University of Milan-Bicocca – Department of Psychology

Doctoral Program: Experimental Psychology, Linguistics and Cognitive Neuroscience

XIX Cycle

# Attitude formation and change via intersecting regularities: Peculiarities, boundaries, and practical applications of the Self-Referencing task.

PhD candidate:

**Simone Mattavelli**

Registration number:

**706202**

Supervisor:

**Prof. Marco Perugini**

Co-supervisor:

**Dr. Juliette Richetin**

Coordinator:

**Prof. Maria Teresa Guasti**

Academic year: 2016

# Table of Contents

# Figures

# Tables

# Abstract

The present work centers on associative attitude change via the recently discovered intersecting regularities (IR) mechanism as well as on the role of the self as an attitudinal source. We reviewed and tested the Self-Referencing (SR) task, an associative learning task where self-object links are operationalized through IR that occur at the level of the action required to categorize stimuli (i.e., pressing the same key for self-stimuli and one target versus pressing another key for other-stimuli and a contrast target). As a result, the target categorized through the same action as the self is better liked. The SR task is introduced by a single-lab meta-analysis that quantitatively summarizes its effectiveness in changing implicit and explicit attitudes (Chapter 1). After that is a series of fourteen empirical studies ($N = 1856$) articulated in five lines of research. We first focused on the role of the self in driving attitude change, by testing what properties possessed by self-related stimuli determine the SR effect (Chapter 2). Then we shifted to the IR mechanism and tested the possibility to exploit it to generate indirect SR effect by building complex chains of intersections between stimuli. In this series of studies we also looked at the impact of target stimuli similarity on the formation of relations between them, which is meant to affect the transfer of liking (Chapter 3). The other chapters tested the application of the SR task in real-life contexts, with a focus on the health domain. We looked at its effectiveness and strength in changing attitudes, identification and purchase intention towards eco-food products (Chapter 4). To bolster the potential of the SR task, in Chapter 5 we tried to use it in combination with a persuasive message to change attitudes and intentions towards green vegetables, in a population with initial negative inclination towards them. In chapter 6 we tested the SR on attitudes, attribution of properties and consumption of soft drinks in an innovative way that used negative self-stimuli (fattened versions of one's face) to increase the relevance of threatening health messages. The discussion is organized around the advances in this field provided by these lines of research with a specific focus on the debate between the associative and propositional interpretation of attitude change via associative procedures.

**General introduction**

In the first Handbook of Social Psychology, Allport (1935) introduced attitude as the most distinctive and indispensable concept in social psychology. The term *attitude* is used to refer to a one's overall evaluation of persons, objects, and issues. The most conventional contemporary definition of attitudes was provided by Eagly and Chaiken (1993): the authors defined an attitude as "*a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor*" (p. 1). Much research has assumed that evaluative processing of information is quite natural, pervasive and fundamental (Markus & Zajonc, 1985; Osgood, Suci, & Tannenbaum, 1957) in large part because of the adaptive and functional purposes of holding attitudes (see Pratkanis, & Greenwald, 1989). Moreover, our evaluations can vary in many ways and be based on different things. First, attitude can origin in different ways, depending on individuals' disposition as well as on the attitudinal target. People evaluate persons, objects, or issues on the basis of either emotional reactions (e.g., "Bob like pizza because eating pizza makes him feel happy"), rational beliefs (e.g., "Mike does not like Donald Trump because he thinks he can destroy US economy"), or past experiences and behaviors (e.g., "John does not like milk because he got sick the last time he drank it"; see Breckler, 1984; Zanna & Rempel, 1988).

After more than eighty years, the concept of attitude remains dominant in modern social psychology. Individuals' attitudes are crucial in determining many types of human behaviors. For instance, voting one or the other candidate at the presidential election might be determined by the extent to which we like the ideas of the political parties they belong to, even if we do not know anything of the candidates themselves. Similarly, if we like a certain apparel brand, we should be more likely to select a tee shirt of that brand when shopping. Thus, given the importance of attitudes in predicting and influencing many types of behavior, it is not surprising that much research has focused on the ways through which attitudes can form and change.

**Implicit and explicit attitudes**

In the last two decades, the focus of attitude research has shifted from understanding explicit attitudes (i.e., attitudes that people can report and for which activation can be consciously controlled)

to examining implicit attitudes (i.e., attitudes for which people do not initially have conscious access and for which activation cannot be controlled). The distinction between implicit and explicit attitudes has become a major theme in current social and experimental psychological research (Greenwald & Banaji, 1995). This distinction finds its origin in the rapid growth of implicit social cognition and the simultaneous development of measures of implicit attitudes. Measures of implicit attitude are based on the concept of automaticity and on the idea that attitudes can be treated as the expressions of automatic processes that occur spontaneously and outside of people's awareness or control (Moors & De Houwer, 2006). Therefore, explicit attitudes that are usually equated with deliberative, self-reported evaluations, are contrasted with automatic, implicit attitudes that are inferred from people's performance on response latency measures, such as the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), affective priming (Fazio, Jackson, Dunton, & Williams, 1995), semantic priming (Wittenbrink, Judd, & Park, 1997), the go/no-go association task (Nosek & Banaji, 2001), the Extrinsic Affective Simon Task (EAST, De Houwer, 2003), and the Affect Misattribution Paradigm (AMP, Payne, Cheng, Govorun, & Stewart, 2005). To get a vivid example, consider a researcher who wants to measure the implicit attitude towards Coke and Pepsi using an IAT, which can be considered as the most known and used among the indirect measures of attitudes. The IAT consists of a computerized classification task of two target categories and two attribute categories (e.g., positive versus negative stimuli) with a 7- or 5-block structure. The underlying logic of the IAT refers to response interference or compatibility. If one has an implicit preference for one Coke over Pepsi, it should be easier to classify positive words and Coke with a single key and negative words and Pepsi with another key than to classify negative words and Coke with the same key and positive words and Pepsi. One's implicit attitude towards Coke and Pepsi reflects the easiness of the task as evaluated through reaction times and error rates; the IAT effect is defined as the difference between these two response conditions.

Prior studies have indicated that implicit evaluations can be even better than explicit evaluations at predicting certain types of automatic or spontaneous behavior (e.g., Dovidio,

Kawakami, & Gaertner, 2002; Friese, Hofmann, & Schmitt, 2009; Perugini, Richetin, & Zogmaister, 2010; Fazio, 1990; Greenwald, Poehlman, Uhlmann, & Banaji, 2009) and play a crucial role in a number of important psychological phenomena including psychopathology (Roefs et al., 2011), addiction (Wiers & Stacy, 2006), and social interactions (Fazio & Olson, 2003). Relying on implicit rather than explicit attitudes can circumvent self-presentational motives (e.g., Dunton & Fazio, 1997) and can often uniquely predict spontaneous behaviors (e.g., McConnell & Leibold, 2001). Therefore, it is of particular importance to understand how both explicit and implicit evaluations originate and change.

Dispositional theorists conceptualized implicit and explicit attitudes as two dissociated, non-interacting types of evaluation simultaneously held toward the same object. Unlike explicit attitudes that develop in response to recent information, automatic evaluations were thought to reflect mental associations formed through early socialization experiences (e.g., De Hart, Pelham, & Tennen, 2006; Rudman, 2004). Once formed, these associations were assumed to be highly robust and resistant to change, as well as stable across both context and time. This idea did find support in many empirical studies reporting that explicit attitudes showed higher malleability to experimental manipulations than the implicit counterparts (Gawronski & Strack, 2004; Gregg, Seibt, & Banaji, 2006). However, more recent findings showed that novel implicit and explicit evaluations toward fictitious social groups can be experimentally induced (e.g., Hughes & Barnes-Holmes, & De Houwer, 2011). Such findings support the sensitivity of both explicit and implicit attitudes to current contextual influences and challenge the view that implicit attitudes reflect highly stable trait-like associations. From a more constructivist perspective, attitudes are seen as constructed associations. Thus, if an attitude is the association between the target concept and a representation of either positive or negative valence, the ease with which this valenced representation activates in mind can be influenced by contextual factors and manipulations.

**What drives implicit and explicit attitudes?**

Besides demonstrating that implicit attitude can also be malleable, research on attitudes has extensively focused on the attempt to understand the mental processes that underlie implicit and explicit attitudes. There are three important theoretical accounts that have tried to provide an interpretation of mental processes responsible for the expression of attitudes at their implicit and explicit levels. One is represented by the single-process associative models, which assume that both implicit and explicit evaluations are based on the formation and activation of associations (e.g., Baeyens, Eelen, Crombez, & Van den Bergh, 1992; Fazio, 2007). Implicit evaluations are assumed to reflect the automatic activation of particularly strong associations. In contrast, explicit evaluations may also depend on the more effortful activation of other associations. In line with the MODE model (Fazio, 1990), the extent to which these more complex associations activate depends on both internal and external conditions: when a person has the opportunity and motivation to consider additional information s(he) will likely use further associations (Olson & Fazio, 2009). A second and different way to look at attitudes with respect to their underlying processes is proposed by dual process models, such as the Reflective Impulsive Model (RIM, Strack & Deutsch, 2004). This model postulates that people process information via two different cognitive routes. One is the fast, frugal, and effortless route, based on automatic association formation processes. The other system is slow, effortful, and entails the rule-based formation of propositional representations. Whilst associations reflect the strength of the relation between concepts, propositions are structured representations that include information about the nature of their relation (e.g., "A causes B" or "A is a consequence of B"; see Lagnado, Waldmann, Hagmayer, & Sloman, 2007). It is often assumed that explicit evaluations reflect the operation of propositional processes such as the activation and validation of propositions whereas implicit evaluations are determined only by the activation of associations (Strack & Deutsch, 2004). In line with this idea, early theorists of dual process models proposed that explicit attitudes are more likely affected by information provided in the form of propositions, while implicit attitudes are more sensitive to repeated pairings between the attitudinal target and valences stimuli (Rydell & McConnel, 2006). The Associative Propositional Evaluation model (APE, Gawronski &

Bodenhausen, 2006; 2011) is probably the most prominent among the dual process models. This model is innovative in its class as it is based on the idea that associative and propositional processes are not mutually exclusive in affecting implicit and explicit attitudes. Conversely, the authors proposed a series of ways in which associative and propositional processes interact in affecting implicit and/or explicit attitude (see Gawronski & Bodenhausen, 2006). A third and final recent evaluation model assumes that both implicit and explicit attitude are affected by a single type of process, which is propositional in nature (De Houwer, 2009, 2014; Mitchell, De Houwer, & Lovibond, 2009). The idea of this propositional model is that both types of attitude – implicit and explicit – are the direct consequence of propositions formed about the relations between stimuli. Thus, implicit attitudes are based on propositions that are more easily available in mind, while explicit attitudes might require further validation process over these available propositions (De Houwer, 2014).

**Attitude change as an instance of learning: A functional perspective.**

If associative and propositional processes affect attitudes at distinct levels of expression, then interventions and manipulation tapping on distinct processes should affect differently implicit and explicit attitudes (Gawronski & Bodenhausen, 2006). Among the pathways that lead to change in attitudes, persuasion and associative learning represent two alternative and (apparently) opposed examples. Persuasive communication involves changes in the set of propositions that are considered to be relevant for an evaluative judgment. In persuasion studies, participants are usually exposed to a persuasive message containing either strong or weak arguments about an attitude object and different kinds of peripheral or heuristic cues, such as the expertise of the source, the likability of the source, or consensus information (for reviews, see Chen & Chaiken, 1999; Petty & Cacioppo, 1986). Adopting either a central or a peripheral analysis of the persuasive communication might depend on a number of dispositional and situational factors. Therefore, attitude change through the persuasion route is meant to require an effortful processing of words and sentences. On the other hand, attitude can form through the associative route. The prototypical case of attitude change through associative

procedure is evaluative conditioning (EC, Levy & Martin, 1975). EC refers to a change in liking towards a neutral object (i.e., conditioned stimulus, CS) due to its repeated presentation in contingency of space and time with a valenced object (i.e., unconditional stimulus, US, see De Houwer, Thomas, & Baeyens, 2001, and Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010, for reviews). Contrary to attitude change resulting from persuasion, EC effects on attitude should require little cognitive effort and thought.

The focus of the present dissertation is on associative attitude change. I would like to point out that the term *associative* could be misleading. In fact, the term does not apply to the underlying process, rather to the way in which an experimental manipulation is characterized. For instance, EC effects result from a procedure in which stimuli are linked through newly formed associations based on spatio-temporal contingency. In line with this view, a functional approach to attitude change helps in clarifying that the effect of a given procedure should not be equated to the underlying mental process. In fact, from a functional perspective, both persuasion and associative attitude change (e.g., EC) should be treated as instances of learning, where learning is defined as "*the changes in the behavior of an organism that are the results of regularities in the environment of that organism*" (De Houwer, Barnes-Holmes, & Moors, 2013, p. 633). Therefore, any attitude change can be seen as a change in behavior that results from environmental regularities. For instance, an IAT score that measures the implicit preference towards a sport brand reflects in functional terms the change in a behavior (i.e., categorizing valences and neutral stimuli with the same key) that might result from an advertisement in which an expert explains the advantages of wearing apparel of that brand (i.e., persuasion) or from the repeated exposure to the brand logo together with your favorite soccer player (i.e., evaluative conditioning). In both cases, the newly expressed attitude is a function of environmental regularities. The concept of *regularity* encompasses all states in the environment of the organism that entail more than the presence of a single stimulus or behavior at a single moment in time (De Houwer, Barnes-Holmes, & Moors, 2013).

**Associative evaluative learning: from 'simple' to 'intersecting' regularities.**

Over the last thirty years, researchers have studied a number of avenues through which novel preferences may be formed and existing ones altered through associative procedures (De Houwer, 2007; De Houwer, Barnes-Holmes, & Moors, 2013). Beyond EC, environmental regularities are responsible for attitude change via mere exposure, as well as operant evaluative conditioning. In mere exposure, people change their attitude towards an object in a favorable way after being repeatedly presented with the object itself (e.g., Zajonc, 1968; Moreland & Topolinski, 2010). Based on a logic that resembles EC, operant evaluative conditioning relies on the idea that relating one neutral stimulus with a valenced behavioral response (e.g., approach or avoidance movement) changes one's attitude towards the object (Krieglmeyer, Deutsch, De Houwer, & De Raedt, 2010). Thus, a regularity can be the presence of a single stimulus at multiple moments in time (i.e., mere exposure), or the presence of multiple stimuli at multiple (or single in the case of one-trial learning) moments in time (i.e., evaluative conditioning) or even the presence of stimuli and behavior at multiple moments in time (i.e., operant evaluative conditioning). Because each of these three pathways involves changes in liking that are due to regularities in the environment, each of them can be regarded as a specific type of evaluative learning (for a detailed treatment of how different types of learning involve different types of regularities, see De Houwer, Gawronski & Barnes-Holmes, 2013).

However, attitude change may also origin from more complex pathways that involve regularities. Hughes, De Houwer, and Perugini (2016) came up with the idea that learning to like and dislike an object can be the consequence of intersections between environmental regularities. Imagine you just moved to a new city and while having a coffee in a bar you met Mike. When you start introducing yourselves, Mike says that he comes from a small town where you used to spend your holidays with your kids ten years ago. From now on, the small town represents the intersection that connects Mike and an enjoyable event of your life. As a consequence, even though you do not know anything else about him, you will probably look at him with some initial degree of liking. The situation described is an everyday example of attitude change (or formation) via intersecting regularities (IR). Hughes and colleagues (2016) showed that attitude change via IR can be also

induced through experimental manipulations based on associative procedures. For instance, in one of four studies they presented, participants performed a simple categorization task in which they used the same key (e.g., "F") when presented with equal probability, a neutral brand name or positively valenced stimuli and a different key (e.g., "J") when presented another neutral brand name or negative valenced stimuli. The regularities (i.e., pressing key upon stimulus presentation) regarding the first neutral brand and positive stimuli thus intersected in terms of response as well as the ones regarding the second neutral brand and negative stimuli. In other words, an intersection was established through a common response between the valence and the neutral stimulus in either of the two sets of stimuli (neutral brand and positively valenced vs. alternative neutral brand and negatively valenced). The results demonstrated increased implicit and explicit liking towards the brand sharing the response with the positively valenced stimuli relative to the alternative brand. In other words and in line with the hypothesis, the establishment of intersecting regularities allowed for a transfer of the evaluative properties from the valenced source to the neutral target (see also Ebert, Steffens, von Stulpnagel, & Jelenec, 2009).

Changing attitude via IR is a new associative pathway characterized by the fact that it does not require people to experience any direct relation between stimuli. In EC, stimuli appear together on the same place for multiple times. In learning via IR, people merely get to learn that stimuli relate because of a third element they have in common. Therefore, while EC requires a direct pairing between stimuli, IR works through indirectly established relations between stimuli. In the EC literature, researchers have shown that preferences towards target object do not necessarily result from the direct pairing between a valenced and a neutral object. Let us consider for instance *second order conditioning* and *sensory preconditioning* (Hammerl & Grabitz, 1996; Walther, 2002; Gast & De Houwer, 2012). Second order conditioning is a change in liking towards an object (i.e., CS2) that results from two distinct phases. In the first phase, a US is paired with a target object (i.e., CS1) and in the second, CS1 is paired with CS2. Sensory preconditioning is similar but based on a reversed procedure: pairing between CS1 and CS2 comes first, followed by the pairing between CS1 and US.

In both cases, a change in liking towards CS2 occurs without any direct pairing with the US, but it is instead due to the fact that both the US and CS2 have been paired with the same stimulus (i.e., CS1). Thus, from a broader perspective, evaluative learning via IR can define previously demonstrated attitude change effects based on indirect relationships between valenced and non-valenced stimuli.

**Using the self to change attitudes.**

Over the years, the idea that objects related to the self are better liked has been largely demonstrated and interpreted under distinct, albeit related, theoretical interpretations. For instance, Nuttin (1985) observed that letters that appear in an individual's name are generally preferred (*name letter* effect) over the others. This effect has been reframed in terms of *mere ownership* (Beggan 1992; Nuttin, 1987). We observe a mere ownership effect from a process where an object possessed or given to an individual acquires greater liking compared to when that object is not owned. Based on a similar principle, the *implicit egotism* effect (Pelham, Mirenberg, & Jones, 2002) occurs when people's positive associations about themselves spill over into their evaluations of objects related to the self. Gawronski, Bodenhausen, and Becker (2007) showed that even arbitrary relationships between one's self and an unknown object can be nonetheless perceived and result in higher liking for the latter. In one of their studies, participants were told that they would be given one of two color prints as a gift in gratitude for their participation. In order to determine which of the two pictures a participant would receive, the experimenter rolled a dice. Results showed that, even randomly assigned objects acquire greater liking than those that are not (Gawronski et al., 2007, Study 4).

To explain how self-object relations can lead to the formation of positive attitudes toward target objects, Greenwald, Banaji, Rudman, Farnham, Nosek, and Mellott (2002) posited that self-object relationships may occur unconsciously through the development of balanced triads of associations in a fashion similar to that of *cognitive consistency theory* (Festinger, 1957). These balanced triads include the *self*, an *object* that is associated with the self-concept and a mental representation of *valence*. A measured link between the self and valence is defined as *self-esteem* (Greenwald and Farnham, 2000); the link between an object and valence is an attitude; a

measured link between an object and the self is an implicit *self-identity* (Rudman, Greenwald, and McGhee, 2001). To facilitate the balancing of these objects in memory, Greenwald and colleagues posited a *balance-congruity principle*. Two unlinked objects in memory (e.g., valence and object) that share "first-order" links with a third object (e.g., self-concept) should develop a mutual association (Greenwald, Banaji, et al. 2002). Assuming an a priori link between self and valence, conceptualized as implicit self-esteem (Greenwald & Banaji, 1995), an association created between the self-concept and an object in the environment should produce a new link between the object and that valence. To the extent that implicit self-esteem is positive in the vast majority of cases (Yamaguchi et al., 2007), it is thought that this newly formed self-object association should be positive in nature.

**Propositional versus associative processes of attitude change via the self.**

Although the balance congruity principle offers a good explanation of attitude change driven by the positivity of the self, it does not really account for the potential underlying mechanism of the phenomenon. Drawing on the APE model (Gawronski & Bodenhausen, 2006), implicit and explicit evaluations should be understood in terms of their underlying mental processes (see also Strack & Deutsch, 2004). The first source of evaluative tendencies resides in associative processes, which build the basis for implicit attitudes. Evaluations resulting from associative processes are best characterized as affective reactions consequential of the particular associations that are activated automatically upon encountering a relevant stimulus. The second source of evaluative tendencies comes from propositional processes, which build the basis for explicit attitudes. Evaluations resulting from propositional processes can be characterized as evaluative judgments that are based on syllogistic inferences from any information relevant for a given judgment. This has direct relation with changing attitude using the positivity of the self. Starting from the idea that people have a generally positive and high self-esteem (Bosson, Swann, & Pennebaker, 2000; Yamaguchi et al., 2007), there are at least two potential ways that can account for the phenomenon and that subsume distinct mental processes. On the one hand, people might preserve a positive view of themselves by adopting some

sort of cognitive consistency strategies. Cognitive consistency is exclusively a concern of propositional reasoning (Gawronski & Bodenhausen, 2006). Applied to attitude change via the self, one should use the proposition *"I have a positive view of myself"* and *"The object X relates to myself"* to infer the proposition *"The object X is positive"*. In fact, the opposite evaluative proposition (e.g., *"The object X is negative"*) would be dissonant with the two premises. On the other hand, it is also possible that attitude change results from a spread of activation that takes place in a network like manner. Given that a relation between an object and the self has been established (e.g., by telling people that they own that object), the view of the objects triggers the activation of the self which in turn activates the concept of positivity. In line with this reasoning, it seems plausible that distinct ways to create self-object relations might activate mental processes that affect differently implicit and explicit attitude change.

