 0.02$, $p = 0.699$, 95% CI [−0.10; 0.16], and a significant interaction term, $b = 0.22$, $p < 0.001$, 95% CI [0.09; 0.34]. The SR effect was significant for participants with correct memory, $b = 0.30$, $p < 0.001$, 95% CI [0.14; 0.45], and not significant and descriptively in the opposite direction for those with either incorrect memory or no memory of the intersecting regularities, $b = −0.16$, $p = 0.137$, 95% CI [−0.37; 0.05].

Table 4: Correlations – Experiment 2.

	1	2	3	4	5	6	7
1. Experimental Condition (SR)	1						
2. IAT score – Time 1	0.33**	1					
3. IAT score – Time 2	0.25**	0.58**	1				
4. Explicit Attitude – Time 1	0.03	0.26**	0.34**	1			
5. Explicit Attitude – Time 1	0.05	0.27**	0.33**	0.53**	1		
6. Food choice – Time 1	0.11	0.33**	0.33**	0.49**	0.36**	1	
7. Food choice – Time 2	0.13*	0.28**	0.30**	0.39**	0.49**	0.51**	1

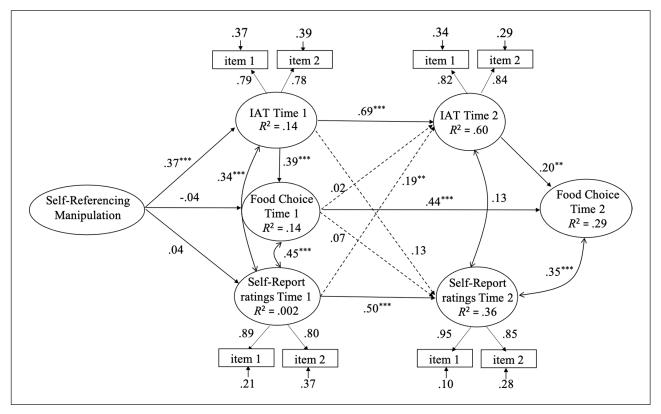*Note:* ** $p < 0.01$. * $p < 0.05$.

Figure 2: Full model for Experiment 2.

Discussion

Experiment 2 replicated the main findings observed in Experiment 1. Namely, we confirmed the lasting effect of the SR task on automatic evaluations revealed by the IAT. The non-significant effect on self-reported evaluations and the significant (lasting) effect on the food choice task confirmed that the experimental paradigm might be especially suitable to change instances of behavior expressed under conditions of automaticity. The temporal stability of the IAT was also good (0.69, $p < 0.001$) in the full SEM and in a separate (not-preregistered) measurement model including only the IAT scores at time 1 to predict IAT scores at time 2 (0.74, $p < 0.001$). Finally, the moderating role of IR memory is of theoretical relevance. It shows that the observed effect cannot be reduced to a mere training-testing effect, according to which performing the SR somehow prepares individuals to perform better in one critical block of the IAT. Instead, we showed that the effect observed after one week emerged only for participants who correctly remembered the intersecting regularities learned via the SR task. Thus, it was not the performance but rather the rule the individual extracted from that performance that determined automatic evaluations and choices.

General Discussion

The IAT has been used extensively to capture automatic evaluations within the context of food-related preferences and behaviors (e.g., Conner et al., 2007; Richetin et al., 2007). Whereas associative learning manipulations have proven reliable in producing immediate changes in automatic evaluations reflected by the IAT (Hollands et al., 2011), such changes typically fail to last over time

(Gawronski et al., 2017). Here we investigated the SR task's impact in producing immediate and lasting change (i.e., after one week) on the IAT. A first study offered initial evidence that categorizing one food brand with the same action as the self (as opposed to another target brand categorized with others) produced immediate and lasting changes on the IAT, but not on self-reported evaluations. To test the robustness of such findings, we conducted a well-powered and pre-registered second study that mirrored the first experiment's procedure, targeting two alternative food brands. Moreover, in this second study, we introduced a measure of automatic food choice. The main pattern of results was replicated. The SR task's effect emerged on the IATs immediately after and one week after, whereas no effect was found on self-reported evaluations. Interestingly, we found that the immediate and the lasting effects of the SR task generalized to the automatic food choice measure.