Past research on attitude change resulting from the transfer of liking from the self to an object has been dominated by two alternative manipulation routes. On the one hand, studies on ownership regard the evaluation of objects that have been either chosen or assigned during the experimental procedure (Gawronski et al., 2007, study 3 and 4; Gawronski & Lebel, 2008, study 3). On the other hand, the effect of the self on attitude change has been operationalized through self-object links that resulted from associative learning procedures where self-related stimuli are presented simultaneously with target objects (EC, Walther & Trasselli, 2003; Zhang & Chan, 2009).

**Integrating the IR principle to attitude change via the self: The Self-Referencing task.**

The previous paragraphs focused on two specific aspects of attitude change. One referred to *how* implicit and explicit attitude change, the other described *what* can change implicit and explicit attitude. In answering the '*how*' question, the IR principle was introduced as a novel pathway through which attitude indirectly form and change based on an associative procedure. Concerning the *'what'* question, the self was described as a special source of attitudinal change, whose effects on implicit and explicit attitude might be driven by distinct processes and through distinct manipulations. Beyond EC, another associative pathway through which the self has shown effective in changing attitudes is

the Self-Referencing (SR, Prestwich, Perugini, Hurling, & Richetin, 2010; Perkins & Forehand, 2012). The SR task represents a prime example of an associative learning paradigm that relies on a specific operationalization of the IR principle. Although a more detailed description of the paradigm in its standard form and in most of its possible variations will be provided throughout the next chapters, it is important to highlight what are the differences between a SR task and a standard EC that uses the self as source of attitude change. In EC, the transfer of attitudes from the self to an object requires repeated direct pairings between self-related stimuli and object exemplars (Walther & Trasselli, 2003) whereas in the SR task participants are never exposed to any self-object direct pairings. Instead, they learn that the self and an object (and that 'others' and another object) are related because they are categorized through the same action. Operationalized through the IR principle, the SR task represents an associative procedure through which people acquire stimuli relations in an indirect fashion. What they learn from the completion of this categorization task is that two classes of stimuli (e.g., the self and Target 1) share a common behavioral response. Inferring that the two objects are related because of this shared response requires a further step in which people confer a symbolic meaning to the stimuli and their relation. In particular, De Houwer & Hughes (2016) recently proposed the idea of the symbolic meaning of associative evaluative learning. Based on their interpretation, people use *proximal cues* (regularities) to detect a relation between stimuli. Through *distal cues* that they retrieve from their life's history, they qualify this relation in a certain way. For instance, in the case of the SR task, one notices that self-stimuli and the exemplars of a target share something (e.g., an action). Given that, normally, things that share something are also related or similar in some senses, one might treat them as related. Applying this distal knowledge to the environmental regularities through which objects are linked results in a transfer of properties from the self to the target object. The resulting effect of this IR based procedure is that people develop higher preferences for the target that is categorized through the same action as the self.

**The present dissertation.**

This dissertation focuses on associative evaluative learning via IR when the self is used as an attitudinal source of change. The SR task will be first introduced by describing its characteristics and reviewing all studies conducted by our lab in the last six years to shed light on the robustness of the effect and to provide important indications concerning the potential boundary conditions of the SR task. Thus, chapter 1 presents a meta-analytical review of the SR task conducted across 53 studies. This meta-analysis is based on Mattavelli, Richetin, Gallucci and Perugini (currently under review after a revision in *Journal of Experimental Social Psychology*). This review has the unique advantage of summarizing all the empirical findings gathered across different SR studies. It offers the opportunity to look at the effectiveness of the paradigm at both implicit and explicit attitude level from a higher perspective, before going into more specific details through the lines of research presented in the following chapters. At the same time, by treating methodological variations that occur across different primary studies as potential moderators of the SR effect, this meta-analytical review is informative for all those researchers interested in using the paradigm for future studies, in that it describes the conditions under which the magnitude of the effect varies.

Then, the attention shifts towards two specific lines of research that exploit the two main characteristics of the SR task – the use of the self to change attitude and the IR principle – to propose the idea of *indirect attitude change* through modified versions of the task itself.

Chapter 2, that will eventually become a manuscript, tests whether the SR effect is determined by either the valence or the relevance of the stimuli used to represent the self-concept. Across four studies, we tailored the SR task in a way that slightly diverges from its standard form. Rather than self- and others-related pronouns (e.g. me, my, mine), we used personality traits that are either relevant (considered as representative of one's self) or irrelevant (considered as non-representative of one's self) for the subjects and asked them to categorize those traits with the same keys as two fictitious brands in the SR task. In the first three studies, traits adjectives differed only in terms of their relevance for the self. In a fourth study, they also varied in valence within the SR task: self-relevant traits were negative, while self-irrelevant traits were positive. Testing whether either valence

or relevance drives the SR effect when personality traits are used as source stimuli has important implications for understanding what type of mental mechanism underlies the effect. While a valence-driven effect would be more in line with an associative explanation of attitude change, a relevance-driven SR effect would imply the higher-order ability to form relations between stimuli, which is a pre-requisite of propositional reasoning.

Chapter 3, based on Mattavelli, Richetin, and Perugini (in preparation), centers on indirect transfers of liking but from a different perspective. Although the characteristics of the self remain central, it focuses more specifically on the IR principle, testing the potential impact of exploiting this learning mechanism through multiple indirect chains between stimuli and events. Across five studies, we tested indirect attitude change through multiple intersecting regularities. In essence, we tested whether objects never presented in the SR task and thus never directly paired with the self can acquire the valence of the self simply because they share a common action with the object targeted in the SR task. Moreover, this series of studies is important from the perspective of the target object on which researchers want to change attitude. The five studies tested the impact of indirect SR effect across stimuli the same domain, when a brand [group] is paired with the self in the SR task and then with another brand [group] in a further categorization task, or across different domain, when a brand [group] is paired with the self in the SR task and then with another group [brand] in a further categorization task. The chapter therefore focuses on the indirect effect of the SR task and on the role of target stimuli similarity in qualifying the effect.

The last three chapters are organized around the potential concrete application of the SR effect. Chapter 4, based on Richetin, Mattavelli, and Perugini (2016), capitalizes on the important role of the self in consumer or pro-environmental behavior, to test implicit and explicit attitudes change toward organic food brands using the self. Across two studies, we also tested the SR task with respect to its capability of changing hypothetical shopping behavior, brand identification, and attitude change persistence. Participants' implicit and explicit attitudes, hypothetical choice of products and identification with either eco-brands were the dependent variables. These two studies are important

because they provide evidence of the ecological validity of the SR task by testing its effectiveness on a behavioral intention towards the target of change. Moreover, in one of the studies the persistence of the SR effect was tested, in order to see whether participants like and identify more with the brand originally paired with the self even after removing the pairing with the self.

Chapter 5 reports a paper by Mattavelli, Avishai, Perugini, Richetin, and Sheeran (accepted for publication in *Annals of Behavioral Medicine).* It presents a single study contribution in which a green vegetables tailored version of the SR task was administered in combination with a persuasive communication on a sample of American participants who did not like green vegetables. The study is a unique attempt in health psychology research to use an associative based procedure that relies on the positivity of one's self to increase implicit and explicit liking in the context of food. The study also measured participants' readiness to change their behavior as a function of either the SR or the persuasive manipulation. The way in which this study was designed and the target population tested make this study relevant from the perspective of the potential of the SR to change attitude and intentions towards a health related behavior and the effect of the associative task when combined with an alternative form of attitude change.

Finally, chapter 6, which is based on Mattavelli, Perugini, Richetin, and Sheeran (in preparation), also presents a test of the SR task in the health domain. Critically however, this study goes further. It used a totally innovative version of the SR task, where self-stimuli were represented by either a normal or a fattened version of one's face. Moreover, it tested the transfer of extra-evaluative properties from the self to the target object. In particular, two studies tested separately whether pairing a soft drink with a fattened version of one's own face result in lower liking towards the drink, compared to another drink paired with the unaltered self-face. This effect was measured on implicit and explicit attitudes in the first study as well as on implicit and explicit transfer of attributes from the self to the targeted drinks in the second study. Taken together, these studies offer a novel, effective way to exploit the properties of one's self to determine both attitudes and the attribution of properties towards existing objects.

Each of these chapters is introduced by a summary containing the theoretical framework, the rationale and the aim behind the work. Given that all these chapters are based on individual manuscripts that are either published, accepted, under review or in preparation, the reader will notice some overlap among the theoretical frameworks that introduce the distinct chapters.

Finally, the general discussion summarizes the findings presented in the previous chapters, elaborating on their implications from both a theoretical and an applied perspective. Moreover, through this chapter the reader has the opportunity to understand the progresses provided by this series of empirical research with respect to the possibility to form and change attitudes via the self in an associative paradigm based on intersecting regularities. As a final point, based on the results of each line of research presented in the dissertation, the general discussion raises new empirical questions and proposes either focused studies or more broad lines of research to test these questions in future.

The work presented in this dissertation would not have been possible without the help of many people. The studies described in each chapter have been conducted in collaboration with Marco Perugini, Juliette Richetin, Marcello Gallucci, Paschal Sheeran and Aya Avishai.

I would like to express all my deepest gratitude to my supervisors, Marco Perugini and Juliette Richetin, for their patience in teaching me most of the things I know, for the precious advices they gave me in the last three years and for their help in improving the quality of this dissertation. But above all, I want to thank both of them for passing down to me their passion and enthusiasm in doing research.

I am also deeply grateful to Paschal Sheeran, for giving me the opportunity of spending five months working in his lab at the University of North Carolina, Chapel Hill. I really appreciated his trust in my (little) expertise and the freedom he gave me in applying my knowledge to his lines of research.

A huge 'thank you' also goes to Jan De Houwer, probably the most brilliant and inspiring mind I had the pleasure to work with, for giving me the chance of staying for four months in his lab

at the University of Ghent. Among all the Liplab people, I will be always thankful to Sean Hughes, for his help, his advices and his effort in teaching me things from a different perspective.

I wish to thank all the other persons I had the chance to collaborate with or who helped me in making the chapters of dissertation possible: Aya Avishay and Katelyn Rose (from UNC, Chapel Hill), Marcello Gallucci and my friend and colleague Giulio Costantini (from Milan).

Last but not least, I cannot forget to thank all the bachelor and master students who helped me in running experiments and collecting the data that I am about to present and discuss in the following pages.

# Chapter 1.

# The Self-Referencing task: Theoretical overview and empirical evidence.[1]

---

[1] This chapter is based on Mattavelli, S., Richetin, J., Gallucci, M., & Perugini, M. The Self-Referencing task: Theoretical overview and empirical evidence. The manuscript is currently under review in *Journal of Experimental Social Psychology* after an invitation to revise and resubmit it.

**Overview**

This chapter provides both a theoretical framework and a quantitative summary of the empirical evidence on a recent evaluative learning procedure, namely the Self-Referencing (SR) task. The SR task is introduced by describing its key features, with emphasis on the intersecting regularities principle as its underlying learning mechanism and on the potential advantages related to use the self as a source of evaluative learning. Then, across 53 studies[2], we meta-analyzed the SR effect on implicit and explicit evaluative and extra-evaluative domains. The meta-analytical technique also allowed us to test for boundary conditions of the effect. We identified potential moderators related to either general, specific, or task-unrelated characteristics of the SR paradigm and tested their power to account for variations in the effect, with special attention on the role played by memory of the intersecting regularities. Overall, findings suggest that the SR task is effective in leading to both evaluative and extra-evaluative learning. We discussed the robustness of the effect and some relevant findings pertaining to the moderators. In particular, we focused on a) the fact that the effect can be qualified by the type of stimuli used in the task as either source or target of attitudinal change and b) the importance of processing the intersecting regularities, which is as a key driver that qualifies the magnitude and the direction of the effect. Both practical and theoretical implications from the moderation analyses are also discussed.

---

[2] The fifty-three studies included ten out of the fourteen primary studies presented across the following chapters of the present dissertation. We excluded from the current meta-analysis study 4 presented in chapter 2 and study 5 in chapter 3 (results were not available at the time of the elaboration of the submitted manuscript) and both studies presented in chapter 6 (excluded because the version of the SR task was too different from its standard implementation).

**Introduction**

In many daily life situations, individuals' preferences drive their behavior. One central question thus is how they come to like certain things and dislike others. Even though research postulates the existence of innate and evolutionary-relevant preferences (see Poulton & Menzies, 2002 for a non-associative acquisition of fear), most of our likes and dislikes result from processes that involve learning (Rozin & Millman, 1987). Therefore, understanding the ways through which individuals learn their preferences is an important issue. Theories of evaluative learning[3] assume that both implicit and explicit preferences can be acquired through alternative routes (e.g., Bodenhausen & Gawronski, 2013; Gawronski & Bodenhausen, 2006, 2011). A more deliberate route involves the scrutiny of descriptive information about an object. For instance, we can learn to like a car through a commercial ad in which persuasive arguments emphasize its quality. A less deliberate route implies the association of the attitude object with positive or negative features of elements of the context in which it is encountered (Bodenhausen & Gawronski, 2013; Chen & Chaiken, 1999; Petty & Wegener, 1999). Evaluative conditioning (Hoffman, De Houwer, Perugini, Baeyens, & Crombez, 2010), which consists in the change in valence of a target object after its repeated presentation with a valenced stimulus, has been the predominant example of research on associative evaluative learning. Thus liking can result from the fact that the car has been paired with a celebrity in the ad.

Still through the less deliberate route, a newly identified associative learning mechanism, namely the Intersecting Regularities (IR) principle (Hughes, De Houwer, & Perugini, 2016) has been

---

[3] The term 'evaluative learning' subsumes both attitude change and attitude formation. Attitude change is a change in the particular quality of the evaluative response evoked by a stimulus as the result of new experiences related to that stimulus. It differs from attitude formation, in which the stimulus initially does not evoke an evaluative response but later does. Even though the terms belong to different traditions and literatures, both change and formation of an attitude can be seen as instances of evaluative learning, in that they both imply new learned evaluative responses to given objects.

proposed as an alternative pathway to formation or change in objects evaluation. The IR principle underlies the Self-Referencing (SR) task, an evaluative learning task also characterized by its reliance on the self as the positive source. Hence, a SR effect is conceptualized as the result of a learning procedure where target stimuli are categorized through the same behavioral response as self-related stimuli (i.e., pressing the same key or moving a joystick in the same direction). In the following, we will first present the characteristics of the SR task and discuss its underlying learning mechanism and the advantages of using the self as an attitudinal source to induce both implicit and explicit evaluative and extra-evaluative changes. Second, through a meta-analysis on a substantial set including all studies conducted by our research group, as detailed later, we will provide a quantitative summary of the SR effect assessing the power of the effect as well as its homogeneity. Results from moderation analyses organize the discussion about the potential boundary conditions of the SR effect.

**Learning to like through regularities**

The pathways through which people can learn to like and dislike are various. The valence of an object can increase together with its familiarity (i.e., mere exposure effect, Zajonc, 1968). Moreover, evaluative conditioning (EC hereafter), probably the most known among the associative learning pathways, conceptualizes the change in valence of a stimulus due to a procedure that involves the repeated presentation of that stimulus (i.e., conditioned stimulus, CS) with a valenced stimulus (i.e., unconditioned stimulus, US) either in contingency of space and time or in a sequential manner, such that the presentation of one stimulus predicts the appearance of the other one (De Houwer, 2007; Hoffman et al., 2010).

Recently, Hughes et al. (2016) proposed that preference may also origin from more complex learning mechanisms and identified the intersecting regularities (IR) principle. To get a concrete idea of how the IR principle operates, let us consider the first experiment in Hughes et al. (2016). Participants performed a simple categorization task in which they used the same key (e.g., "F") when presented, with equal probability, a neutral brand name or positively valenced stimuli and a different key (e.g., "J") when presented another neutral brand name or negative valenced stimuli. The

regularities (i.e., pressing key upon stimulus presentation) regarding the first neutral brand and positive stimuli thus intersect in terms of response as well as the ones regarding the second neutral brand and negative stimuli. In other words, an intersection was established through a common response between the valence and the neutral stimulus in either of the two sets of stimuli (neutral brand and positively valenced vs. alternative neutral brand and negatively valenced). The results demonstrated increased implicit and explicit liking towards the brand sharing the response with the positively valenced stimuli relative to the alternative brand. In other words and in line with the hypothesis, the establishment of intersecting regularities allowed for a transfer of the evaluative properties from the valenced source to the neutral target (see also Ebert, Steffens, von Stulpnagel, & Jelenec, 2009 for another exemplification of learning via IR).

**The Self-Referencing task**

The Self-Referencing (SR) task, initially conceived as a peculiar case of EC (Prestwich, Perugini, Hurling, & Richetin, 2010), represents a prime example of an evaluative learning paradigm that relies on a specific operationalization of the IR principle, additionally characterized by the use of the self as a source (valenced) stimulus. The standard version of the SR task (Prestwich et al., 2010) consists of four blocks of 40 trials each. In the first two blocks participants categorize stimuli belonging to a Target A and words related to one source (i.e., the self) to one response key (e.g., 'E') and Target B stimuli and words relating to the contrast source (i.e., others) to a different response key (e.g., 'I'). Participants then repeat the two blocks of 40 trials switching the keys assigned to the categories (i.e. Target A and Self-words assigned to the 'I' key, and Target B and Other-words to the 'E' key). This is done to prevent any systematic pairing between keys and stimuli, hence keys do not acquire any specific evaluative meaning. In case of incorrect classification, a red-X appears on screen and remains until correction to emphasize the need to learn the correct intersecting regularities between pairs of stimulus categories throughout the whole task. A SR effect from the SR task relies on the sharing of specific features between the two pairs of stimuli categories involved in the task. More specifically, there are four contingencies underlying the SR task. Two of them (i.e., "*if self-*

*related, press the 'E' key"* and *"if first neutral stimulus, press the 'E' key"*) intersect in terms of a shared response and outcome, as in both contingencies pressing the 'E' key is correct. Likewise, two other contingencies (i.e., *"if other-related, press the 'I' key"* and *"if second neutral stimulus, press the 'I' key"*) also intersect insofar as they also share a response and outcome.

The first SR studies (Prestwich et al., 2010) used a control condition in which participants did not perform any task. However, this asymmetry could have raised doubts about the underlying mechanism of the SR effect (e.g., salience of the self). Therefore, the most used (and recommended) control condition of the SR task requires participants to perform an identical categorization task where the opposite source-target relationships are established via intersecting regularities. Specifically, the target categorized with the same key as the self in the experimental condition has to be categorized through the same key as 'other' in the control condition and vice-versa for the contrast target. This control task has been used in most of the targeted studies as it rules out any alternative explanation of the SR effect (see Chapter 5 for another type of control task the eliminates any confounding explanation of the SR effect).

Although the effect of the SR task might be interpreted as the joint consequence of both an intersection between one target and the self and between a contrast target and the category 'Other', a set of target studies has shown that that the SR effect is primarily driven by the common action required to categorize one target object and the self. Perugini, Zogmaister, Richetin, Prestwich, and Hurling (2013) showed in three studies that the category 'others' did not reveal any impact on the implicit attitude measures neither when examined in isolation (study 1) nor when defined positively by using the most positive idiosyncratically chosen exemplars of the category (study 2 and study 3).

Given the reliance on the IR principle, the effect from SR task is meant to be more likely if the individual learns that the two elements share the same response. A memory question administered after the SR task thus checks for individuals' ability to recollect the correct IR pairing resulting from the learning procedure at the level of the action (i.e., *"Throughout the task you pressed the same key to classify the self and one object. Do you remember which one?").* Participants are classified as 1)

correct IR memory if they indicate the correct target object, 2) incorrect, if they indicate the wrong target and 3) without IR memory if they opt for the 'I do not remember' option. This question is somehow the conceptual equivalent of contingency awareness (or memory) in EC (see Sweldens, Corneille, & Yzerbyt, 2014 for a review of the role of awareness in EC).

**The positivity of the self and attitude change**

The other peculiarity of the SR task is the use of the self as positive source to generate changes in objects' evaluation. People have a generally positive and high self-esteem (Bosson, Swann, & Pennebaker, 2000; Yamaguchi et al., 2007) and develop strategies to preserve a positive view of themselves. Over the years, the idea that objects related to the self are better liked has been largely demonstrated and interpreted under distinct, albeit related, theoretical interpretations. For instance, Nuttin (1985) observed that letters that appeared in an individual's name are generally preferred (*name letter* effect). This effect has been interpreted in terms of *mere ownership* (Beggan 1992; Nuttin, 1987), conceptualized as the result of a process where an object possessed or given to an individual acquires greater liking compared to when that object was not owned. Based on a similar principle, the *implicit egotism* effect (Pelham, Mirenberg, & Jones, 2002) occurs when people's positive associations about themselves spill over into their evaluations of objects related to the self. Gawronski, Bodenhausen, and Becker (2007) showed that even arbitrary relationships between one's self and an unknown object can be nonetheless perceived and result in higher liking for the latter.

Past research on attitude change resulting from the transfer of liking from the self to an object has been dominated by two alternative manipulation routes. On the one hand, studies on ownership regards the evaluation of objects that have been either chosen or assigned during the experimental procedure (Gawronski et al., 2007, study 3 and 4; Gawronski & Lebel, 2008, study 3). In other studies, self-object links result from associative learning procedures where self-related stimuli are presented simultaneously with target objects (EC, Walther & Trasselli, 2003; Zhang & Chan, 2009). From the perspective of the Associative-Propositional Evaluation model (APE, Gawronski & Bodenhausen, 2006) implicit and explicit attitude change can result from different underlying processes, which are

associative and propositional processes. It is therefore assumed that distinct manipulations might have a differential impact on either implicit or explicit attitude change. For instance, the APE model proposes that, under specific circumstances, associative learning manipulations (e.g., EC) can produce associative evaluations that are inconsistent with subjectively valid propositions: such manipulations result in a change in implicit liking without affecting the explicit component. Therefore, to be effective on both implicit and explicit attitude change, a manipulation should act at both the peripheral and the central routes. The former implies changes at the level of mental associations, while the second requires a propositional validation of the newly generated links. Although both ownership and associative learning manipulations showed a general effect on liking, less is known about their specific impact on measures of implicit and explicit attitude. Gawronski & Lebel (2008, study 3) demonstrated that the effects of self-object links resulting from an ownership manipulation influenced implicit but not explicit attitudes. Specifically, these links led to a preference for owned over non-owned objects on the explicit measure only when participants focused on their feelings for these objects, but not when they thought about reasons for their preference. Drawing on Strack and Deutsch's (2004) Reflective–Impulsive Model, people might transform their affective reactions into propositional format (e.g., a positive affective reaction to object X is transformed into the proposition "I like X"), but the resulting proposition is subject to syllogistic inferences that assess its validity for an evaluative judgment (Bodenhausen & Gawronski, 2013). On the other hand, all the studies using EC to establish self-object links have solely focused on the effect on explicit liking, therefore not allowing us to draw any conclusion about a potential differential impact of associative manipulations on implicit or explicit liking. By systematically assessing liking at both levels, the SR studies meta-analyzed in this work will provide robust empirical demonstration of the impact of an associative learning procedure involving the self and based on the IR principle on both implicit and explicit attitude change.

**The self beyond attitude**

Greenwald, Banaji, et al. (2002) posited that self-objects relationships may occur unconsciously through balanced triads of associations, in a fashion similar to that of *cognitive consistency theory* (Festinger, 1957; Heider, 1958; Osgood & Tannenbaum, 1955). The transfer of positivity from the self to an object is therefore conceptualized as the result of balanced triads that include the *self*, an *object* associated with it, and a mental representation of *valence*. Assuming an a priori link between self and valence, a novel relationship between the self-concept and an object in the environment should produce a new link between the object and that valence. To the extent that self-esteem is positive for most people (Farnham, 1999; Yamaguchi et al., 2007), the newly formed associations between the self and an object lead to a positive valence of this object.

However, the self is characterized not only by the valence dimension. In fact, one's self-concept is thought to be best represented as a complex dynamic knowledge structure, consisting of multiple self-relevant attributes, traits, and roles, organized into a multifaceted and interconnected network that then filters and guides experience and behavior (e.g., Markus & Wurf, 1987; McConnell et al., 2011). One such prominent aspect is identification, that is, the level at which people tend to perceive themselves as close to either groups, categories, or objects. The notion that people incorporate social relationships and group memberships into their extended self-concept has received ample theoretical (e.g., Brewer & Gardner, 1996; Markus & Kitayama, 1991) and empirical considerations (Aron, Aron, & Smollan, 1992; Tropp & Wright, 2001). Although research in social psychology and intergroup relations have shown that overlapping between self and objects reduces biases towards the latter (Galinsky & Moskowitz, 2000), only a handful of studies have focused on the ways to establish links between the self and an object (see Phills, Kawakami, Tabi, Nadolny, & Inzlicht, 2011 for an example of approach-avoidance manipulation). By relying on an associative procedure that works through intersecting regularities between the self and targeted objects, the SR task can be also viewed as a novel intervention strategy to strengthen the perceived relation between the self and an object in a way that differs from all previous attempts. It can be argued that once individuals have learnt that regularities intersect the self and an object in the SR task, the level of

closeness or identification with the target object increases, following the same principles of Greenwald et al.'s (2002) balanced triads. As far as we are aware, there is only one published study showing that the SR task can lead to increased identification with the target object (Richetin, Mattavelli, & Perugini, 2016).

**A meta-analysis of SR studies**

In standard meta-analyses, the objective is to depict a quantitative review of the entire existing literature (considering data from all the available published and unpublished studies) in a given field. Specifically, when a meta-analysis is performed on a given set of studies, it serves a) to highlight effects that might deserve to be re-examined in future studies and b) to inspire new predictions and theoretical developments which require additional data and it can be the starting point for examining possible confounds or alternative theoretical explanations (Lakens, Hilgard, & Staaks, 2016). The main aim of the present chapter is to provide an exhaustive record of all the empirical evidence of the effect of the SR task gathered by our research group. For this reason, we intentionally excluded from this meta-analysis any published (e.g., Perkins & Forehand, 2012) or unpublished study conducted by external researchers. Although this selection could restrict the generalization of the conclusion on the effect of the SR task, a single-lab meta-analysis has an important value. Specifically, in such a type of meta-analysis the overall effect size estimates is unbiased because all results are analyzed. This issue of biased effect size estimates in meta-analysis is of increasing concern (Dwan, Gamble, Williamson, & Kirkham, 2013). A meta-analysis of a full single-lab disclosure instead produces unbiased estimates that are not affected by common threats such as publication bias or file-drawer bias.

We first aim at showing the overall effect of the SR effect across studies on both evaluative (i.e., implicit and explicit liking) and extra-evaluative (i.e., identification) domains. Because the studies under investigation differ in their characteristics, a heterogeneous overall effect was expected, hence with the need to investigate potential boundary conditions. The meta-analysis method is useful in structuring results and particularly suitable for three main reasons. First, it provides the empirical

basis for resolving uncertainties about different interpretations of findings observed in isolated studies. Second, it has a direct impact on future research: researchers interested in administering the SR task can take advantage of any information related to potential variables affecting the magnitude of the effect and thus can be more accurate in their predictions. Finally, single studies often prevent researchers from drawing comparisons among subgroups, especially if some of them suffers from restricted sample size. For example, in the SR studies, examining the effect of memory of the IR between the self and the target is statistically inappropriate in a single study given the small size of the incorrect memory sample. By taking into account a large pool of studies, the meta-analysis circumvents this problem and allows the researcher to look at the impact of such variables from a broader and a more reliable perspective. We organized and examined the potential moderators according to the heuristic framework for learning research (De Houwer, 2009a). We coded the SR studies' features on the basis of 'abstract' (general) and 'concrete' (specific) aspects that are supposed to have an impact on learning. Factors not directly related with the SR task itself were also included as potential moderators. Taken together, results provide important indications about the robustness of the SR effect on implicit, explicit attitudes and identification, and allow us to identify the most crucial aspects in qualifying the magnitude of the overall effect.

**Method**

The chapter involves all studies using the SR task that we ran from January 2006 until December 2015 (including undergraduate projects), with the exception of only three studies due to their special design. In fact, two studies focused on the persistence of the SR effect on attitude change after either an extinction or a counterconditioning procedure, while another one introduced a Matching-To-Sample procedure (Dymond & Whelan, 2010) between the SR task and the attitude measures. For all three cases, on a priori ground we reasoned that the additional manipulations

rendered the studies inadequate for estimating a mean overall effect.[4] This preliminary selection led to an overall amount of fifty-three studies, which represents the basis of our meta-analytical investigation. Twelve of these fifty-three studies have been published (Perugini, Richetin, & Zogmaister, 2012; 2014; Perugini et al., 2013; Prestwich et al., 2010; Richetin, Mattavelli, & Perugini, 2016). Among the unpublished studies, five are included into two manuscripts that are currently under review, other ten are in three manuscripts in preparation, and the remaining twenty-six studies are simply unpublished[5].

All included studies were experimental and assessed the effect of one main factor (SR condition: self+target1 vs. self+target2) on at least two dependent variables, measuring the effect on attitude at implicit and explicit levels (only one study tested the effect at the explicit level).

**General and specific moderators**

Following the heuristic framework (De Houwer, 2009a), the general aspects pertain to the statistical properties of the stimulus-response relationships involved in the task and all the possible variations in their administration and performance. On the other hand, specific aspects involve the implementation of these relationships. We considered characteristics related to the type of stimuli (and the relationship between them), the type of behavioral response required to categorize them and SR task's contextual factors as specific moderators. We also included an additional category including aspects unrelated to the SR paradigm, i.e., participants' gender, the characteristics of the

---

[4] The inclusion of the three studies ($K = 56$) in the standard meta-analysis does not affect the overall mean effect of the SR task observed in the main analyses ($K = 53$) as shown by the trivial difference in Cohen's $d$ estimators for implicit ($\Delta = 0.007$) and explicit ($\Delta = 0.003$) attitude change.

[5] Note that the overall mean effect of the SR task on implicit and explicit attitude is not affected by whether the studies have been published, as suggested by the non-significant moderations when considering 'study publication' as a factor (implicit attitude: $Q_B(1) = .188$, $p = .664$; explicit attitude: $Q_B(1) = .002$, $p = .968$).

outcome variables, and the type of target-familiarization task administered before the SR task. The general aspects encompass two main moderators (see Table 1). For the number of trials involved in the SR, we distinguished between the standard task with 160 trials (four blocks with 40 trials each) from those with a reduced number of trials. This allowed seeing whether the SR effect varied as a function of the number of times each stimulus response combination is operationalized and thus the number of times the source (self) and the target (e.g., A) intersect through these combinations. If the number of intersections between stimulus-response contingencies matters, then the higher the number of trials, the stronger the effect. Furthermore, we considered the percentage of errors made in the task as a moderator. In a single SR block of 40 trials, participants should provide one response (e.g., pressing the E key) whenever either self-stimuli (10 trials) or target A stimuli (10 trials) appear on screen. We considered the percentage of errors in the SR trials as an indicator of participants' ability to apply the IR principle when categorizing the stimuli. Participants' average error rate was thus treated as a continuous indicator of the impact of participant's performance on the distinct outcome measures.

Specific characteristics were sorted in three main classes (see Table 1). The nature of the stimuli and the type of behavioral response used to categorize them fall into the first and the second class of specific moderators, respectively. Regarding the sources stimuli, we tested the impact of using either personally irrelevant (i.e., self/other pronouns) or relevant (i.e., participants' names or self-describing personality traits) stimuli to define the self and the contrast source 'others' separately, hypothesizing that personally relevant stimuli might increase the self-reference and thus the SR effect. We checked for any effect driven by the relation between sources valence, comparing the standard procedure where stimuli related to the self are normatively positive and those related to the contrast source are neutral with any different type of valence relation (e.g., both positive, both negative, etc.). In line with previous research showing that the role of the others category valence and thus its relation with the self category valence does not impact the effect (Perugini, Zogmaister, Richetin, Prestwich, & Hurling, 2013), we did not expect any substantial moderation effect.

Table 1. *Coded Studies Characteristics*

| Variable | Definition | Coding levels |
|---|---|---|
| **Specific Properties of the SR task** | | |
| **a. Nature of the Stimuli** | | |
| Main source | What kind of main source is used? | 1 = Self-related pronouns<br>2 = Idiosyncratic stimuli (e.g., first name, personality adjectives) |
| Contrast source | What kind of contrast source is used? | 1 = Other-related pronouns<br>2 = Different non-self stimuli |
| Relationship between source valence | What is the relationship between sources valence? | 1 = Main Source = Positive vs. Contrast Source = Neutral<br><br>2 = Other relationships |
| Target prior knowledge | Is the target know or unknown? | 1 = No<br><br>2 = Yes |
| Target type | What type of unknown target is used? | 1 = Groups<br>2 = Brands |
| Relation between target valence | What is the relationship between targets valence? | 1 = Neutral – Neutral<br><br>2 = Other relationships |
| **b. Nature of the Response** | | |
| Type of behavioral response | What behavioral response is to categorize stimuli? | 1 = Joystick left-right movement<br><br>2 = Keyboard pressing<br>3 = Colored response box pressing |
| **c. Contextual factors** | | |
| Time of IR memory question | If present, does IR memory precede or follow the measures? | 1 = After measures<br><br>2 = Before measures |
| Type of control task | What type of task is performed in the control condition? | 1 = Same as the SR but opposite relations<br>2 = Different task |
| **General Properties of the SR task** | | |
| Number of trials | How many stimulus-response trials are in the SR? | 1 = less than 160 trials<br><br>2 = 160 trials |
| Number of errors in the SR | How many errors did participants make throughout the SR completion? | Continuous |
| **Properties not related with the SR task** | | |
| **a. Nature of the individuals** | | |
| Gender | Is the participants male or female? | 1 = male<br>2 = female |

Table 1. (*Continued*)

| Variable | Definition | Coding levels |
|---|---|---|
| **a. Nature of the Evaluative Measurement Procedure** | | |
| Indirect measure | What type of indirect measure is used? | 1 = IAT-like task (IAT, brief IAT, ST IAT)<br>2 = AMP |
| Direct measure | What type of direct measure is used? | 1 = Semantic differential<br>2 = Olson & Fazio's liking measure or others. |
| Pre-post measure | Does the study involve a pre-post attitude measure? | 1 = Yes<br><br>2 = No |
| Time of measurement | Is attitude measured right after the paradigm? | 1 = Yes<br><br>2 = No (preceded by either a measure or a manipulation) |
| Familiarization task | Is the SR preceded by familiarization trials? | 1 = No task<br><br>2 = Brief categorization task (with different behavioral response)<br>3 = Standard categorization task |

Target stimuli were qualified into two main variables. Target previous knowledge distinguished between known objects and those never encountered before the SR task. Moreover, given that most of our SR studies ($K = 50$) used either groups or brands as target object, on this subset of studies we tested whether the effect differed due to the type of target. Another variable accounted for variations in the effect of the SR task when both target stimuli were neutral rather than when at least one of them was not (e.g., positive vs. negative, both positive) or when only one object was targeted without any contrast category. The effect of the behavioral response required to intersect sources and target was inspected by distinguishing between three distinct actions (i.e., right-left joystick movement, keyboard keys pressing, response box colored key pressing). Two additional contextual aspects were included in the specific moderators. First, given the hypothesized central role played by IR memory, we checked for the impact of administering the IR memory right after the SR task or at the end of the study. The last specific moderator was the type of task participants performed in the control condition.

We differentiated between a task that was perfectly symmetrical to the SR task (with participants categorizing self-stimuli with the target paired with the category 'other' in the SR task) from the condition in which a control task was either absent or did not involve any intersecting regularities between stimuli.

Among the moderators that were unrelated to the task, the first pertained to participants' characteristics. Because all studies but one involved a population of university students, only gender was considered and coded as the proportion of female participants. In the second subclass of the unrelated moderators were all the characteristics related to the measures used to assess the SR effect. As said before, one of the main limits of previous studies testing the effect of the self in increasing liking towards objects lies at the measurement level. First, many studies focused only on implicit or on explicit attitude change: the SR studies avoid this pitfall by assessing liking at both implicit and explicit level. Second, within the two attitude levels (i.e., implicit and explicit) the meta-analysis allows us to test whether the use of different instruments of measure has an impact on the SR effect. Implicit liking was tested through either the Implicit Association Test (IAT, Greenwald, McGhee, & Schwarz, 1998), the Brief Implicit Association Test (BIAT, Sriram & Greenwald, 2009), the Single Target Implicit Association Test (ST-IAT, Bluemke & Friese, 2006), and the Affect Misattribution Procedure (AMP, Payne, Cheng, Govorun, & Stewart, 2005). The effect of the SR on explicit evaluative change also involved distinct instruments of measure: gut feelings measures (see Olson & Fazio, 2001), semantic differentials, and feelings thermometer. For both implicit and explicit attitude, we tested whether distinct instrument of measures qualify the SR effect. Two additional aspects related with the outcome measures were considered. First, we distinguished between those studies where the evaluation of the target was assessed prior and after the manipulation and those with a single assessment after the manipulation. Second, we inspected the effect of the time of administration of the dependent variable (i.e., either right after the manipulation or preceded by an additional task). Finally, familiarization with target stimuli prior the SR was also considered as an independent moderator. Familiarization is a categorization task where only target stimuli are used. Across studies,

familiarization had three different levels: present with two blocks of 20 trials using response keys identical to those used in the SR task (most frequent), present but brief with 10 trials using the mouse, and absent.

**The impact of IR memory**

As said before, we assumed that the participants' ability to remember the correct intersecting regularities established throughout the SR task might be an important qualifier of the effect. In their meta-analysis of EC, Hofmann et al. (2010) considered contingency awareness as a continuous moderator variable of the EC effect because contingency awareness was not administered in all the EC studies and because there was not a univocal method to assess it. Responses to the standard IR memory question distinguish among three distinct groups of participants (i.e., correct, incorrect, and without IR memory). Therefore, for each study assessing participants' IR memory ($K = 44$) we dummy coded responses and considered the factor as a potential moderator of the SR effect with potential variations at the study-level.

**Meta-analytic computation**

The meta-analysis was conducted employing the original datasets of the 53 studies, using participant-level data. A series of mixed linear models were estimated to assess the effects of the SR condition on the experimental outcomes (implicit, explicit, and identification), its differential effect across outcomes, and, in subsequent estimations, the effects of the moderators. The use of subject-level data allowed estimating the effects of the independent variables and their moderators while taking into account the dependency among data, at both the individual and the study levels. At the individual level, the estimated models took into account the covariance among the experimental outcomes and possible sources of variation due to individual differences across participants. At the study level, the models captured possible heterogeneity within samples and study idiosyncrasy.

To make the comparison between explicit and implicit outcomes interpretable, participants' outcomes were standardized within each study. This choice has the advantage of removing spurious effects due to the interplay between the individual and study level (Raudenbush & Bryk, 2002), eases

the interpretation of the model coefficients in terms of standardized scale, and simplifies the estimation because it guarantees a simpler random structure, particularly with regards of the random intercepts. After standardization, the outcomes have been treated as a single variable score. For each participant, the information concerning which score corresponded to each dependent variable (implicit, explicit, or identification) has been encoded into a set of dummy variables (hereinafter, outcome level). In this way, the interaction between any experimental independent variable or moderator and type of outcome signals whether there is a differential effect across outcomes. Furthermore, by opportunely setting the reference level of the type of outcome variable (Cohen et al. 2013), we can estimate the effect of the independent variables on each outcome separately, while still taking into account the covariance among them. The approach we used is akin to other estimation methods used for multivariate multilevel (mixed) models (e.g., MacCallum, Kim, Malarkey, & Kiecolt-Glaser, 1997) and random coefficients mediation analysis (Bauer, Preacher, & Gil, 2006). The technique of slicing the effect of the independent variable for each experimental outcome level is equivalent to the simple effects analysis, a standard method within the linear models application (e.g. Tabachnick, & Fidell, 2007).

**Analyses plan**

We first estimated a mixed model with the outcome score as dependent variable, SR condition, type of variable, and their interaction as predictors. Given that only fourteen studies tested the SR effect on the identification measure, we ran the analyses in two steps, first focusing on the comparison between implicit and explicit attitude and second including identification. We then estimated the same effects for each outcome level. We added moderators in subsequent analyses. In all models, the clustering variables were participants and studies. Keeping with modern treatments of mixed models (Barr, Levy, Sheepers, & Tily, 2013), we estimated each model by first establishing the maximal random structure, which is the model with the largest possible number of random coefficients. When data did not support a full random structure, we ensured that the crucial fixed effects were tested over their random distribution. This guarantees a proper Type I error rate (Barr et al. 2013). Thanks to the

standardization of the outcome variables, intercepts across studies were fixed as they are all equal to zero.

In order to provide important indications related to the effect size of the SR task on all the outcome levels, we also meta-analyzed data at the study level following a standard meta-analysis procedure. The assessment of a SR effect primarily involved a comparison between two means. For this reason, we chose the standardized mean difference (Cohen's $d$) as the effect size statistic and, as recommended by Lipsey and Wilson (2001), we adjusted effect sizes for bias in small samples using the correction formula proposed by Hedges (1981). Effect sizes were computed from group means, standard deviations, and sample size.

Finally, for the studies included in the meta-analyses with all participants on the three dependent variables we inspected the evidential value of our findings through a p-curve analysis (Simonsohn, Nelson, & Simmons, 2014a; 2014b) to test the distribution of significant results ($p <$ .05) among the studies.

**Results**

The raw data were analyzed using R (R Development Core Team, 2013). For the mixed-models analysis, we used the lme4 and lmerTest packages to perform our mixed-effect model analysis in R (Bates, Maechler, & Bolker, 2014).

**The SR task on implicit and explicit attitude**

The sample included 53 studies ($N = 6113$). At the first step, we tested whether the SR affected attitudinal change and whether this hypothesized effect did vary for implicit and explicit attitude. The SR manipulation had a largely significant impact on attitude change, $\beta = .43$, $SE = .02$, $p < .001$. Moreover, we observed a significant interaction between SR manipulation and the attitude level at which the effect was estimated, $\beta = .21$, $SE = .04$, $p < .001$. Looking separately at the effect of the manipulation on implicit and explicit attitude change, the effect was significantly larger on implicit than on explicit attitude, $\beta = .53$, $SE = .02$, $p < .001$ and $\beta = .33$, $SE = .02$, $p < .001$, respectively. Target objects categorized through the same action as the self in the SR task are the better liked than

those categorized with a contrast attitudinal source (e.g., 'Others'). This effect showed reliably on both implicit and explicit attitude change, although it was significantly stronger at the implicit level.