Taken together, these results offered robust evidence in favour of the SR task as a powerful tool to affect automatic evaluations over time. Recent research has shown that training participants to respond to a target food versus to inhibit a response to an alternative food creates preferences for the target food even one week after the intervention (Chen et al., 2019). The SR goes beyond the action versus inaction effect, proving the potential of learning via intersecting regularities. In particular, performing the same action (e.g., pressing one key response to categorize stimuli) in response to alternative classes of stimuli acquires distinct meaning depending on the type of intersections created by the action itself. When the action produced in response to a neutral food brand is identical to that produced in response to self-stimuli, that

action puts the two classes in connection. In so doing, it creates the conditions for the transfer of valence from the source to the target (see De Houwer et al., 2019 and Hughes et al., 2020, for articulated reasoning on how people assume similarity between stimuli based on their shared features).

The lasting effects of the SR task are even more remarkable when considering that other associative manipulations failed to produce similar effects on the IAT (Lai et al., 2016). Then, what makes the SR task stronger than other associative paradigms in changing IAT scores? We propose that a key role is played by the environmental conditions that form stimuli relationship (i.e., acquisition) and those that affect performances (i.e., measurement). In the SR task, participants are trained to respond with a common key to categorize one stimulus and the self. Thus, participants learn that two stimuli are related because they share a common response (i.e., intersecting regularities). Just like intersecting regularities set the occasion for learning in the SR task (i.e., if I respond with the same action to Stimulus A and the Self, then I learn that the two classes of stimuli share something), they are also at play in the IAT performance (i.e., If Stimulus A shares something with a positive class of stimuli, then it should be easier for me to perform a task where Stimulus A and positive [vs. negative] stimuli go with the same key). Therefore, we speculate that this matching in the learning principle both the acquisition and the testing are based on, eases the lasting effect of the former on the latter. This reasoning paves the way for future investigations. For instance, researchers might want to compare the impact of associative manipulations based on distinct learning pathways on the IAT directly. Or, applying the same logic at the measurement level, researchers might compare the lasting effect of the SR task on the IAT and an alternative indirect measure.

Claiming that the features shared by the SR and the IAT might account for the observed effects could cast doubts on what drives the effect of the manipulation on the outcome measure. One might argue that the SR task serves as a practice for one IAT block mapping, leading to shorter response latencies on the measurement level (i.e., IAT). However, if this was the case, one should expect a stronger SR effect on the IAT measured immediately after the learning paradigm, when response mapping has just been acquired, than one week later. Our results from both studies clearly showed that this is not the case. In both studies, we found no significant difference between the immediate and lasting effects on the IAT scores. However, this does not rule out potential issues related to the construct validity of SR task. Namely, performing the SR task might produce both immediate and lasting non-evaluative effects on the IAT. We anticipated this issue in designing Experiment 2. We decided to introduce a choice task that could provide an additional measure of behavior performed towards the two food brands under automaticity conditions (i.e., a binary choice made under time constraints). Notably, we observed SR effects in this food choice task after one week. The lasting effect on the food choice measure and its significant correlation with the IAT scores supports the validity of the SR manipulation

in producing evaluative changes capable of influencing automatic decisions and behaviors.

Another important finding replicated across the two studies is the non-significant effect of the SR manipulation on self-reported evaluative ratings measured at time 1 and time 2. Participants did not exhibit an overt preference for one brand over the other. Together with the significant findings observed on automatic outcome measures (IAT and food choice task), such a pattern of results has implications for the mental mechanisms that might mediate the impact of the SR manipulation. Based on a dual-process account of attitude change (Gawronski & Bodenhausen, 2014), associative interventions, like the SR task, are more likely to affect evaluations captured by indirect, as opposed to direct, measures. This assumption is largely based on the idea that this type of manipulation can alter mental associations between concepts without requiring any form of propositional reasoning. Alternatively, a propositional account (De Houwer, 2009, 2014, 2018) proposes that evaluations reflected by indirect measures like the IAT are also mediated by propositional reasoning. A thorough analysis of the findings from Experiment 2 seems to support this latter account. We showed that the SR effects on the IAT and the food choice task measured after one week were qualified by participant's ability to recall which brand was categorized through the same action as self one week before. When this was not the case, no SR effect was detected on either automatic measure. In essence, merely performing the SR task was the necessary, but not sufficient, condition to change automatic performances. What was necessary for such changes was learning (and recalling) the intersection of regularities that arose from performing the SR task.