**Is the SR task effect moderated by specific aspects of the task?**

*The role of source stimuli.* This analysis tested whether the effect of SR varied as a function of the type of stimuli used to represent a) the category 'self' (i.e., the main source: personal pronouns vs. self-relevant personality adjectives or names), b) the category 'other' (i.e., the contrast source: personal pronouns vs. self-irrelevant personality adjectives or someone else's name), and c) the relation between the main source and the contrast source with respect to their valence. For theoretical reasons, the three factors were included simultaneously in the model. For all three variables, we tested their main effects and their interactions with the SR manipulation and with the type of outcome. Neither the type of main source, nor the type of contrast source, nor their relation in valence showed a main effect on attitude change ($ps > .868$). Importantly, whilst the type of contrast source and source valence relation did not interact with the SR manipulation ($\beta = -.12$, $SE = .10$, $p = .208$ and $\beta = .02$, $SE = .10$, $p = .836$, respectively), the type of main source did significantly ($\beta = .38$, $SE = .09$, $p < .001$). The SR effect was stronger when self-relevant personality traits or names ($\beta = .75$, $SE = .08$, $p < .001$) were used as main source in the SR task, rather than personal pronouns ($\beta = .37$, $SE = .03$, $p < .001$). The lack of significant three-way interaction when the type of attitude level was included ($ps > .226$) showed that the main source qualified the effect of the SR manipulation both on implicit and explicit attitude.

*The role of target stimuli.* We tested whether specific characteristics of the target had any impact on the SR effect. At a first step, we included target previous knowledge (i.e., known vs. unknown target objects) and the relation between the two targets used in the task in terms of their valence (i.e., both targets are neutral vs. one of the two is not neutral). The two factors were entered simultaneously as fixed factors. First, the model confirmed the significant effect of the SR manipulation ($\beta = .35$, $SE = .01$, $p < .001$) as well as its interaction with the attitudinal level ($\beta = .20$, $SE = .06$, $p < .001$) and revealed no main effect of the two target-related variables ($ps > .911$).

Moreover, the model showed an interaction between the SR manipulation and target knowledge ($\beta$ = -.28, $SE$ = .07, $p$ < .001) and a three-way interaction between SR, target relations in terms of valence, and the attitudinal level ($\beta$ = .26, $SE$ = .12, $p$ = .028). Decomposing the first interaction term shows that the SR effect was significant on both unknown and known objects, though to a stronger extent for the former ($\beta$ = .49, $SE$ = .03, $p$ < .001 and $\beta$ = .20, $SE$ = .05, $p$ < .001). The second interaction indicated that when at least one of the two targets is not neutral, the effect of the SR decreased on explicit (effect with two neutral targets: $\beta$ = .36, $SE$ = .03, $p$ < .001; effect on at least one non-neutral target: $\beta$ = .20, $SE$ = .07, $p$ = .010), without affecting implicit attitude ($\beta$ = .54, $SE$ = .04, $p$ < .001 and $\beta$ = .60, $SE$ = .09, $p$ < .001 respectively).

Since most of the studies ($K$ = 50) used either groups or brands as target objects, we tested the impact of target type on this subsample, by entering target type as fixed factor in the model. We did not observe an interaction between the SR manipulation and the type of target ($\beta$ = -.04, $SE$ = .04, $p$ = .305). However, a significant three-way interaction ($\beta$ = .27, $SE$ = .07, $p$ < .001) suggested that the SR effect differed on groups and brands depending on whether attitude change was measured implicitly or explicitly. The SR effect on explicit attitude was stronger on groups than brands ($\beta$ = .40, $SE$ = .04, $p$ < .001 and $\beta$ = .30, $SE$ = .03, $p$ < .001 respectively), while the opposite pattern hold for implicit attitude ($\beta$ = .46, $SE$ = .04, $p$ < .001 and $\beta$ = .63, $SE$ = .03, $p$ < .001 respectively).

***The type of action performed.*** We tested whether the effect of the SR manipulation was moderated by the type of action required to categorize stimuli in the task. The analysis confirmed the significant effect of SR and its interaction with attitude level ($\beta$ = .45, $SE$ = .02, $p$ < .001 and $\beta$ = .19, $SE$ = .04, $p$ < .001, respectively) while none of the other main effect or interaction was significant ($p$ > .076). This effect of the SR manipulation thus is not affected by the type of action performed to categorize stimuli.

***Contextual factors.*** We tested for the impact of the SR on attitude change depending on a) the type of task participants performed in the control condition and b) whether the IR memory question was administered either prior after the attitude measure. Two separate linear mixed-models

were conducted, one for each potential moderator. The inclusion of the type of control task (i.e., symmetrical version of the SR vs. absent or different task) in the model resulted in a main effect of the SR manipulation ($\beta = .38$, $SE = .03$, $p < .001$), in no main effect of the type of control task ($\beta = .00$, $SE = .03$, $p = .931$), and in no significant interaction between the two factors ($\beta = -.11$, $SE = .07$, $p = .101$). None of the other effects involving the type of control task reached the significance level ($p$s $> .213$). This clarifies that the SR has the power to affect attitudes regardless of the type of task performed in the control condition. The same analysis was conducted to test the impact of the IR memory question assessed either prior or after the attitude measures. None of the parameters, other than that estimating the main effect of the SR and its interaction with attitude level, were significant predictors of attitudinal change ($p > .087$). Although non-significant, the three-way interaction ($\beta = .12$, $SE = .07$, $p = .088$) suggested that measuring IR memory before the measures tends to amplify the effect on implicit, but not explicit, attitude change.

**Is the SR task effect moderated by general aspects of the task?**

We first focused on whether the number of errors performed by participants during the completion of the task qualified the SR effect. The analysis was conducted on forty-nine studies, since for the remaining four studies the relevant information was not available. Given that we used versions of the SR task that differed in their total number of trials, the number of errors performed during the SR task by each participant was reported as the overall percentage of incorrect trial-responses ($M = 6.27$, $SD = 5.45$). The significant interaction between SR manipulation and the number of errors ($\beta = -.02$, $SE = .003$, $p < .001$) revealed that the effect was stronger when participants performed less errors in categorizing stimuli (i.e., when the IR principle was correctly applied). The analyses conducted with the number of SR trials (considering all the studies) as potential moderator showed neither a main effect ($\beta = .00$, $SE = .04$, $p = .976$) nor an interaction of this factor with the SR manipulation ($\beta = -.10$, $SE = .09$, $p = .257$). This indicates that the SR effect depends on one's ability to correctly apply the IR principle throughout the task (i.e., the number of errors), but not on the number of times participants categorize source and target stimuli.

**Is the effect moderated by factors that are independent from the SR task?**

We investigated the impact of either individual characteristics or aspect related to the design and the procedure adopted in the studies on the overall effect. Gender was the only individual characteristic we took into account. We included the variable in the model as fixed and random factor (at the level of the study), conducting linear-mixed models on fifty-two studies (no gender information was available in one of the studies). We found a main effect of gender ($\beta = .05$, $SE = .02$, $p = .011$). More relevant for the purpose of the current analysis, its interaction with the SR manipulation was not significant ($\beta = -.06$, $SE = .04$, $p = .136$) neither was the three-way interaction with the outcome level included. Therefore, on both implicit and explicit attitude change, the effect of the SR was not affected by the gender of the participant.

The type of categorization task through which participants get familiar with the stimuli was included as a moderator. We discriminated between studies with a standard version of the task (i.e., 20 trials), a brief version (e.g., 10 trials), and no familiarization task. We found no main effect of the type of familiarization task ($\beta = -.002$, $SE = .01$, $p = .838$), nor an interaction with the SR manipulation ($\beta = .00$, $SE = .03$, $p = .954$).

Next, we focused on study design. The analysis revealed no main effect of study design ($\beta = .001$, $SE = .03$, $p = .952$) and a significant interaction with the SR manipulation ($\beta = -.14$, $SE = .05$, $p = .009$). The effect of the SR manipulation was stronger in only-between, rather than mixed (within-between), designs ($\beta = .45$, $SE = .02$, $p < .001$ and $\beta = .31$, $SE = .08$, $p < .001$, respectively). This finding seems to suggest that when attitudes were assessed only after the manipulation, the effect of the SR was stronger than when they were assessed before and after the manipulation.

*Measurements characteristics.* The last three moderators pertained to the attitudinal measures. First, we inspected whether administering the outcome measure right after the SR manipulation or preceded by an additional task qualified the effect. The analysis conducted with time of administration of the measure as fixed factor confirmed the main effect of the SR manipulation and its interaction with the attitudinal outcome. More central to our concerns, there was no main

effect of the hypothesized moderator (β = -.002, *SE* = .02, *p* = .917) and no significant interaction with the SR task (β = -.02, *SE* = .05, *p* = .690). For both implicit and explicit attitude, we tested whether the type of measure used to assess attitude change mattered. First, we focused on the type of implicit measure. Given that in most of our SR studies implicit attitude was measured before explicit, we maintained attitude level as fixed factor in the model, to see whether the type of implicit measure affected either implicit or explicit attitude, or both. SR manipulation, the attitude level, the type of measure of implicit attitude (i.e., IAT vs. AMP) and their interactions were included in the model as fixed effects. The model revealed a main effect of SR manipulation (β = .38, *SE* = .02, *p* < .001), while its interaction with the outcome level (i.e., implicit vs. explicit) was not significant (β = .02, *SE* = .07, *p* = .736). There was no main effect of the type of implicit attitude measure (β = .00, *SE* = .04, *p* = .905), no interaction with the SR manipulation (β = -.14, *SE* = .11, *p* = .217), but a significant three-way interaction with SR manipulation and outcome level (β = -.43, *SE* = .14, *p* = .004). The type of implicit attitude measure qualified the SR effect on implicit, but not explicit, attitude change. On implicit attitude the effect of the SR on an IAT-like measure was significant, β = .59, *SE* = .04, *p* < .001, but it was not on the AMP, β = .21, *SE* = .14, *p* = .135. To test the effect of the explicit attitude measure, we focused exclusively on the explicit level and included the SR manipulation, the type of explicit attitude measure, and their interaction as fixed factors. There was a significant SR effect (β = .24, *SE* = .04, *p* < .001), no main effect of the type of explicit attitude measure (β = .00, *SE* = .04, *p* = .994), and a significant interaction (β = -.22, *SE* = .08, *p* = .008). There was a significant effect on the semantic differential (β = .24, *SE* = .04, *p* < .001) but not on the other types of explicit attitude measures (β = .13, *SE* = .08, *p* = .109).

### What is the impact of IR memory on attitude change?

As we said in the introduction, IR memory is meant to play an important role in the SR effect. This variable was included in a mixed model analysis considering forty-four studies (*N* = 5488) in which an IR memory question was administered either after the SR task or at the end of the experimental session. IR memory was included as a fixed factor to investigate a main effect and an

interaction with the SR manipulation. We also maintained outcome level as a fixed factor to test whether any hypothesized variation in the SR effect due to participants' memory was also qualified by the level (implicit vs. explicit) at which attitude was measured (i.e., testing a three way interaction). With the inclusion of IR memory, the model showed a significant main effect of the SR manipulation ($\beta = .43$, $SE = .02$, $p < .001$), no main effect of the outcome level ($\beta = -.002$, $SE = .02$, $p = .919$), and no main effect of IR memory ($\beta = -.02$, $SE = .02$, $p = .538$). The interaction between SR manipulation and attitude level was significant ($\beta = .20$, $SE = .04$, $p < .001$) and, crucially, so was the interaction between the SR and IR memory ($\beta = .21$, $SE = .04$, $p < .001$). We found no interaction between the outcome level and IR memory ($\beta = -.03$, $SE = .04$, $p = .414$) nor between the three variables ($\beta = -.03$, $SE = .08$, $p = .691$). Specifically, the SR effect on participants with correct IR memory was significant, $\beta = .51$, $SE = .02$, $p < .001$, decreased on participants with no IR memory, $\beta = .32$, $SE = .06$, $p < .001$, and was not significant (and negative) for participants with incorrect IR memory, $\beta = -.08$, $SE = .06$, $p = .167$. Participants' ability to recall the correct intersecting regularities after the SR task qualified the SR effect irrespective of whether attitude change was measured at the implicit or explicit level.

**The SR task effect on identification.**

The final set of meta-analytical computations focused on the level of identification with the target paired with the self. A target-tailored version of the Inclusion of Other in the Self scale (IOS, Aron, Aron, & Smollan, 1992) was used as identification measure for all primary studies. In a single item, a series of seven pairs of Eulero Venn diagrams was used, one representing the self and the other representing the target object at issue. The participants choose the pair of circles that best represented their level of identification with each target object separately. The difference between the two scores was the indicator of relative identification with the focal target object. The analyses considered a total of fourteen studies ($N = 1336$). We first estimated the general effect of the SR manipulation. We observed a significant impact of the SR on the level of identification with the target object paired with the self, $\beta = .32$, $SE = .10$, $p = .005$. Besides changing implicit and explicit attitude,

the SR paradigm is thus a reliable task to increase one's level of identification with target objects categorized with the self.

**What factors moderated the SR task effect on identification?**

Given the limited number of studies, we included in the analyses only moderators with at least three studies for each level. First, we focused on the specific characteristics to see whether the effect of SR varied as a function of the type of stimuli used to represent the 'self', the category 'other', and the relation between the main and the contrast sources with respect to their valence. As we did for attitude, the three factors were included simultaneously in the model. Their main effect and their interaction with the SR manipulation were fixed factors. None of the three source-variables showed a significant main effect ($p$s > .867), neither the type of contrast source nor the relation of the sources in terms of their valence qualified the SR effect, ($\beta$ = -.45, $SE$ = .41, $p$ = .304 and $\beta$ = .32, $SE$ = .33, $p$ = .355, respectively). Although the interaction between the type of self-stimuli and the SR manipulation did not reach the conventional level of significance ($\beta$ = .67, $SE$ = .31, $p$ = .067), we decomposed the interaction and found a stronger SR effect when using personally relevant stimuli compared to standard pronouns ($\beta$ = .73, $SE$ = .15, $p$ < .001 and $\beta$ = .21, $SE$ = .08, $p$ = .032, respectively). This finding is consistent with what we observed on implicit and explicit attitude. The use of personally relevant stimuli tends to increase the effect of the SR task. The other two moderators considered were gender and the number of errors performed during the SR task. With the inclusion of gender, the effect of the manipulation remained significant ($\beta$ = .33, $SE$ = .10, $p$ = .005), there was a marginally significant main gender effect ($\beta$ = .11, $SE$ = .06, $p$ = .049) but no significant interaction ($\beta$ = .01, $SE$ = .12, $p$ = .932). The same analysis was conducted with the number of errors performed in the task. The effect of the SR effect remained significant ($\beta$ = .39, $SE$ = .11, $p$ = .004), but there was no main effect of the number of errors nor an interaction with the SR manipulation ($\beta$ = .004, $SE$ = .004, $p$ = .371 and $\beta$ = -.002, $SE$ = .09, $p$ = .784, respectively).

**What is the impact of IR memory on identification?**

As we did for the effect on attitude, we also estimated the impact of IR memory on the effect of the SR on identification. Given the restricted number of studies, the number of participants in both the incorrect and no IR memory groups was very small ($N = 123$ and $N = 78$, respectively). Therefore, we merged these two IR memory subgroups and compared the effect of the SR manipulation between this new group and the one with correct IR memory ($N = 975$). We used a mixed model analysis, with SR manipulation, IR memory, and the interaction as fixed factors. In addition, IR memory was also included in the model as random factor at the study level. Results showed a main effect of the SR manipulation ($\beta = .31$, $SE = .08$, $p < .001$) and a main effect of IR memory ($\beta = .17$, $SE = .08$, $p = .022$). More central to our concerns, we found a significant interaction term ($\beta = .33$, $SE = .15$, $p = .032$). As we found for attitude change, participants' memory of the correct intersecting regularities moderated the impact of the manipulation on identification. The effect was significant for correct IR memory participants, $\beta = .47$, $SE = .06$, $p < .001$, while it was not significant on the incorrect group, $\beta = .15$, $SE = .14$, $p = .296$.

**Results from a standard meta-analytical approach: Effect size estimations and recommendations for SR research.**

In addition to the mixed model analyses, we ran a series of standard meta-analyses of the 53 studies to provide information about the effect size of the SR at each outcome level and for the three IR memory subgroups. Moreover, we use the results to provide recommendations related to the sample size required to detect a reliable SR effect in future research.

For the effect on implicit attitude, the random-effects model yielded a mean estimated effect size ($d$) of 0.590, $SE = 0.041$, 95% CI [0.509, 0.671]. The minimum and maximum study effect sizes were -0.018 and 2.086, respectively (see Supplementary Materials Chapter 1 – Appendix B for the forest plots on the three dependent variables considering all participants). Three independent meta-analyses showed a significant strong SR summary effect size for participants with correct IR memory ($K = 44$), $d = 0.685$, $SE = 0.045$, 95% CI [0.597, 0.773], a non-significant effect for participants with incorrect IR memory ($K = 32$), $d = 0.050$, $SE = .087$, 95% CI [-0.120, 0.221], and a significant effect

for participants with no memory of the intersecting regularities ($K = 30$), $d = 0.356$, $SE = 0.090$, 95% CI [0.180, 0.532]. The overall meta-analysis on explicit attitude involved 52 studies. The mean effect size ($d$) was 0.347, $SE = 0.027$, 95% CI [0.294, 0.400]. Compared to the overall analysis, we detected an increased effect size for participants with correct IR memory ($K = 44$), $d = 0.428$, $SE = 0.039$, 95% CI [0.350, 0.505]. The meta-analysis of SR on explicit attitude including participants with no IR memory ($K = 30$) lead to a summary effect, $d = 0.325$, $SE = 0.090$, 95% CI [0.148, 0. 503]. Interestingly, the meta-analysis conducted on participants with incorrect IR ($K = 32$) showed a negative summary effect, $d = -0.205$, $SE = 0.087$, 95% CI [-0.377, -0.034]. The mean effect size for identification was $d = 0.351$, $SE = 0.088$, 95% CI [0.179, 0.524]. The effect also increased for those participants with correct IR memory ($K = 12$), $d = 0.444$, $SE = 0.09$, 95% CI [0.260, 0.629]. The meta-analysis conducted on participants with incorrect memory of the intersecting regularities ($K = 9$) showed a negative, yet not significant, summary effect, $d = -0.065$, $SE = 0.297$, 95% CI [-0.647, 0.518]. In a similar way, the effect of the SR on identification was not significant for participants with no IR memory ($K = 8$), $d = -0.022$, $SE = .215$, 95% CI [-0.444, 0.400].

A power analysis conducted with G-Power and based on the overall effect revealed that with an overall sample of 74 participants, researchers can detect the effect of the SR on implicit attitude with a power ($1 – \beta$) of .81. Nevertheless, it is important to bear in mind that we estimated the effect size from studies where implicit attitude was assessed with an IAT. The effect might decrease with other measures of implicit attitudes (e.g., AMP). Given that the overall effect sizes on explicit attitude and identification were almost identical, a power analysis revealed that with an overall sample of 200 participants, researchers could detect the effect of the SR on both the outcome measures, with a power ($1 – \beta$) of .80. Given the effect is stronger on the semantic differential ($d = 0.375$), thus a sample of 174 participants would suffice in case that measure is used. The required sample sizes decrease substantially when considering participants with correct IR memory (as the effect sizes increase). Specifically, 54 participants would suffice to have a power of .80 on implicit attitude and 130 participants on explicit attitude and identification. These numbers can be further pushed down (or

power can be increased) by focusing on conditions that make the SR effect stronger. Indeed, from a more general perspective, although information about overall effect sizes is important for future studies, researchers should be aware of the relevant role played by many characteristics that might be either related to the SR task or to the experimental design in qualifying the magnitude of the SR effect. To this aim, we encourage researchers to take into account both the analyses presented in the previous paragraphs and the summary results of the effect of the SR based on distinct levels of moderators (see Tables 2 and 3).

**Evidential value tests**

The "*p*-curve" method involves examining the skewness of the frequency distribution of statistically significant *p* values ($p < .05$) for a set of research findings (Simonsohn et al., 2014a). If no true effect exists, and no p-hacking has occurred, a *p*-curve will be uniform, with equal frequencies of p values between .04 and .05, .03 and .04, .02 and .03, and so on. A left-skewed p-curve would indicate a relatively high frequency of *p* values closer to .05 that the effect is actually produced by *p*-hacking. Conversely, a right skewed *p*-curve with relatively high frequency of *p* values close to 0 would indicate no evidence of lack of evidential value in the nonzero true effect. The *p*-curve analyses conducted on the present findings reveal that on each dependent variable (i.e., implicit attitude, explicit attitude and identification) there is evidential value (right-skewed, all *p*s $< .032$) and no evidence of lack of evidential value and that the observed p-curves are not flatter than the null at 33% power (all ps $> .990$), so we can rule out the hypothesis that the evidential values are inadequate (see Supplementary Materials Chapter 1 – Appendix A for the p-curve graphs for each dependent variable). Along with a p-curve analyses, we also inspected any bias in the distribution of the estimated effect sizes for each of the outcome variable (implicit, explicit, and identification) through funnel plots (Sterne & Egger, 2001). The graphical distribution of the studies' effect sizes based on their own standard errors revealed the absence of bias in the estimation of an overall effect due to underpowered studies (see Supplementary Material Chapter 1 – Appendix B for the funnel plots for each dependent variable).