Finally, this research introduced the power of the self in lasting evaluative changes. Beyond the intersecting regularities principle, the SR is different from other associative paradigms because it uses the self as a positive evaluative source. Abundant empirical evidence has shown that the self is peculiar when it comes to producing evaluative change. Self-relevant information gains advantages within other important cognitive domains, such as memory (Cunningham et al., 2008). We propose that such benefits can make the self an ideal source for generating evaluative changes that last over time. Future research should test more directly the key role of this specific feature of the SR task in producing lasting changes in automatic evaluations towards food stimuli. For instance, as past studies demonstrated the role of individuals' self-esteem in moderating the SR effect (Mattavelli, Richetin, & Perugini, 2019; Prestwich et al., 2010), future studies could investigate the role played by this variable when the SR effect is tested over time.

In summary, we showed that once food items are related to the self via a common response, automatic positive evaluations towards food stimuli reflected by the IAT follow and last at least one week. These effects were obtained using the SR task, an associative paradigm that relies on the assumption that stimuli that share a common feature become related (i.e., intersecting regularities). Combining the power of the self and learning via intersecting

regularities, the influence of the SR is remarkable. Even a seemingly meaningless common feature (i.e., the act of pressing a common response key) sufficed to relate food items to the self and ultimately resulted in automatic positive evaluations that persisted over time and predicted automatic food choice. The key role of intersecting regularities memory in qualifying these lasting effects suggests potential generalizability to a vast range of alternative self-food commonalities. Any feature shared by a desirable food and the self (e.g., a brand name) can potentially alter how individuals respond to the former under conditions of automaticity.

Notes

[1] Because the inclusion of such a measure in the analysis did not determine effects that were relevant for the current investigation, we decided neither to desribe it further nor to include it in the analyses presented here.

[2] The values of the degrees of freedom are based on Huynh-Feldt correction, which was chosen because the Mauchly sphericity test showed significant. The values without Huynh-Feldt correction are very similar and do not change any of the observed effect.

[3] For two participants who successfully completed both the sessions, no data for the time2 self-reported ratings were saved. This explains the reduced degrees of freedom for the analyses on self-report ratings.

[4] We also tested the same model without the cross-lagged effects, $\chi^2 (35) = 33.47$, $p = 0.542$. A direct comparison showed that the full model was significantly better, $\chi^2 (4) = 13.21$, $p = 0.01$.

Competing Interests
The authors have no competing interests to declare.

Author Contributions
Simone Mattavelli: investigation, data curation, formal analysis, methodology, writing original draft, revision of the manuscript.

Juliette Richetin: conceptualization, investigation, data curation, formal analysis, methodology, writing, revision of the manuscript.

Marco Perugini: conceptualization, investigation, methodology, writing, revision of the manuscript.

References

Chen, Z., Holland, R. W., Quandt, J., Dijksterhuis, A., & Veling, H. (2019). When mere action versus inaction leads to robust preference change. *Journal of Personality and Social Psychology, 117*(4), 721–740. DOI: https://doi.org/10.1037/pspa0000158

Conner, M. T., Perugini, M., O'Gorman, R., Ayres, K., & Prestwich, A. (2007). Relations between implicit and explicit measures of attitudes and measures of behavior: Evidence of moderation by individual difference variables. *Personality and Social Psychology Bulletin, 33*(12), 1727–1740. DOI: https://doi.org/10.1177/0146167207309194

Corneille, O., & Stahl, C. (2019). Associative attitude learning: A closer look at evidence and how it relates to attitude models. *Personality and Social Psychology Review, 23*(2), 161–189. DOI: https://doi.org/10.1177/1088868318763261

Cunningham, S. J., Turk, D. J., Macdonald, L. M., & Macrae, C. N. (2008). Yours or mine? Ownership and memory. *Consciousness and Cognition, 17*(1), 312–318. DOI: https://doi.org/10.1016/j.concog.2007.04.003

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology, 10*(2), 230–241. DOI: 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. DOI: https://doi.org/10.3758/LB.37.1.1

De Houwer, J. (2014a). A propositional model of implicit evaluation. *Social and Personality Psychology Compass, 8*(7), 342–353. DOI: https://doi.org/10.1111/spc3.12111

De Houwer, J. (2014b). Why a propositional single-process model of associative learning deserves to be defended. In J. W. Sherman, B. Gawronski & Y. Trope (Eds.), *Dual processes in social psychology*. Guilford; NY.