Table 2. *Moderator Analyses for the SR effect on Implicit Attitude*

| Moderator | | ES d/β | C | SE | K | $Q_B$ (df) | $Q_w$ (df) | p |
|---|---|---|---|---|---|---|---|---|
| **Main source** | | | | | | 6.35 (1) | 109.09 (51) | .012 |
| Self-related pronouns | | .561 | a | .041 | 49 | | | |
| Idiosyncratic stimuli | | .946 | b | .147 | 4 | | | |
| **Contrast source** | | | | | | .85 (1) | 119.65 (51) | .358 |
| Other-related pronouns | | .573 | a | .045 | 43 | | | |
| Different non-self stimuli | | .672 | a | .098 | 10 | | | |
| **Relationship between sources valence** | | | | | | .95 (1) | 119.52 (51) | .329 |
| Positive-Positive | | .573 | a | .045 | 44 | | | |
| Other relationships | | .683 | a | .103 | 9 | | | |
| **Target previous knowledge** | | | | | | 6.14 (1) | 112.67 (51) | .013 |
| Known targets | | .348 | a | .105 | 9 | | | |
| Unknown targets | | .629 | b | .043 | 44 | | | |
| **Target type** | | | | | | 4.37 (1) | 97.19 (49) | .037 |
| Brands | | .681 | a | .052 | 30 | | | |
| Groups | | .514 | b | .061 | 20 | | | |
| **Relation targets valence** | | | | | | .33 (1) | 122.89 (51) | .564 |
| Neutral-Neutral | | .648 | a | .045 | 45 | | | |
| Other relationships | | .580 | a | .108 | 8 | | | |
| **Type of response** | | | | | | .53 (2) | 122.11 (50) | .766 |
| Joystick movement | | .534 | a | .209 | 3 | | | |
| Keyboard pressing | | .556 | a | .069 | 19 | | | |
| Colored keys pressing | | .616 | a | .054 | 31 | | | |
| **IR memory time** | | | | | | 5.51 (1) | 106.98 (51) | .019 |
| After measures | | .500 | a | .054 | 28 | | | |
| Before measures | | .683 | b | .056 | 25 | | | |
| **Type of control task** | | | | | | .31 (1) | 122.92 (51) | .580 |
| Same as SR task | | .598 | a | .044 | 47 | | | |
| Different task | | .519 | a | .136 | 6 | | | |
| **Number of trials** | | | | | | .11 (1) | 123.24 (51) | .743 |
| Less than 160 trials | | .531 | a | .185 | 3 | | | |
| 160 trials | | .594 | a | .043 | 50 | | | |
| **Gender** | | β = .004 | | .004 | 51 | .80 (1) | 120.59 (50) | .371 |
| **Indirect measure** | | | | | | 5.14 (1) | 107.59 (51) | .023 |
| IAT-like measure | | .610 | a | .040 | 50 | | | |
| AMP | | .233 | b | .161 | 3 | | | |
| **Pre-post measures** | | | | | | 1.13 (1) | 123.41 (51) | .287 |
| Post only | | .606 | a | .044 | 48 | | | |
| Pre-post | | .461 | a | .128 | 5 | | | |
| **Time of measurement** | | | | | | .03 (1) | 123.41 (51) | .870 |
| Right after SR | | .594 | a | .046 | 44 | | | |
| Preceded by task | | .576 | a | .097 | 9 | | | |
| **Familiarization task** | | | | | | 4.85 (2) | 112.52 (50) | .088 |
| No task | | .497 | a | .074 | 17 | | | |
| Brief categorization task | | .820 | a | .128 | 5 | | | |
| Standard categoriz. task | | .597 | a | .052 | 31 | | | |

*Note. ES (d = Cohen's d* effect size estimate for categorical moderators*; β =* standardized regression coefficient for continuous moderators*); C =* contrast index: different subscripts indicate significant differences (*p* < .05) as indicated by contrasts*; SE =* standard error*; K =* number of study effect sizes for a given moderator category*; $Q_B$ =* analysis of variance (ANOVA) between groups/regression sum of squares*; $Q_w$ =* ANOVA/regression sum-of-squares residual*; p =* significance level of between-groups effect (ANOVA) or regression coefficient.

Table 3. *Moderator Analyses for the SR effect on Explicit Attitude*

| Moderator | | ES d/β | C | SE | K | $Q_B$ (df) | $Q_w$ (df) | p |
|---|---|---|---|---|---|---|---|---|
| **Main source** | | | | | | 4.91 (1) | 54.96 (50) | .027 |
| Self-related pronouns | | .331 | a | .028 | 48 | | | |
| Idiosyncratic stimuli | | .575 | b | .106 | 4 | | | |
| **Contrast source** | | | | | | .20 (1) | 60.13 (50) | .654 |
| Other-related pronouns | | .341 | a | .032 | 42 | | | |
| Different non-self stimuli | | .377 | a | .072 | 10 | | | |
| **Relationship between sources valence** | | | | | | .84 (1) | 59.36 (50) | .361 |
| Positive-Positive | | .336 | a | .031 | 43 | | | |
| Other relationships | | .411 | a | .075 | 9 | | | |
| **Target previous knowledge** | | | | | | 11.54 (1) | 48.88 (50) | .001 |
| Known targets | | .377 | a | .080 | 8 | | | |
| Unknown targets | | .090 | b | .027 | 44 | | | |
| **Target type** | | | | | | 2.74 (1) | 48.63 (47) | .098 |
| Brands | | .324 | a | .037 | 29 | | | |
| Groups | | .415 | a | .040 | 20 | | | |
| **Relations target valence** | | | | | | 4.08 (1) | 55.73 (50) | .043 |
| Neutral-Neutral | | .370 | a | .030 | 44 | | | |
| Other relationships | | .213 | b | .072 | 8 | | | |
| **Type of response** | | | | | | 2.43 (2) | 57.62 (49) | .297 |
| Joystick movement | | .445 | a | .160 | 3 | | | |
| Keyboard pressing | | .292 | a | .046 | 19 | | | |
| Colored keys pressing | | .377 | a | .037 | 30 | | | |
| **IR memory time** | | | | | | .31 (1) | 60.07 (50) | .577 |
| After measures | | .332 | a | .040 | 27 | | | |
| Before measures | | .364 | a | .042 | 25 | | | |
| **Type of control task** | | | | | | 3.66 (1) | 56.31 (50) | .056 |
| Same as SR task | | .361 | a | .029 | 47 | | | |
| Different task | | .149 | a | .107 | 5 | | | |
| **Number of trials** | | | | | | 1.28 (1) | 58.80 (50) | .257 |
| Less than 160 trials | | .200 | a | .133 | 3 | | | |
| 160 trials | | .354 | a | .029 | 49 | | | |
| **Gender** | | β = .002 | | .003 | 50 | .39 (1) | 59.81 (49) | .535 |
| **Direct measure** | | | | | | 7.28 (1) | 52.78 (50) | .007 |
| Semantic differential | | .372 | a | .028 | 44 | | | |
| Other measures | | .138 | b | .082 | 8 | | | |
| **Pre-post measures** | | | | | | 4.13 (1) | 55.87 (50) | .042 |
| Post only | | .368 | a | .029 | 47 | | | |
| Pre-post | | .192 | b | .081 | 5 | | | |
| **Time of measurement** | | | | | | .24 (1) | 60.07 (50) | .626 |
| Right after SR | | .354 | a | .032 | 43 | | | |
| Preceded by task | | .318 | a | .066 | 9 | | | |
| **Familiarization task** | | | | | | 23.42 (2) | 37.08 (49) | <.001 |
| No task | | .134 | a | .051 | 16 | | | |
| Brief categorization task | | .456 | b | .083 | 5 | | | |
| Standard categoriz. task | | .415 | b | .032 | 31 | | | |

*Note. ES (d = Cohen's d* effect size estimate for categorical moderators*;* β *=* standardized regression coefficient for continuous moderators)*; C =* contrast index: different subscripts indicate significant differences (*p* < .05) as indicated by contrasts*; SE =* standard error*; K =* number of study effect sizes for a given moderator category*; $Q_B$ =* analysis of variance (ANOVA) between groups/regression sum of squares*; $Q_w$ =* ANOVA/regression sum-of-squares residual*; p =* significance level of between-groups effect (ANOVA) or regression coefficient.

**Discussion**

Across 53 studies, our meta-analytical findings showed that the SR is a genuine and robust evaluative learning effect on both implicit and explicit attitude as well as on identification. Distinct from previous classes of evaluative learning effect, the SR takes advantage of the general positivity of one's self concept (Bosson, Swann, & Pennebaker, 2000; Yamaguchi et al., 2007) to use it as a positive source to increase liking towards target objects. The reference to the self has been demonstrated as a powerful way to change how individuals evaluate and react to objects (Gawronski et al., 2007; Pelham et al., 2002). The SR task goes further in operationalizing this link between an object and the self with a learning procedure based on the intersecting regularities principle (Hughes et al., 2016) through a common action.

**The SR task on attitudes and identification: implications for behavior change**

The meta-analyses conducted on implicit and explicit attitude change both support the strength of the SR task in influencing preferences towards target objects. Prior studies on EC demonstrated that when the self is paired with an object, this object acquires positive valence (Walther & Trasselli, 2003; Zhang & Chan, 2009). However, none of these studies assessed implicit attitudes. As far as we are aware, the SR task is the first associative learning procedure, among those that rely on the self as source of attitude change, to show effectiveness on implicit attitude change. Given the robustness of this effect and the role of implicit attitudes in shaping behaviors in many different contexts, such as food choice (e.g., Friese, Hofmann, & Wänke, 2008), purchase behavior (e.g., Maison, Greenwald, & Bruin, 2004), and social interactions (e.g., Dovidio, Kawakami, & Gaertner, 2002), the SR candidates as a powerful tool to increase the likelihood of affecting behaviors through the attitude route. In support of this hypothesis, Richetin, Mattavelli, and Perugini (2016, study 1) demonstrated the mediating role of implicit and explicit attitudes in a serial manner in the SR effect on behavioral intention. Once an implicit attitude toward a food brand changes as a function of the SR task, the effect ramifies in a domino-like manner on explicit attitude and intention to buy that brand's products.

A further important finding is in the effectiveness of the SR to increase one's level of identification (or closeness) with a given target object. Asking people to categorize both self- and one target-related stimuli through the same action increases their perception of closeness to the target. The extent to which people identify with something (e.g., social groups or organizations) and the value of the particular membership tend to co-vary (Ellemers, 1993; Tajfel & Turner, 1979) because positive groups or categories of objects may contribute to create a positive social identity. The interplay between positive evaluation and identification processes thus might suggest that the same effect would emerge if positive stimuli other than the self were used as source in the paradigm (e.g., after being exposed to positive pictures paired with Group A and neutral/negative pictures with Group B, people would probably identify more with Group A). However, we advocate for a superiority of the self for the potential behavioral consequences of perceiving others as closer the one's self. Let us consider *defensive responding*, the human tendency to defend others that we perceive as close to ourselves when they are threatened. Research has shown that an important moderator of defensive responding is the relative salience and activation of the self-concept carried by the target. When the self is activated, people are more concerned with their self-evaluation and become more attentive to self-relevant information (Carver & Scheier, 1981; Duval & Wicklund, 1972). It seems therefore plausible that the self, used as positive source in a learning paradigm, acts as a booster for increasing the perceived closeness with the target and that this might have important consequences on behavior.

**The boundary conditions: Summarizing the effects of distinct moderators**

Moderator analyses revealed some relevant findings at the level of the two main classes of stimuli involved in the task, that is, the source and the target. The effectiveness of the SR effect on both implicit and explicit attitude is qualified by type of target, such that when it comes to known target objects the SR effect is weaker. This latter aspect seems to indicate that the paradigm is very effective in forming an attitude towards newly encountered objects, while changing existing attitudes is more challenging. The fact that the SR effect is weaker when a measure of participants' attitude towards the target object is administered before the SR manipulation might stand as a further evidence

for this point. In fact, one could say that after completing an attitude measure, participants might have higher accessibility to their preference towards the target (Powell & Fazio, 1984). Thus, once this preference becomes accessible, changing it via the SR task is more challenging. However, an alternative explanation seems plausible too. Given the reduced time interval between the attitude measures taken before and after the manipulation, participants could tend to be consistent in the evaluation provided across the two measures, therefore reducing the potential impact of the SR manipulation. To address this point, future studies could assess participants' baseline liking towards the target well before (e.g., one week) the actual experimental session, in order to reduce the chance of anchoring the new evaluation to the previous one.

The type of source stimuli used to represent the self-concept is another important moderator of the effect. In fact, on both implicit attitude change and (marginally) identification, stimuli with higher self-relevance have the power to increase the effect of the SR compared to possessive pronouns. Our data therefore show that stimuli exert an influence in qualifying the effect of the SR task. This might suggest that increasing the centrality of the stimuli for the self-concept is crucial to amplify the positive effect of the general category 'self'. The underlying idea is that idiographic stimuli should raise the focus on the core self more than generic stimuli do (see Bluemke & Friese, 2012). Even though we acknowledge that there were only three studies with idiosyncratic sources stimuli to test for this moderation, this promising finding suggests that when using particularly self-evoking stimuli, the SR effect increases. Future research should aim at clarifying whether enhancing the reference of the source to the self can lead to a change, rather than a formation, in the evaluation of objects with pre-existing attitudes.

When focusing on the general properties of the task, the SR effect is not qualified by the number of trials included in the task, while it is by the number of errors made throughout the performance of the task. This suggests that the SR effect primarily relies on participants' ability to learn the rule that underlies the task and to apply it (i.e., the self and target A share a common feature), rather than on the experienced stimulus-response contingencies. Future investigations should explore

further this issue. For example, one could investigate the SR effect partialing out the role played by instructions. In the standard procedure, participants might learn through intersecting regularities that are both operationalized in the stimulus-response contingencies and instructed throughout the task. A study implementing an 'instructions-only' (participants are exposed to the task rule but never actually perform it) versus a 'performance-only' (participants have to figure out the rule by themselves throughout the completion of the task) version might provide important information about the nature of the effect.

Some important considerations raised from the effect of the SR on different attitude measures. When the type of implicit attitude measure was considered as a moderator of the effect, the effect was weaker and did not reach the significance level for the AMP whereas it was for the IAT. A quite likely explanation is that the IAT has better psychometric properties than the AMP (Bar-Anan & Nosek, 2014). However, one might also argue that this could be due to the fact that the SR and the IAT share some method-variance. We will return to this issue below.

Finally, we want to shed light on another aspect related to the type of target stimuli used in the task. We found that the effect of the SR task on implicit attitude change is stronger for brands than for groups, while the opposite pattern emerged on explicit attitude. This finding might be interpreted in light of a propositional, rather than associative, interpretation of attitude change via the SR effect. Pressing the same key to categorize the self and an object results in links between the two types of stimuli. From an associative perspective, these links should result into associations between the source and the target concept, such that the view of the target concept should activate the source (i.e., the self) and its valence. In fact, an association "*simply connects the mental images of a pair of events in such a way that activation of one image causes activation (or inhibition) of the other*" (Shanks, 2007, p. 294). However, from a propositional perspective, these links between the self and the target objects might be qualified in a relational way (De Houwer, 2009b). To the extent to which implicit and explicit attitude are affected differently by distinct propositional relations (e.g., *"I possess that brand"* or "*I am part of that group"*), it seems reasonable that the observed difference

between the SR effect at the two attitude levels based on the type of target is a function of different propositions inferred from the nature of the stimuli. Future studies should be targeted to better understand the role of propositional factors in SR effects.

**Learning and recalling the IR principle: Symbolic meaning construction throughout the SR task**

The present findings support the importance of participants' ability to use the general features of the task to extract an underlying rule from it. As a related remark, the analyses showed that for both attitude and identification, participants' ability to recall the intersecting regularities between the self and a target stimulus was a crucial moderator of the effect. These findings were consistent in both the mixed models analysis at the level of the participants and in the standard meta-analysis procedure, where the separate effect sizes were estimated on different subsamples for each outcome variable. The effect on implicit attitude was not significant for participants with incorrect memory of the IR and was significantly bigger for participants with correct IR memory than for those with no memory. At the explicit level, the SR effect for participants with correct memory was larger but not significantly different from those with no memory, and participants with incorrect IR memory showed a significant effect but in the opposite direction (i.e., opposite to the objective pairings but in line with their subjective recall of the pairings). These findings advocate the importance of the IR memory question as an indicator of participants' ability to acquire the principle that underlies SR task. Note also that participants who responded "I don't remember" to the IR memory question might include participants who were not confident in their answer. Hence in principle they could have implicit but not explicit memory of the correct pairing. Future focused studies would be needed to better understand this issue.

The fact that the SR effect on the attitude measures is qualified by IR memory has also important theoretical implications. IR memory for the SR effect can be treated as the conceptual equivalent of contingency awareness in EC. In a meta-analysis, Hofmann et al. (2010) showed that EC effects strongly decrease when participants are not able to recall the correct US-CS pairing

acquired throughout the EC paradigm. The observation that contingency awareness is such a strong moderator of EC cannot be accounted for by existing associative models of EC, according to which, CS–US associations are formed and influence liking regardless of whether participants are aware of the CS–US contingency. Instead, the strong impact that contingency awareness has on EC is in line with a propositional account of EC. Participants need to form a conscious proposition about the CS–US relation before this relation can influence liking of the CS (De Houwer, 2007, 2009). We believe that the pattern of results observed in the present meta-analysis with regard to the effect of IR memory is in line with the previous findings obtained on EC. Though associative in the terms of the procedure of its paradigm, the SR effect seems to rely on participants' ability to infer relations between stimuli from learned intersecting regularities.

It might remain less intuitive why, even in the optimal condition in which one learns and recalls in memory that the self shares something with an object, the latter should acquire the properties carried by the former. Derived from De Houwer and Hughes' (2016) reasoning, our hypothesis is that, once they learn the rule, participants make use of distal cues to convey a specific meaning to the relationship between stimuli. The authors explained this hypothesis with reference to EC. Following a symbolic interpretation of EC, people should learn from a proximal cue (i.e., stimuli pairing) that two object occurs together. Then, people might apply to this a simple heuristic rule formed throughout the history of their life that tells them that things that occur together are often similar. In fact, in their learning history, often things that share something are also similar. We believe that a symbolic meaning construction can similarly explain evaluative learning via IR. In the case of the SR effect, the proximal cue is the common action between the self and a target. From their life history people should derive that things that share something are somehow related (see Hughes, De Houwer, & Barnes-Holmes, 2016 for a discussion on the role of distal regularities on evaluative learning).

**The construct validity of the SR manipulation**

Although the data provide robust empirical evidence concerning the effectiveness of the SR manipulation, it is possible to raise questions concerning its construct validity. Specifically, one might

argue that the SR task serves as a practice for one IAT block mapping that leads to shorter response latencies on the measurement level (IAT) without necessarily affecting the targeted construct (implicit attitude), whereas the effect of the SR task on explicit attitudes and identification might be interpreted as a demand effect. Several lines of empirical evidence do not support these possibilities. Concerning the IAT, our meta-analysis reports that the effect of the SR manipulation on implicit attitude was not affected by the way in which intersecting regularities between the self and target are established in the manipulation phase. Specifically, in three studies participants categorized self/other-stimuli and target objects through joystick movements (left and right; up and down). This way to implement the SR task reduces the method-variance overlap between the manipulation and the implicit attitude measure given that participants are trained to relate stimuli via a different action (joystick movement) than in the IAT block (key presses). Yet, the effect sizes of the SR modified manipulation were not reduced compared to the standard manipulation, despite that the method-variance overlap was arguably reduced, suggesting therefore that the effect of the SR manipulation must imply more than speeding up responses in the IAT because of practicing the mapping of one critical block in the SR. Additional evidence comes from an ancillary analysis of our meta-analytical data-set. A response mapping explanation of the effect of SR on the IAT implies that the two outcomes should be correlated if all they share is method rather than construct variance. We therefore calculated a performance score for the SR manipulation (mean response latencies divided by their standard deviation). For those studies in which the focal information was available and/or considered as relevant for the experimental design ($K = 32$, $N = 4220$)[6] we tested whether this performance score was related with the IAT. To test this relation, we adopted a mixed model approach, with SR

---

[6] For instance, no SR performance score was calculated for those studies that used a SR task that did not imply key pressing ($K = 3$), nor for other variations of the SR task related to the number of stimuli categories ($K = 3$). Studies assessing implicit attitude change through the AMP ($K = 3$) were also excluded.

performance score treated as fixed and random (at the study level) factor and IAT score entered as criterion. Although significant, the observed relationship was minimal, $\beta = .05$, $SE = .02$, $p = .032$. On the one hand, the significant relation suggests that the performance of the SR task could in fact serve as a practice phase that facilitates the subsequent performance of the IAT. Critically, however, the low magnitude of the beta coefficient compared to the magnitude of the general effect of the SR task on implicit attitudes ($\beta = .53$, $SE = .02$) indicates that this issue is relatively marginal and does not challenge the genuineness of the effect of the SR task on implicit attitude change.

Concerning possible demand effects, there are experimental results that do not support this explanation. For example, the experimental set-up described in detail in Perugini et al. (2013) was such that positive others (e.g., best friends, boy/girl-friend, inspirational figures) were very salient. Participants were asked to generate a short list of their most admired people (study 2) and of their single most admired person (study 3). These stimuli were then used as source stimuli for the category other in a subsequent SR task. In terms of a potential demand effect explanation, one could have well argued that it was clear to participants that we were interested in showing an effect of positive others. Yet, the results showed a strong SR effect, hence in opposite direction to a demand effect hypothesis prediction.

In further support of construct validity, overall implicit attitude correlated significantly with explicit attitudes ($r = .26$) and with identification ($r = .33$). These significant correlations suggest that the SR manipulation does actually change attitudes. Finally, findings from recent studies provide further evidence. For example, in the first study of Richetin, Mattavelli, and Perugini (2016, see chapter 4 of the present dissertation) participants, after completing a SR task in which they paired two alternative fictitious eco-brands with either the self or the category others, completed an IAT, a semantic differential and a shopping choice task. The findings showed the effectiveness of the manipulation on all three outcome variables. Moreover, the three dependent variables were correlated with each other (all $r$s > .18). Furthermore, a serial mediation analysis showed that a) the effect of the SR on the IAT fully mediated the effect of the SR on explicit attitude and that b) the effect of the SR

on the IAT and explicit attitude in series fully mediated the impact of the SR manipulation on one's behavioral choice (see Mattavelli et al., 2016, presented in chapter 5 for an example of a SR manipulation affecting behavioral intentions). To summarize, both experimental and correlational evidence support the construct validity of the SR manipulation which should be further corroborated by future research.

**Limitations, recommendations and new directions**

The main goal of the current meta-analysis was to provide a quantitative summary of the empirical evidence on the SR task, as well as to identify potential boundary conditions of the effect. To this end, it was based on published and unpublished studies from our own research group regardless of whether they did obtain significant effects. The database of the meta-analysis is therefore limited given that it did not include studies from other labs using the same paradigm nor different paradigms for answering broader questions concerning for instance the role of the self in attitude formation and change. Nonetheless, the substantial number of studies and participants allowed to investigate reliably issues in a way that would not have been feasible with a handful of studies as well as to present robust empirical evidence concerning the SR task.

One may wonder why so many studies in this meta-analysis were unpublished. Some studies because a less than perfect pattern of results would have made them difficult to publish (Giner-Sorolla, 2012), a few because we decided to gather additional evidence on specific issues, and others are included in manuscripts currently under review or in preparation. Many of them, however, are unpublished simply because we did not find enough time to write them up as manuscripts. We believe however that is worthwhile to share with the research community all what we have learned by using the SR task during these years, regardless of whether the results have been already published. This manuscript can be considered as an example of a lab-disclosure contribution on a line of research that summarizes all available data gathered during a certain time period and circumvents the pitfalls of the file-drawer effect or of the publication bias that can lead, among others, to systematic overestimation of effect sizes (Ioannides, 2008; Open Science Collaboration, 2015). In particular, the

present meta-analyses showed that, in light of its effectiveness in affecting implicit and explicit evaluation of attitudinal objects, the SR candidates as an associative learning task that any researcher can use to induce changes in attitude towards neutral stimuli. The higher importance of the specific aspects, rather than general ones, offered important indications from two perspectives. First, to make accurate predictions, researchers must be aware that the effect of the SR is higher in forming attitudes towards novel objects, while the transfer of positivity is weaker when it comes to change existing attitudes. Second, when the aim is to maximize the power of the paradigm, it is important to select carefully source stimuli: the higher their personal relevance, the more participants can transfer the properties of the self to the target.

The central role played by IR memory in determining the magnitude (and the direction) of the effect has also some practical implications on how to analyze and present SR data. We believe that to be fully informative, each SR study should report the analysis of the effect both on the full sample and on the correct IR memory participants, the latter representing the optimal condition for the effect to occur. Moreover, this suggests that manipulations or means that can increase a correct IR memory among participants would also likely increase the overall SR effects.

As a final remark, we observed that the SR task is genuinely effective on an extra-evaluative domain, that is, identification. In addition to this, Richetin, Perugini, & Mattavelli (2016) recently showed that target objects characterized through the same action as the self acquire higher cognitive accessibility. Taken together, these findings stand for an early demonstration that other cognitive properties carried by the self-concept might transfer to related stimuli via intersecting regularities. In fact, if we look at changes in liking through the SR task as the consequence of learning to act as if one novel stimulus is equivalent to the self, then one might attribute other psychological properties of the self to the novel stimuli related to it. Future studies should better address this point and exploit the special status of the self to broaden the effect of the SR task on other domains.

**Chapter 2.**

**Self-relevance is more important than valence in the self-referencing effect.[7]**

---

[7] This chapter is based on a work that is currently in preparation and will eventually result in a manuscript.

**Overview**

After having reviewed the SR task by considering the results of the empirical studies in which it has been used, this chapter considers the important role played by the self-concept in driving the SR effect. We started from the assumption that the self must be treated as a complex stimulus that comprises the most relevant characteristics of each individual (Symons & Johnson, 1997). Given that the characteristics that describe one's self can vary along both the *relevance* and the *valence* dimension, we investigated which of the two dimensions drives the SR effect. The meta-analysis showed that using idiographic stimuli lead to a stronger effect, but the main aim of these studies using idiographic stimuli was to test relevance vs. valence. We tested this aspect across four experiments ($N = 463$) in which we tailored the SR task with idiographic characteristics of the individuals, using self-relevant personality traits as stimuli that represented the self, while the self-irrelevant ones described the category 'others'. The categories 'self' and 'others' were represented by either equally moderately positive (Study 1) or moderately negative (Study 2 and 3) personality traits. In Study 4, we used moderately negative personality traits to represent the self, and moderately positive adjectives were used for others. The effect of this personalized version of the SR task was tested on implicit and explicit attitude change and identification towards two fictitious brands of food products. Results showed the indirect effect of the general positivity of the self on attitude change and identification when self-relevant stimuli are used. In essence, objects paired with self-relevant stimuli were preferred over those paired with the self-irrelevant ones, irrespective of their valence. Moreover, a meta-analytical comparison between the effect of the standard SR and this novel version of the task in changing attitudes and identification with brands showed the substantially stronger effectiveness of the latter. We discussed the results in light of their implications for the underling processes of the SR effect and, more in general, of associative learning effects based on the self.

**Introduction**

Attitudes are defined as the tendency to evaluate an object with some degree of favor or disfavor (Eagly & Chaiken, 2007) and can have profound effects on behavior by influencing the construal of the current situation, eliciting spontaneous approach-avoidance tendencies, and guiding the formation of deliberate action plans (Fazio, 1990; Strack & Deutsch, 2004). Preferences also shape what we pay attention to (Smith, Fazio, & Cejka, 1996), how we recall past events (Bohner & Dickel, 2011) as well as the judgments and decisions that we arrive at. Consequently, how, when and why likes and dislikes are established and changed is relevant for many aspects of psychological life.

**Associative attitudes change**

Empirical evidence suggests that likes and dislikes are learned, rather than innate (Rozin & Millman, 1987). One way through which attitudes can form and change is via associative procedures. Over the last thirty years, different types of procedures through which novel preferences may be formed and existing ones altered have been proposed (De Houwer, 2007). The prime example of an associative procedure to induce likes and dislikes is evaluative conditioning (EC). EC corresponds to changes in liking that are due to the repeated pairing between an unconditioned stimulus (US) and a conditioned stimulus (CS, see Hofmann, De Houwer, Perugini, Baeyens & Crombez, 2010). As a result of these repeated pairings, the CS tends to acquire the valence carried by the paired US.

Recently, Hughes, De Houwer, and Perugini (2016) proposed that likes and dislikes may also origin from learning mechanisms other than stimuli pairings, and identified the intersecting regularities (IR) principle. This learning principle is based on the idea that people learn that two objects are somehow related by merely realizing that they share something. The consequence of this newly acquired relation is the transfer of liking from the valenced source to the neutral target. In one of their experiments, the authors asked participants to perform a simple categorization task in which they used the same key (e.g., "F") when presented, with equal probability, a neutral brand or positively valenced stimuli and a different key (e.g., "J") when presented another neutral brand name or negative valenced stimuli. The regularities (i.e., pressing key upon stimulus presentation) regarding

the first neutral brand and positive stimuli, as well as the ones regarding the second neutral brand and negative stimuli, thus intersect in terms of response. In other words, an intersection was established through a common response between valenced and neutral stimuli in either of the two sets of stimuli. The results demonstrated increased implicit and explicit liking towards the brand sharing the response with the positively valenced stimuli relative to the alternative brand. Associative learning via IR is possible in many different ways. It can involve multiple and complex intersections between stimuli and actions (Hughes et al., 2016, Study 4; Mattavelli, Richetin, & Perugini, 2016) and be guided by distinct type of valenced sources.

**Changing attitudes via the self**

Extant research suggests that an important source of attitude change is the self. Studies on attitude formation and change have focused on the idea that objects that are linked to the self-concept can be better liked. Nuttin (1985) proposed the *name letter effect*, observing that letters that appeared in an individual's name were generally better liked. Similarly, Beggan (1992) observed that objects given to an individual attained greater liking than expected positive evaluations compared to when that object was not owned by that individual (*mere-ownership effect*). Evidence for these assumptions can be found in a series of studies by Gawronski, Bodenhausen, and Becker (2007). Gawronski and colleagues found that choosing between two equally attractive pictures changed participants' implicit evaluations of these pictures, such that they evaluated the chosen pictures more positively after than before the decision.

This advantage of the self has been exploited also in the field of associative learning. Given that most people show positive evaluations of themselves (Bosson, Swann, & Pennebaker, 2000; Greenwald & Farnham, 2000; Koole, Dijksterhuis, & Van Knippenberg, 2001), the creation of a new association between an attitude object and the self should lead to more positive evaluations of the attitude object. Walther, Nagengast, and Trasselli (2005) showed that using the self as a unconditional stimulus to pair it with a neutral conditioned stimulus results in increased linking of the latter (Walther & Trasselli, 2003; Zhang & Chan, 2009).

The Self-Referencing task (SR, Prestwich, Perugini, Hurling, & Richetin, 2010) also relies on the self as a positive stimulus to be paired with a target via intersecting regularities (IR). In the SR task, participants have to respond to a first neutral stimulus and self-related pronouns (e.g., the word "me") by pressing one response key (e.g., yellow key) and to another neutral stimulus and other-related pronouns by pressing another response key (e.g., red key). As a result, participants evaluate the first neutral stimulus (which was assigned to the same key as self-related words) more positively than the second neutral stimulus (Prestwich at al., 2010; Perugini, Zogmaister, Richetin, Prestwich, & Hurling, 2013; see also Ebert, Steffens, von Stülpnagel, & Jelenec, 2009). In the Self-Referencing task, the valence of one stimulus (e.g., "self") transfers to another (neutral) stimulus when they both signal that the same response should be emitted. Given the fact that one's ability to learn that self-stimuli and target objects relates because of a common action is crucial for the SR effect to occur, in SR studies participants are asked to indicate which target has been categorized through the same action as the self. This question, which is asked at the end of the SR task, is an IR memory question and can be considered as the conceptual equivalent of the most common way to detect contingency awareness in EC (see Sweldens, Corneille, & Yzerbyt, 2014).

**The self, its positivity and the motives behind it.**

The links between objects and the self are meant to affect one's social knowledge structures such as the ones that Greenwald, Banaji, and colleagues modelled in their 'unified theory' (Greenwald et al., 2002). The model incorporates three elements – the self-concept, its valence and an attitudinal object – and it is based on triangular associations between them: the association between the self-concept and valence is one's self-esteem, that between the object and the self (which in the case of the SR task is experimentally induced) is defined as objects identity, while the association between the object and valence is the resulting attitude. However, the self-concept is a complex and very articulated structure that defines our person (Symons & Johnson, 1997). It plays a key role in memory when we filter information, construct knowledge, or evaluate objects and people, including ourselves (Markus, 1977; Greenwald & Pratkanis, 1984; Mussweiler & Strack, 2000). Therefore, rather than

automatic, the link between the self-concept and positivity might reflect one's need to maintain a positive view of (her)himself. For instance, when asked to evaluate themselves, individual have the tendency to be self-serving. They believe that they are more trustworthy, moral, and physically attractive than others. The better-than-average effect is a prime example of a self-serving bias showing that people seek to maintain positive images of themselves relative to others. It can be viewed as a type of self-serving bias in which people evaluate their characteristics more favorably than those of others. This phenomenon has been demonstrated with respect to many individual characteristics, like behavior ratings (Allison, Messick, & Goethals, 1989), perception of risk and misfortune (Perloff & Fetzer, 1986) and, more central for the present research, personality trait ratings (Alicke, 1985; Dunning, Meyerowitz, & Holzberg, 1989). Social comparison theory (Festinger, 1954) offers a useful framework to understand self-serving biases like the better-than-average effect. Social comparisons entail three fundamental elements: a target with whom comparisons are made, the particular dimension being compared, and a motivation for self-evaluation. This motivation reflected in the comparison is self-enhancement, which is achieved by viewing one's traits and prospects more favorably than those of others.

**Self-relevance and valence: Implications for associative learning effects**

The self-concept can be represented by many distinct characteristics. Crucially, these characteristics may vary in their closeness to the self-concept's core. The extent to which a given characteristic is close to the self-concept core reflects its relevance for one's self, in turn defined as "*the relative importance that various aspects bear in a person's conceptualization of objects in a given cognitive domain*" (Scott, Kline, Faguy-Coté, & Peterson, 1980, p. 12). This has important implications for associative tasks that make use of stimuli to represent the self-concept. The relation between stimuli relevance and the self is based on the idea that the more central a trait is, the more conceptually relevant the associative echo from the stimulus prompt will be. Therefore, the relevance of the stimuli for the self should help people to focus on the concept in question (i.e., the self). The importance of stimuli relevance for the activation of the self-concept in associative task has been

tested by Bluemke and Friese (2012) in the context of IAT research. The authors showed that the use of idiographic stimuli (e.g., first name, family name, date of birth) focused participants' attention on the core features of the self, hence, tapping into self-related associations to a stronger degree than generic stimuli (e.g., pronouns). To explain the observed phenomenon, Bluemke and Friese referred to the *core-concept model*. Based on this model, idiographic stimuli increase the validity of implicit measures involving the self-concept. This model is based on research on semantic network models and considers self-schemas as cognitive–affective structures (Markus, 1977). It assumes that when the self-concept is represented by relevant stimuli, the mental representation will be centered more strictly on the core self.

It seems therefore plausible that the relevance of the stimuli used to represent the self-concept should determine the activation of the self-concept upon the encountering of those stimuli. Thus, we reasoned that this could have important implications also from the perspective of attitude change driven by the self-concept. In fact, if self-relevant stimuli are more effective than general stimuli in activating the self-concept, and the activation of the self-concept is a crucial pre-requisite for the transfer of the positivity of the self to related objects, then the effect of the self on attitude change should be stronger when self-relevant stimuli are used. The higher activation of the general self-concept driven by self-relevant stimuli would result in a stronger transfer of liking from the self to the related object.

**Aims of the research**

Previous studies on the SR task are in line with the idea that evaluative learning driven by the self can be better understood if one considers characteristics other than valence as key drivers of the effect. For instance, Perugini, Zogmaister, Richetin, Prestwich, & Hurling (2013) showed that people prefer objects related to the self even when the category 'others' is represented by positive idiosyncratically chosen exemplars (e.g., the person they admired the most). Further evidence of the superiority of the self on attitude change also comes from Perugini, Richetin, & Mattavelli (2016, study 3). The authors used a tailored version of the SR task where the category self was opposed to a

more positive source category (i.e., mammals) that was shown in a previous study to be highly effective in changing implicit and explicit attitudes. Implicit attitude was in favor of the stimuli paired with the self, as it was explicit attitude for participants with positive self-esteem and positive orientation. Taken together, extant SR studies provide evidence that the SR effect cannot be easily reduced to a matter of valence only.

Using a modified version of the SR task, we tested whether evaluative learning effects through intersecting regularities are primarily driven by the valence or the self-relevance of personality traits. Concretely, we designed a SR task in which the categories 'self' and 'others' were represented by either equally moderately positive (Study 1) or moderately negative (Study 2 and 3) personality traits. By manipulating stimuli valence (across studies) and relevance (within studies) we tested which of the two aspects drives the SR effect on the targeted objects. To strengthen the robustness of this test, in Study 4 we designed a SR task where moderately negative personality traits represented the self, whilst moderately positive adjectives were used for others. In doing so, we manipulated both the valence and self-relevance of the stimuli within the same task. We tested the effect of this personality SR on implicit attitude, explicit attitude, and identification. Attitudinal objects were two fictitious brands of food products.

Moreover, in line with the idea that people tend to be self-serving in evaluating themselves, we were also interested in seeing whether individuals showed any self-serving bias towards self-relevant traits. In essence, we tested whether the adjectives, once selected as relevant for one's self and used as source stimuli in SR task, were also considered as more positive compared to those that were chosen as non-relevant for one's self. This was also done to test the extent to which a self-serving bias could contribute in determining the potential preference for objects related with self-relevant traits.

We used the SR task as an associative learning procedure because it relies on the self to generate attitude change. Therefore, this paradigm is ideally suited to investigate how manipulations at the level of the properties of the self-stimuli (i.e., valence and relevance) can determine evaluative

learning effects. Second, the task has an inherent comparative structure. It makes people link one object with a source and another object with the contrast source 'others'. As we said, an additional aim of the research is to look at self-serving biases towards self-relevant traits. In this sense, the structure of the paradigm provides a target category that induces social comparison, which is a pre-requisite for a self-serving bias to occur.

**Study 1**

With this study we tested the effect of SR task using personality traits that are moderately positive and either relevant or non-relevant for the self-concept on attitude change (implicit and explicit) and identification towards two fictitious brands. We manipulated stimuli relevance by asking participants to select five adjectives representative of themselves and five that were not. Although equal in terms of valence, the two lists of adjectives would produce differential effects on liking and identification towards the paired brands depending on whether the stimuli activated the self-concept. We also measured whether, despite of their assumed equal valence, adjectives liking increased after being selected as representative of one's self and if this possible increased liking did moderate evaluative learning effects.

**Normative pre-selection of moderately positive adjectives**

We generated an initial list of 60 adjectives containing moderately positive personality traits (from Caprara & Perugini, 1991). Thirty-five individuals (9 men, 26 women, $M$age = 24.46, $SD$ = 3.37) rated these adjectives on two dimensions in an online questionnaire. First, they indicated the valence on 11-point scales from 0 (very negative) to 10 (very positive) and then they indicated the extent to which each of the adjectives was useful to describe someone's personality on 11-point scales from 0 (very useless) to 10 (very useful). From the two sets of ratings, we selected 20 adjectives with a mean valence rating of 5.96 ($SD$ = 0.64). A one sample $t$-test revealed that the average score significantly differed from the midpoint of the scale, $t(34) = 8.77$, $p < .001$. The mean usefulness rating was 6.50 ($SD$ = 0.86), also significantly different from the midpoint, $t(34) = 10.36$, $p < .001$. To increase the number of adjectives to be used in the main study, we added 5 adjectives that received

moderately positive rating from the pretest conducted to identify moderately negative traits (see Study 2), *M* valence = 5.84, *SD* = 0.98 and *M* usefulness = 6.13, *SD* = 1.29. Both the valence and usefulness average scores were significantly different from the midpoint, $t(37) = 36.59$, $p < .001$ and $t(37) = 29.19$, $p < .001$, respectively.

**Method**

### Participants and Procedure

A hundred and twenty students (71 women, 49 men, *M* age =22.24, *SD* = 2.58) took part individually to a one-session study. After giving informed consent, participants took a familiarization task to learn to assign products to the brand they belonged to and then completed an adjectives selection task.[8] All participants then took the self-referencing task. For half of the participants, the manipulation consisted of pairing one brand with self-relevant moderately positive adjectives and the contrast brand with self-irrelevant moderately positive adjectives, whereas for the other half was the opposite. At that point, all participants completed the intersecting regularities memory measure, followed by an IAT, an explicit evaluation of the brands, and a measure of identification. Then, participants completed an evaluation of the adjectives used in the SR task. Finally, participants completed demographics, were thanked, debriefed, and given course credit.

### Materials

---

[8] In order to identify two potential brand names and logos to be used in the main studies, we ran a pilot test. Two different pairs of fictitious brands and their products (i.e., Pecivo vs. Sabedo and Lestea vs. Sabea) were selected as the most neutral among a set of twelve potential candidates were presented in the pilot test. Thirty-three (10 men, 23 women, *M*age = 25.15, *SD* = 4.31) individuals completed two IATs and also indicated their explicit evaluation of the four brands on four different dimensions (i.e., logo, brand, products and name). We chose the two brands Lestea and Sabea as their differential implicit evaluation did not differ from 0, $t(31) = 1.46$, $p = .155$, as it happened also for the explicit differential score on the four dimensions listed above, $t(31) < 1.07$, $p > .294$.

*Familiarization task*. Participants indicated, as quickly as possible, the logo/brand to which a product belonged by pressing one of two keys. On each trial, pictures of products appeared in the center of the screen. Labels for both brands remained on the upper portion of the screen throughout the task, each placed on the same side as the key used to classify products. The familiarization consisted of a total of 20 trials (5 pictures for each brand) divided into two blocks of 10. In each block, half the trials featured products of one brand, half products of the other, intermixed at random. The positions (left vs. right)for the two brands were alternated block by block. The order in which participants completed these blocks was counterbalanced. For each of the two target categories, five pictures were used. The products were used with the logo Lestea for half of participants and with the logo Sabea for the other half of participants (see Supplementary Materials Chapter 2 – Appendix E).

*Adjectives selection task*. Participants were asked to indicate the 5 adjectives that describe them the most and the 5 adjectives that describe them the least (not characteristic of them at all, see Supplementary Materials Chapter 2 – Appendix D) among a list of 25 adjectives.