De Houwer, J. (2018). Propositional models of evaluative conditioning. *Social Psychological Bulletin, 13*(3), 1–21. DOI: https://doi.org/10.5964/spb.v13i3.28046

Demartini, E., De Marchi, E., Cavaliere, A., Mattavelli, S., Gaviglio, A., Banterle, A., ... & Perugini, M. (2019). Changing attitudes towards healthy food via self-association or nutritional information: What works best? *Appetite, 132*, 166–174. DOI: https://doi.org/10.1016/j.appet.2018.08.001

Eitam, B., & Higgins, E. T. (2010). Motivation in mental accessibility: Relevance of a representation (ROAR) as a new framework. *Social and Personality Psychology Compass, 4*(10), 951–967. DOI: https://doi.org/10.1111/j.1751-9004.2010.00309.x

Friese, M., Hofmann, W., & Wänke, M. (2008). When impulses take over: Moderated predictive validity of explicit and implicit attitude measures in predicting food choice and consumption behaviour. *British Journal of Social Psychology, 47*(3), 397–419. DOI: https://doi.org/10.1348/014466607X241540

Gawronski, B. (2019). Six lessons for a cogent science of implicit bias and its criticism. *Perspectives on Psychological Science, 14*(4), 574–595. DOI: https://doi.org/10.1177/1745691619826015

Gawronski, B., & Bodenhausen, G. V. (2014). Implicit and explicit evaluation: A brief review of the associative–propositional evaluation model. *Social and Personality Psychology Compass, 8*(8), 448–462. DOI: https://doi.org/10.1111/spc3.12124

Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of implicit and explicit

measures: A longitudinal analysis. *Personality and Social Psychology Bulletin, 43*(3), 300–312. DOI: https://doi.org/10.1177/0146167216684131

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. DOI: https://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. DOI: https://doi.org/10.1037/0022-3514.85.2.197

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

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. DOI: https://doi.org/10.1037/a0022261

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. DOI: https://doi.org/10.1037/xge0000100

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., ... & Simon, S. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General, 145*(8), 1001–1016. DOI: https://doi.org/10.1037/xge0000179

Maison, D., Greenwald, A. G., & Bruin, R. (2001). The Implicit Association Test as a measure of implicit consumer attitudes.

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, 51*(4), 511–518. DOI: https://doi.org/10.1007/s12160-016-9874-9

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. DOI: https://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. DOI: https://doi.org/10.1002/ejsp.2518

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). E-research: Ethics, security, design, and control in psychological research on the Internet. *Journal of Social Issues, 58*(1), 161–176. DOI: https://doi.org/10.1111/1540-4560.00254

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*, 142–156. DOI: https://doi.org/10.1086/662069

Prestwich, A., Hurling, R., & Baker, S. (2011). Implicit shopping: Attitudinal determinants of the purchasing of healthy and unhealthy foods. *Psychology & Health, 26*(7), 875–885. DOI: https://doi.org/10.1080/08870446.2010.509797

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. DOI: https://doi.org/10.1002/ejsp.610

Richetin, J., Perugini, M., Prestwich, A., & O'Gorman, R. (2007). The IAT as a predictor of food choice: The case of fruits versus snacks. *International Journal of Psychology, 42*(3), 166–173. DOI: https://doi.org/10.1080/00207590601067078

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of Statistical Software, 48*(2), 1–36. DOI: https://doi.org/10.18637/jss.v048.i02

Schakel, L., Veldhuijzen, D. S., Middendorp, H. V., Dessel, P. V., Houwer, J. D., Bidarra, R., & Evers, A. W. (2018). The effects of a gamified approach avoidance training and verbal suggestions on food outcomes. *PloS one, 13*(7), e0201309. DOI: https://doi.org/10.1371/journal.pone.0201309

Van Dessel, P., Cummins, J., Hughes, S., Kasran, S., Cathelyn, F., & Moran, T. (2020). Reflecting on twenty-five years of research using implicit measures: Recommendations for their future use. *Social Cognition.* DOI: https://doi.org/10.1521/soco.2020.38.supp.s223

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological bulletin, 132*(2), 249–268. DOI: https://doi.org/10.1037/0033-2909.132.2.249

Wentura, D., & Rothermund, K. (2014). Priming is not priming is not priming. *Social Cognition, 32*, 47–67. DOI: https://doi.org/10.1521/soco.2014.32.supp.47

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. DOI: https://doi.org/10.1111/j.1467-9280.2007.01928.x