*Self-Referencing task*. The Self-Referencing task (Prestwich et al., 2010) consisted of four blocks of 40 trials. In the first two blocks participants categorized, as quickly as possible, Lestea versus Sabea product pictures and self-irrelevant adjectives to one response key (e.g., "blue") on the response box and Sabea versus Lestea product pictures and self-relevant adjectives to a different response key (e.g., "yellow"). Participants then repeated the two blocks of 40 trials with the keys to which the target categories were assigned switched. The order in which participants completed these two blocks was counterbalanced. When an exemplar was incorrectly classified, a red-X appeared on screen and remained until it was correctly classified. The inter-trial interval was 400ms. For each of the two target categories, 5 pictures were used (see Supplementary Materials Chapter 2 – Appendix E). The category labels 'self' and 'others' remains visible throughout the completion of the task. Like for the familiarization task, the products were used with the logo Lestea for half of participants and with the logo Sabea for the other half of participants. If participants had a percentage of errors above

15% in the last block of 40 trials, participants completed 2 additional blocks of 20 trials (one block for each key assignment).

*Intersecting Regularities Memory*. Participants were asked about their recollection of the intersecting regularities between the source and the target. Specifically, the question was the following: "One of the tasks that you have done consisted in classifying with the same key adjectives related to the self and pictures of one brand. Do you remember which brand?" Participants indicated one of the two brands or the option "I don't know". Answers were coded as correct memory (correct response) or not (no recollection or incorrect response).

*IAT*. Participants classified words and pictures that were presented individually and in a random order in the middle of the screen using two colored keys on a response box (i.e., 'yellow' and 'blue'). The target concept was Lestea and its contrast was Sabea, 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 Lesta and Positive being presented first and the other half having the combination Sabea and Positive being presented first. All practice blocks consisted of 20 trials and each critical block consisted of 81 trials (4 blocks of 20 trials including one initial dummy trial for the first two critical blocks). 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, the inter-trial interval of 500ms, and the category labels stayed on the top part of the screen throughout the task. For each attribute category, five words were used (Positive: *positive, joy, happy, paradise, nice*; Negative: *negative, hell, ugly, sad, pain*), and 5 pictures were used for each target and contrast categories. Like for the learning task and the self-referencing task, the products were used with the logo Lestea for half of participants and with the logo Sabea for the other half of participants (see Supplementary Materials Chapter 2 – Appendix E). The D score with 600 ms of penalties for errors was calculated (Greenwald, Nosek, & Banaji, 2003) such that positive values indicate an automatic positive preference for Lestea over Sabea.

*Explicit Evaluation of the brands.* Participants rated each brand (Lestea first when first associated with positive in the IAT vs. Sabea first when first associated with positive in IAT) separately on 4 bipolar dimensions (*ugly-nice, unpleasant-pleasant, worthless-valuable, repulsive-attractive*) on 7-point scales from 1 to 7. As dependent variable, we used the differential mean of the four items used for each brand, such that higher scores indicated a preference for Lestea over Sabea.

*Identification measure.* This measure was adapted from the inclusion of other in the self scale (Aron, Aron, & Smollan, 1992) and the inclusion of the in-group in the self scale (Tropp & Wright, 2001). Participants evaluated their level of closeness with each brand indicating the corresponding pair of circles representing the self and the brand on a scale ranging from 1 (very distant circles) to 7 (overlapping circles). A single score was computed by subtracting the two identification scores such that higher scores indicated a preference for Lestea over Sabea.

*Post-SR adjectives rating.* Participants rated the valence of each of the 10 self-relevant and self-irrelevant adjectives on 11-point scales from 0 (very negative) to 10 (very positive).

**Results**

The data from two participants were excluded from the analyses due to high error rate ($> 25\%$) in either the IAT or the SR task. From the 118 remaining participants, 103 correctly responded to the IR memory (87.3%) and were considered for the focal analysis (see Supplementary Materials Chapter 2 – Appendix C for results on all participants). Both the IAT and the explicit measure showed good internal consistency ($\alpha = .87$ and $\alpha = .86$, respectively). We first looked at the effect of the SR manipulation on implicit attitude change. Statistics are reported in Table 4. We ran a series of one-way-ANOVAs with SR manipulation as independent variable and IAT score, explicit attitude, and identification as dependent variables. We observed a significant SR effect on all three criteria. Participants who pressed the same key for Lestea and self-related moderately positive adjectives showed greater implicit and explicit liking as well as identification for Lestea, compared to those who paired the same brand with non-central adjectives. The three dependent variables were positively and significantly correlated (see Table 5).

Table 4. *SR condition on implicit and explicit attitude and identification for each study (correct IR memory participants).*

| | Lestea Self-Adjectives | | Sabea Self-Adjectives | | *F* | *p* | Cohen's *d* [95% CI] |
|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | | | |
| **Implicit attitude** | | | | | | | |
| Study 1 | 0.10 | 0.44 | -0.40 | 0.49 | 29.26 | < .001 | 1.08 [0.66, 1.49] |
| Study 2 | 0.00 | 0.45 | -0.50 | 0.41 | 32.37 | < .001 | 1.16 [0.72, 1.59] |
| Study 3 | 0.08 | 0.49 | -0.36 | 0.53 | 18.06 | < .001 | 0.86 [0.45, 1.29] |
| Study 4 | -0.07 | 0.54 | -0.49 | 0.34 | 20.67 | < .001 | 1.94 [0.52, 1.36] |
| **Explicit attitude** | | | | | | | |
| Study 1 | 0.23 | 1.41 | -0.88 | 1.33 | 16.65 | < .001 | 0.81 [0.41, 1.21] |
| Study 2 | 0.31 | 1.47 | -0.54 | 1.34 | 8.65 | .004 | 0.60 [0.19, 1.01] |
| Study 3 | 0.25 | 2.15 | -0.83 | 2.23 | 19.00 | < .001 | 0.89 [0.47, 1.31] |
| Study 4 | 0.02 | 1.33 | -0.77 | 1.11 | 10.09 | .002 | 0.65 [0.24, 1.06] |
| **Identification** | | | | | | | |
| Study 1 | 0.51 | 2.08 | -1.10 | 1.79 | 17.63 | < .001 | 0.83 [0.43, 1.23] |
| Study 2 | 0.55 | 1.98 | -1.15 | 2.01 | 17.30 | < .001 | 0.85 [0.43, 1.27] |
| Study 3 | 0.51 | 1.71 | -0.93 | 1.62 | 17.65 | < .001 | 0.84 [0.42, 1.26] |
| Study 4 | 0.30 | 2.18 | -1.32 | 1.68 | 16.89 | < .001 | 0.84 [0.42, 1.25] |

Table 5. *Correlations between SR condition, implicit and explicit attitude, identification and self-adjectives rating for each study (correct IR memory participants).*

|  | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **Study 1 (N = 103)** | | | | | |
| 1. SR manipulation | 1 | | | | |
| 2. Implicit Attitude | .47** | 1 | | | |
| 3. Explicit Attitude | .38** | .40** | 1 | | |
| 4. Identification | .39** | .32** | .52** | 1 | |
| 5. Self-adjectives rating post SR | -.13 | .06 | .00 | .09 | 1 |
| **Study 2 (N = 95)** | | | | | |
| 1. SR manipulation | 1 | | | | |
| 2. Implicit Attitude | .51** | 1 | | | |
| 3. Explicit Attitude | .29** | .32** | 1 | | |
| 4. Identification | .40** | .43** | .67** | 1 | |
| 5. Self-adjectives rating post SR | -.03 | -.03 | .09 | .04 | 1 |
| **Study 3 (N = 96)** | | | | | |
| 1. SR manipulation | 1 | | | | |
| 2. Implicit Attitude | .40** | 1 | | | |
| 3. Explicit Attitude | .41** | .32** | 1 | | |
| 4. Identification | .40** | .34** | .63** | 1 | |
| 5. Self-adjectives rating post SR | .20 | -.02 | .09 | .03 | 1 |
| **Study 4 (N = 97)** | | | | | |
| 1. SR manipulation | 1 | | | | |
| 2. Implicit Attitude | .42** | 1 | | | |
| 3. Explicit Attitude | .31** | .45** | 1 | | |
| 4. Identification | .39** | .41** | .72** | 1 | |
| 5. Self-adjectives rating post SR | -.02 | -.06 | .15 | .11 | 1 |

*Note*.    SR manipulation is coded 0: Sabea+Self, 1: Lestea+Self.
$* p < .05$  $** p < .01$

We then compared the evaluations for the adjectives selected as central or not for one's self both before and after manipulation. After the SR task, the adjectives selected as relevant for one's self where better liked than those that were not self-relevant, $t(102) = 13.80$, $p < .001$, although this

preference did not emerge from their average normative scores, $t(102) = -1.93$, $p = .057$. Moreover, selecting adjectives as representative of one's self also increased their liking compared to a normative evaluation of the same adjectives, $t(102) = 9.72$, $p < .001$: Individuals rated more positively the adjectives used as stimuli compared to how they are generally evaluated. Conversely, selecting adjectives as not representative of one's self decreased their liking compared to a normative evaluation, $t(102) = -14.84$, $p < .001$.

In light of the present findings, we ran a series of moderation analysis to see whether the evaluation of the adjectives selected as representative for one's self qualified the impact of the experimental manipulation on any of the outcome measure. Both the predictor (SR manipulation) and the proposed moderator (differential evaluation of self-related versus unrelated adjectives) were standardized. The interaction between the SR manipulation and the evaluation of the self-relevant adjectives was not significant for implicit attitude and identification ($ps > .378$). For the explicit attitude, the model explained 19.25% of the total variance. There was a main effect of the experimental manipulation ($\beta = .38$, $p < .001$), no significant main effect of the self-adjectives rating ($\beta = .04$, $p = .658$), but a significant interaction between the two factors ($\beta = .22$, $p = .016$). A simple slope analysis revealed a significant SR effect when the self-adjectives rating score was either medium or high ($B = .38$, $p < .001$ and $B = .60$, $p < .001$, respectively) but not significant when it was low ($B = .16$, $p = .233$, see Figure 1).

This study showed that moderately positive adjectives that are relevant for one's self lead to greater liking for and identification with the target brand they were categorized with, compared to another brand categorized with moderately positive adjectives that were not relevant for one's personality. This study is the first to show the crucial role of stimuli self-relevance in determining the SR effect. The effect emerged at both implicit and explicit attitudinal level, as well as on identification with the target brand. Also, compared to the original normative ratings, the evaluation of self-relevant traits increased after the SR manipulation, while the evaluation of self-irrelevant traits decreased.
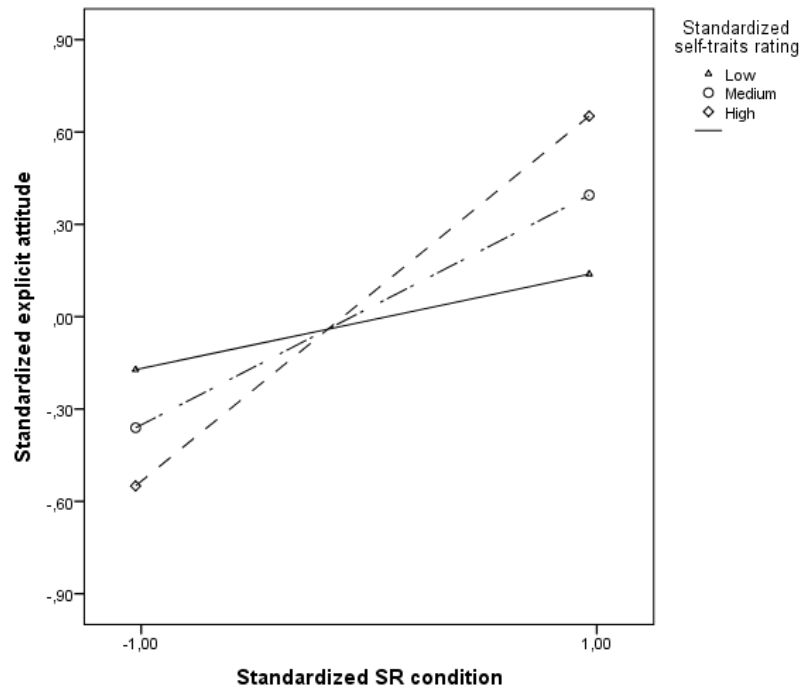
Figure 1. *Simple slope analysis for the interaction between SR condition and self-traits' rating on explicit attitude (Study 1).*

This difference in valence for the self-relevant stimuli moderated the effect of the manipulation for explicit attitude only: the SR effect on explicit attitude was no longer significant when the evaluation of the adjectives after the manipulation decreased. In essence, if people do not show the tendency to be self-serving in rating the adjectives that are relevant to depict themselves (rather they rate them as negative), neither the brand product related with those traits preferred over the alternative one. This suggests that the SR effect on explicit attitude requires some further validation process about the newly formed relation between the target objects and the self-relevant traits. It is plausible that the explicit preference for the brand related with self-relevant traits reflects the motivational need to enhance one's self (Alicke & Sedikides, 2009), which should also result in a self-serving bias in the evaluation of the traits. Thus, the absence of this motivational need might explain the lack of SR effect on explicit attitude.

**Study 2**

Study 1 showed that when the valence of the stimuli used to represent the categories 'self' and 'others' is identical, self-relevance drives the SR effect: people prefer the objects that relate to the

(moderately) positive traits that are relevant for themselves. In the first study, all the stimuli used were moderately positive in valence. Hence, a SR effect could derive from a generic impact of stimuli self-relevance. In the second study we tested whether the preference for a brand related to self-relevant personality adjectives is still detectable when moderately negative (rather than positive) adjectives are used to represent both the self and the contrast category. In this second study the relevance for the self becomes even more crucial in driving the effect: self-relevance must be strong enough a) to create an asymmetry with stimuli that are non-relevant for the self and b) to counteract the negativity carried by the source stimuli. Only if these two conditions are met, participants should exhibit greater implicit and explicit liking and identification towards the brand categorized through the same action as moderately negative self-relevant adjectives. Otherwise, we should expect no SR effect. Whether self-relevant adjectives' evaluation moderates the SR effect also in this case remains an exploratory question.

**Selection of moderately negative adjectives**

We generated a list of 60 adjectives containing moderately negative personality traits (from Caprara & Perugini, 1991). Thirty-eight individuals (8 men, 30 women, $M$age = 26.26, $SD$ = 5.59) rated the adjectives on two dimensions in an online questionnaire. First, they indicated the valence on 11-point scales from 0 (very negative) to 10 (very positive) and then they indicated the extent to which each of the adjectives was useful to describe someone's personality on 11-point scales from 0 (very useless) to 10 (very useful). From the two sets of ratings, we selected 25 adjectives with an average valence of 4.46 ($SD$ = 0.83), which was significantly lower when compared with the midpoint of the scale in a one-sample t-test, $t(37) = -4.05$, $p < .001$. The average usefulness of the selected adjectives was 5.55 ($SD$ = 1.34), and significantly higher than the midpoint, $t(37) = 2.51$, $p = .017$.

**Method**

**Participants and procedure**

A hundred and twelve students (89 women, 23 men, $M$ age =22.72, $SD$ = 3.22) took part individually to a one-session study. As in study 1, participants took a familiarization task to learn to

assign products to the brand they belonged to and then completed an adjectives selection task. All participants then took the self-referencing task. For half of the participants, the manipulation consisted of pairing one brand with self-relevant moderately negative adjectives and the contrast brand with self-irrelevant moderately negative adjectives, whereas for the other half was the opposite. All participants completed the intersecting regularities memory measure, followed by an IAT, an explicit evaluation of the brands, and a measure of identification. Then participants completed an evaluation of the adjectives used in the SR task.[9]

**Results**

The data from one participant was excluded from the analyses due to high error rate (> 25%) in either the IAT or the SR task. The IR memory screening criterion lead to a final sample of 95 participants (85.6%, see Supplementary Materials Chapter 2 – Appendix C for the analyses on the full sample). Both the IAT and the explicit measure showed good internal consistency ($\alpha = .84$ and $\alpha = .85$, respectively). Statistics are reported in Table 4. We observed a significant SR effect on the implicit and explicit attitude, as well as on the identification. Participants who categorized Lestea products and self-related moderately negative adjectives showed greater implicit and explicit liking for and identification to Lestea over Sabea. The three dependent variables were positively and significantly correlated (see Table 5).

We compared the post manipulation evaluation of the self-related adjectives with the normative scores observed in the pretest. Selecting adjectives as representative of one's self also increased their liking compared to a normative evaluation of the same adjectives, $t(94) = 7.89$, $p < .001$. In other words, adjectives were judged as more positive when they were used as stimuli to represent the self than when they were not. When this rating score was used as potential moderator of the SR effect on the outcome measures we did not find any significant interaction ($ps > .222$).

---

[9] Due to a programming error, the evaluation of the adjectives selected by individuals as not representative of themselves was not included.

Study 2 therefore showed that implicit and explicit attitude change as well as identification with a brand increases when the latter is paired with self-relevant traits, even when the traits are negative. Again, the evaluative learning effect resulting from the personality version of the SR task seems to be completely driven by the relevance for the self that the stimuli used in the task possess. This means that traits' self-relevance created an asymmetry with traits that were non-relevant for the self and that this asymmetry counteracted the original negativity of the traits. Critically however, study 2 raised an important issue. In fact, the initial selection of the self-relevant adjectives showed that the rating of stimuli chosen as representative of one's self was on average more positive than that of adjectives selected as non-relevant, $t(94) = 5.55$, $p < .001$. While an increased evaluation of the self-relevant adjectives taken after the manipulation should reflect a post-choice self-serving bias in evaluating one's self, what we observed suggested that also the selection of the adjectives might have been affected by some sort of self-serving bias, such that the adjectives perceived as less negative were more likely selected as representative of one's self. To circumvent this problem we tested whether the potential selection bias observed in Study 2 was still detectable when the adjectives were to be selected from a list of adjectives rated as moderately negative by the participants' themselves.

**Study 3**

In studies 1 and 2, we demonstrated that attitude and identification with an object varies according to the self-relevance of the stimuli the object is paired with. However, the self-serving bias in the selection of the self-relevant adjectives observed in study 2 castes doubts on the explanation of the overall results in terms of self-relevance. In fact, one could argue that valence can also explain the effect: if people select less negative stimuli to describe themselves, then self-relevant stimuli are also higher in valence than those considered as non-relevant for the self. In study 3, we tried to neutralize this self-serving bias in the selection of the self-relevant stimuli by adopting a slightly different procedure that introduces an idiosyncratic rating of the adjectives to be subsequently selected as relevant or irrelevant for themselves. This rating phase ensured that the selection phase

included only adjectives that received the same moderately negative rating, therefore excluding any potential self-serving bias in the selection.

**Method**

### Participants and Procedure

A hundred and eight students (90 women, 18 men, *M* age =22.42, *SD* = 3.70) took part individually to a one-session study. The procedure was similar to the one used for Study 1 and 2 with the exception that participants first completed an adjectives pre-rating task at the very beginning after giving informed consent. Then, participants completed a learning task, an adjectives selection task, and the Self-Referencing task. For half of the participants, the manipulation consisted of pairing one brand with self-relevant adjectives, whereas for the other half it consisted of pairing another brand with the self-relevant adjectives. At that point, all participants completed the intersecting regularities memory measure, followed by an IAT, an explicit evaluation of the brands, and a measure of identification. Then participants completed a post-manipulation evaluation of all adjectives that received a moderately negative score during the initial rating phase (including those used in the SR task). Finally, participants completed demographic questions.

### Materials

All the materials, except for the initial and final adjectives rating and the selection task, were identical to those used in Study 2.

*Adjectives rating*. Participants evaluated a list of 80 adjectives presented individually in a random order on 6-point scales from 1 (very negative) to 6 (very positive, see Supplementary Materials Chapter 2 – Appendix D). The list included the adjectives used in the previous study, with the exception of those obtaining an inadequate evaluation (i.e., too negative or too positive), and the addition of a number of potential adjective candidates. As we were interested in selecting moderately negative adjectives, we reasoned that by avoiding a middle point neutral scale, we would have facilitated participants to classify adjectives as either positive or negative. All the adjectives that received an evaluation equal to 3, corresponding to 'moderately negative', were automatically

selected by the computer. If the participant rated less than 20 adjectives he or she did not proceed further with the study. For this reason, twenty-two participants who failed in meeting this criterion were excluded from the experimental session

*Adjectives selection task*. From the list of the adjectives selected in the previous task (i.e., all adjectives that received a rating equal to 3), a new list was generated and presented to participants for the selection phase. The number of adjectives included in the new list depended on the prior ratings. If participants rated between 20 and 24 adjectives with a rating of 3, the list contained 20 of them drawn randomly. If participants rated 25 or more adjectives, the list included 25 of them. Participants were presented with a list of adjectives and asked to indicate the 5 adjectives that describe them the most and the 5 adjectives that describe them the least (not characteristic of them at all, see Supplementary Materials Chapter 2 – Appendix D).

*Post-SR adjectives rating*. Participants rated the valence of all the adjectives that received a rating of 3 in the initial rating task on 6-point scales from 1 (very negative) to 6 (very positive).

**Results**

The data from three participants were excluded from the analyses due to high error rate (> 25%) in either the IAT or the SR task. IR memory screening criterion lead to a final sample of 96 participants (91.4%, see Supplementary Materials Chapter 2 – Appendix C for analyses on the full sample). Both the IAT and the explicit measure showed good internal consistency ($\alpha = .86$ and $\alpha = .87$, respectively). One-way-ANOVAs revealed a significant SR manipulation on implicit and explicit attitude as well as on the identification towards the two brands (see Table 4 for statistics). Participants who categorized Lestea products and self-related moderately negative adjectives showed greater implicit and explicit liking for and identification to Lestea over Sabea. The three dependent variables were positively and significantly correlated (see Table 5).

We then looked at the evaluation of the personality adjectives. All the adjectives included in the selection list were pre-rated by participants and they all received the same moderately negative rating score (i.e., 3 on a scale from 1 to 6). Therefore, we compared a) the post-manipulation ratings

for the adjectives selected as relevant for the self with those that were not, and b) the post manipulation evaluations of the adjectives with the baseline score (i.e., 3). After SR manipulation, the adjectives selected as representative for one's self where better liked than those that were not self-relevant, $t(95) = 10.61$, $p < .001$. Moreover, self-relevant adjectives ratings increased whereas self-irrelevant selected adjectives ratings decreased after the SR manipulation, $t(95) = 5.54$, $p < .001$ and $t(95) = -9.40$, $p < .001$, respectively.

Post manipulation rating towards self-relevant adjectives was used as a moderator of the effect of the manipulation on the three outcomes. The model explained 23.96% of the total variance on implicit attitude change. We found a main effect of the experimental manipulation ($\beta = .43$, $p < .001$) and no significant main effect of self-adjectives rating ($\beta = -.10$, $p = .289$). Crucially, the interaction term was significant ($\beta = .27$, $p = .005$). We decomposed the interaction with a simple slope analysis (see Figure 2, left panel). The SR effect was significant when the self-adjectives rating score was either medium or high ($B = .43$, $p < .001$ and $B = .70$, $p < .001$) but not significant for low scores ($B = .16$, $p = .205$). An opposite pattern was observed on explicit attitude (see Figure 2, right panel). The model accounted for 20.33% of variance. There was a main effect of the experimental manipulation ($\beta = .40$, $p < .001$) and no effect of self-adjectives rating ($\beta = .00$, $p = .997$). The interaction term was significant ($\beta = -.19$, $p = .005$). The SR effect was significant when the self-adjectives rating score was either low or medium ($B = .59$, $p < .001$ and $B = .40$, $p < .001$) but not significant for high scores ($B = .21$, $p = .130$). The same analysis conducted on identification showed no significant interaction ($p = .900$).

Study 3 confirmed the effect of pairing self-relevant adjectives with a particular brand on implicit and explicit attitude towards as well as identification with the brand. Besides replicating the findings observed in Study 2, this study is important as it shows that the observed effects on the outcome measures cannot be ascribed to any self-serving bias in the selection of the personality traits at the beginning of study.
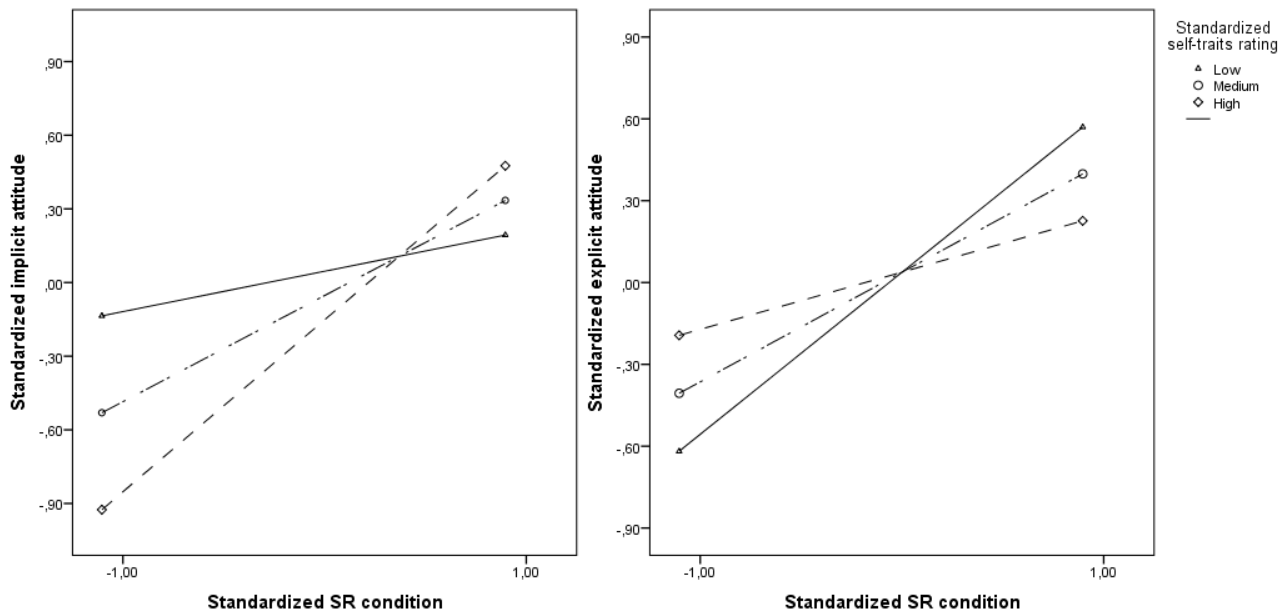
Figure 2. *Simple slope analysis for the interaction between SR condition and self-traits' rating on implicit and explicit attitude (Study 3).*

Results from the moderation analyses are also of interest as they showed a dissociation pattern between implicit and explicit attitude change and at the same time they are not consistent with what observed in Study 1. In fact, this study showed that implicitly, the more positive the evaluation of the self-relevant adjectives, the greater the preference for the brand related with them. Explicitly, the brand related with self-relevant adjectives is more liked than the alternative one when the adjectives received low or medium ratings. This result might imply that moderately negative traits that relate to the self-concept a) affect distinctively implicit and explicit attitude change and b) might induce motivational processes that differ from that induced by

moderately positive traits (Study 1). In fact, assuming that explicit attitudes tap more on motivational reasoning, one could say that in Study 1 positive traits generate explicit attitude change towards the brand when accompanied by the tendency to self-enhance (e.g., *"The more positive I evaluate myself, the higher my preference for the brand that relates to me"*). Conversely, when negative traits are used the opposite is true. A negative self-evaluation might provoke the

need to reaffirm one's positivity via the related objects (e.g., *"The more negative I evaluate myself, the higher my need to be in favor of the brand that relates to me"*). These three studies showed a self-

referencing effect when both lists of adjectives used as sources (self vs. others) are equal in terms of valence. Study 4 goes one step further in trying to demonstrate that an object related with moderately negative adjectives selected as relevant for the self is better liked than another one related with moderately positive, but self-irrelevant, adjectives.

**Study 4**

In the previous studies, we showed that personally relevant adjectives can be used as source to transfer the positivity of the self towards related brands, regardless of the valence of the adjectives. These studies showed the possibility to transfer positivity when the two classes of adjectives (either referring or not referring to the self) were either both positive (Study 1) or negative (Studies 2 and 3). The effect of this personality SR task is driven by whether the stimuli used as source where either relevant or non-relevant for one's self. Moreover, study 3 clarified that the observed effect when moderately negative adjectives are used cannot be explained by a self-serving bias in the selection of the adjectives. Study 4 pushes further the idea that stimuli self-relevance, rather than valence, is the crucial aspect that determines the SR effect. Rather than using self-relevant or non-relevant stimuli of equal valence, we tested whether adjectives with reference to the self may induce transfer of positivity even when they are moderately negative, while self-irrelevant adjectives are moderately positive. We asked participants the rate a list of adjectives, selected those with either moderately positive or negative scores, and asked to select five self-relevant among the moderately negative adjectives and five irrelevant among the moderately positive one. Compared to previous studies, this is the first in which the relation between self-relevant and non-relevant stimuli is unbalanced with respect to their valence. Crucially, this unbalanced relation should be in favor of the self-irrelevant stimuli. Thus, to observe a pattern of findings that is consistent to the one observed in previous studies, the self-relevance of the adjectives must reverse the initial valence asymmetry. Only if this pre-condition is satisfied we can hypothesize moderately negative self-referent adjectives to induce greater liking towards the brand paired with them, compared to moderately positive but self-irrelevant adjectives.

**Method**

**Participants and Procedure**

One hundred and twenty-three students (77 women, 46 men, *M* age = 23.37, *SD* = 3.86) took part individually to a one-session study. The procedure was similar to the one used in Study 3. First, after giving informed consent, participants completed an adjectives rating task, a learning task, then an adjectives selection task. All participants then took the Self-Referencing task. For half of the participants, the manipulation consisted of pairing one brand with moderately negative self-relevant adjectives and another one with moderately positive self-irrelevant adjectives, whereas for the other half it consists of pairing another brand with moderately negative self-relevant adjectives. At that point, all participants completed the intersecting regularities memory measure, followed by an IAT, an explicit evaluation of the brands, and a measure of identification. Then participants completed an evaluation of all the adjectives that received a rating of 3 or 4 in the adjectives rating task done at the beginning of the session. Finally, participants completed demographics question, and the experimenter thanked, debriefed, and gave them course credit.

**Materials**

The manipulation and measures used in the study were identical to those used in Study 3. We slightly modified the list of adjectives used in the adjectives rating task in order to maximize the chance to obtain two lists of moderately positive and negative adjectives for the selection task.

*Adjectives rating task*. Participants rated a list of 80 adjectives on 6-point scales from 1 (very negative) to 6 (very positive, see Supplementary Materials Chapter 2 – Appendix D). The list included both moderately positive and moderately negative adjectives. Adjectives equal to 3 were selected by the computer system in a new list, while those with a rating equal to 4 were selected in a second list.

*Adjectives selection task.* All participants were asked to choose from the list containing moderately negative adjectives (i.e., rating of 3) the five adjectives that describe them most and to choose from the list containing moderately positive adjectives (i.e., rating of 4) the five adjectives

that describe them the least (not characteristic of them at all, see Supplementary Materials Chapter 2 – Appendix D).

**Results**

The data from six participants were excluded from the analyses due to high error rate ($> 25\%$) in either the IAT or the SR task. IR memory screening criterion lead to a final sample of 97 participants (see Supplementary Materials Chapter 2 – Appendix C for analyses on the full sample). Both the IAT and the explicit measure showed good internal consistency ($\alpha = .87$ and $\alpha = .84$, respectively). Statistics are reported in Table 4. One-way-ANOVAs revealed the SR manipulation to be significant on implicit and explicit attitude towards and on identification with the two brands. Participants who categorized one brand with self-related moderately negative adjectives showed greater implicit and explicit liking for and identification to that brand, compared to a contrast brand categorized with self-irrelevant moderately positive adjectives. The three dependent variables were positively and significantly correlated (see Table 5).

We looked at the evaluation of personality adjectives selected. All the self-relevant adjectives received the same moderately negative rating score (i.e., 3 on a scale from 1 to 6) whereas all the self-irrelevant adjectives received the same moderately positive rating score (i.e., 4 on a scale from 1 to 6). Therefore, we compared a) the post-SR scores of the self-relevant adjectives with the baseline score (i.e., 3), b) the post-SR scores of the self-irrelevant adjectives with the baseline score (i.e., 4), and c) the post-SR score between the self-relevant and the self-irrelevant adjectives. Self-relevant adjectives received significantly greater liking after the SR manipulation, $t(96) = 8.43$, $p < .001$, whereas self-irrelevant adjectives received significantly lower liking, $t(96) = -12.27$, $p < .001$. Importantly, a paired-sample t-test revealed that, after being attributed to the self, moderately negative adjectives no longer differed in liking from the self-irrelevant moderately positive ones, $t(96) = -.07$, $p = .222$.

We conducted moderation analyses to test whether the SR effect on implicit attitude, explicit attitude, and identification was qualified by participants' evaluation of the self-referent adjectives. All the interaction terms were not significant ($p$s > .251).

Study 4 showed that the target brands categorized through the same action as moderately negative self-relevant adjectives are implicitly and explicitly preferred and receive higher identification compare to brands paired with moderately positive self-irrelevant adjectives. This happens because of the higher reference of the former ones with the positive category self. Results from this study are important as they provide compelling evidence of fact the reference to the self of the stimuli used in the SR task is more important than their valence. We showed that the self-relevance of the stimuli was capable to reverse the initial asymmetry in valence between the source stimuli, which was in favor of the self-irrelevant (moderately positive) traits. Once again, after being selected as representative of one's personality, the adjectives increased in liking. However, the difference between moderately positive and negative adjectives was not significant, revealing no preference for the self-relevant traits, but instead an overall comparable evaluation between self-relevant and irrelevant stimuli. Therefore, this finding showed that the self-serving post-SR evaluative bias cannot explain the observed SR effect. In other words, the SR effect must be driven by self-relevance, no matter whether there is a self-serving evaluative bias, hence over and beyond stimuli valence even after considering the post-selection self-serving evaluative bias.

**General discussion**

Across four studies, we showed that target brands paired with personality adjectives that refer to the self-concept are better liked than those paired with self-irrelevant personality adjectives. This self-referencing effect occurs irrespective of the valence of the stimuli used as source in the SR paradigm. In particular, when two lists of personality adjectives that were equally positive (Study 1) or negative (Study 2 and 3) in valence were presented in the same associative learning paradigm, those selected as representative of one's self transferred the acquired positivity to the objects categorized through the same action. Results from Study 4 strengthened this idea, showing that people

prefer and identify with the objects related to moderately negative stimuli, rather than those related to moderately positive ones, if the stimuli were relevant in describing their own personality. These findings show that when the self-concept is involved in associative learning paradigm, the valence of the stimuli used is not the key driver of the resulting evaluative effect. Rather, what really matters is the relevance of the stimuli for the self-concept.

Previous research in the context of implicit attitude measures showed that idiographic self-representations render IATs more suitable for the assessment of the implicit self-concept than do generic self-representations (Bluemke & Friese, 2012). It was observed that idiographic items used in the IAT allowed a better representation of invariant person cores, and second, they are more proximal to the core self than generic pronouns. Based on the logic that idiographic stimuli that are close to the self-concept more easily activate the self, we tested their impact in an associative evaluative learning paradigm. We developed the importance of stimuli self-relevance on an associative evaluative learning procedure that uses the self as source of attitude change (i.e., the SR task). We showed that when the transfer of liking towards a target object is driven by the self, the relevance of the stimuli for the self-concept is more important than their valence. These findings support the idea that the self represents a special attitudinal source which goes beyond the valence of its stimuli in determining evaluative learning effects. Previous studies in the context of the SR research provided evidence of the fact that valence is not the only characteristic that drives the SR effect (Perugini, Zogmaister, Richetin, Prestwich, & Hurling, 2013; Perugini, Richetin, & Mattavelli, 2016). However, all previous studies merely focused on the valence of the sources, showing a superior liking for self-related objects even when the contrast category was either equally or more positive than the self. None of the previous contributions tried to explore what factors, other than valence, might drive the SR effect. This was the first attempt to show that an additional mechanism, that is, self-relevance, can account for the SR effect over and beyond valence. In particular, the fourth study clarifies the idea that when the impact of valence and self-relevance on evaluative learning is tested within the same study, the latter drives the effect.

In line with the importance of stimuli self-relevance, we also want to emphasize that the effect of this personality SR task was much stronger in shaping implicit and explicit attitudes as well as identification towards target objects than the standard SR task is. We meta-analyzed the results of the present research and compared them with meta-analytical findings from twenty-six studies (eight of which also measured identification) conducted in our lab using the standard SR task (with possessive pronouns as stimuli for the categories Self and Others) on brands. The personality version of the SR task showed impressively stronger effects than the standard version in changing implicit attitudes (Cohen's $d = 1.00$, $SE = .107$ vs. $d = .731$, $SE = .055$, respectively), explicit attitudes (Cohen's $d = .730$, $SE = .104$ vs. $d = .350$, $SE = .051$, respectively), as well as on identification (Cohen's $d = .832$, $SE = .105$ vs. $d = .268$, $SE = .090$, respectively). Thus, the four studies indicate that products' liking and identification increase when they share something with one's specific personality characteristics. This might be due to the overlapping between ourselves and an object (e.g., a brand) when the latter possesses something we also possess or is similar to how we look like.

In more general terms, these findings are also important from the perspective of the underlying processes of associative learning via the self. What we observed can be considered as an indirect form of Self-Referencing effect. The IR principle states that a transfer of valence occurs among a valenced and a neutral object that share a common feature. If we rely on this definition of IR, we should observe no preference for either brands in the first three studies, given that the valence of the stimuli sharing an action with the brands was equal. Even more important, the IR principle would predict a preference for the brand categorized with moderately positive traits in study 4. Conversely, both attitude change and identification are driven by an indirect relation established between the brands and either the self or others. In our view, such an indirect SR effect requires one's ability to construe relations between observed source-stimuli and their higher-order category. Therefore, we believe that a propositional account (De Houwer, 2009; Mitchell, De Houwer, & Lovibond, 2009) offers the most suitable explanation for the current results. Therefore, the formation of propositions, described as "*statements about the world that contain information about the way in which concepts are related*" (De Houwer,

2014, p. 246) may underlie the observed effect. In particular, the fact that participants constantly showed a preference for target objects paired with self-relevant adjectives seems to imply their ability to form relations between source-stimuli and their core category (e.g., *"The adjective 'shy' relates to myself"*) and some sort of *if-then* validation process upon these relationships is underway (e.g., "*If being shy relates to myself and to the brand X, then the brand X relates to me*"). More important, the effect was consistently in favor of the self on all the three outcome measures considered in the studies (i.e., implicit attitude, explicit attitude, and identification). The observed effect on identification and explicit attitudes is predicted by a dual-model account of associative learning (Gawronski & Bodenhausen, 2011) since both the outcomes should tap more on participants' ability to form and validate the relationship between target objects, source stimuli, and the category that these stimuli represented. Conversely, such an effect on implicit attitude change is of more difficult interpretation if one adopts the perspective of a dual-model account. Based on dual-model theories of attitudes, implicit attitude change resulting from such procedures should reflect a simple associative process that does not imply any higher-order reasoning upon the nature of the stimuli and their relations. Considering the present findings, the effect of the SR driven by stimuli self-relevance seems more in line with a propositional interpretation of implicit attitude change. Having said that, we refrain to claim that the same mechanisms underlies the observed effects on implicit and explicit attitude change. We merely propose the idea that both implicit and explicit attitude change might depend on one's ability to form relations between stimuli. At the same time, we acknowledge that the formation of propositions upon stimuli relations might involve further validation processes, which might differ depending on the attitudinal level.

Related to this point, the research also tested whether a self-serving bias occurred in the evaluation of the adjectives selected as relevant for the self. Moreover, we examined whether this overall evaluation of the self-relevant traits did qualify the SR effect on all the observed outcomes. Although not consistent across the four studies, the pattern of results offered by the moderation analyses provided some interesting indications. First, when moderately positive adjectives were used

for both the self and others (study 1) a change on explicit attitude occurred only when the evaluation of the self-adjectives was medium or high. Second, when personally relevant moderately negative adjectives were used for self (study 3), implicit attitude increased when the evaluation of the adjectives was medium or high, while explicit attitudes did so at low or medium evaluation of the self-adjectives. Taken together, these findings might suggest a) that the impact of stimuli self-relevance does not univocally impact implicit and explicit attitude change and b) that distinct motivational processes might determine explicit attitude change when either positive or negative traits are used as self-relevant stimuli. One could speculate that with moderately positive adjectives, people's explicit preferences for the brands follows the need to advance themselves through the newly related object. Instead, with moderately negative adjectives, the need to reaffirm one's positivity through the related object is more pronounced when self-evaluation is low or medium, rather than high. However, we acknowledge that further empirical evidence is needed to support any speculative reasoning on these findings, especially given the inconsistency of the results across studies (e.g., no moderation effect was observed in studies 2 and 4 on either outcome measure).

This contribution offered an attempt to study the role played by a specific facet of the self on associative learning. Considering the effect of the SR task we demonstrated that changes in liking and identification towards brands do not depend on intersecting regularities between valenced personality traits and the targeted brands, but rather on the relevance that these traits have for the self-concept. We look at these findings as an ancillary proof that evaluative learning via IR, at least when the self is involved, might result from propositional reasoning through which people infer stimuli relationships at both a direct (e.g., traits and brands relate due to the same action) and indirect (self and brands relate due to the common traits they relate with) level. Moreover, we showed a strong impact of self-relevant traits in increasing implicit attitudes, explicit attitudes and identification towards related brands. This effect was stronger than the standard SR effect, observed with personal pronouns as self-referent stimuli. Taken together, these findings highlight the importance of stimuli self-relevance in directing attitude (and identification) change via the SR task. They also offer

important indications in terms of the interpretation of the SR effect from a cognitive perspective and pave the way for further investigations to understand what kind of processes can better explain such an effect.

**Chapter 3.**

**Indirect attitude change within and across domain: Transferring self-positivity via multiple intersecting regularities.[10]**

---

[10] This chapter is based on Mattavelli, S., Richetin, J., & Perugini, M. (*manuscript in preparation).* Indirect attitude change within and across domain: Transferring self-positivity via multiple intersecting regularities.

**Overview**

In the second chapter, we demonstrated that the self can indirectly affect liking and identification with target objects paired with personality traits when these traits are self-relevant for the self-concept independently from their valence. In the present chapter, we introduced an alternative way through which the SR task can indirectly affect the evaluation of objects. We hypothesized that the IR principle can be exploited to spread the valence acquired by a previously neutral object over novel targets in more complex chain of intersections. We tested the transfer of liking generated through the Self Referencing (SR) task. In doing so, we administered a task where targets paired with the self versus others in the SR shared a new action response with novel targets. Throughout five studies ($N = 604$) we used the IR principle to test an indirect SR effect on implicit (IAT) and explicit attitude. Moreover, in this chapter we investigated whether this indirect transfer of liking is qualified by stimuli domain. Specifically, we relied on the idea that evaluative learning through IR occurs when people a) notice that there is a common element between valenced and neutral stimuli and b) use this information as a cue to infer stimuli relations (see De Houwer & Hughes, 2016). Unlike previous EC studies, which manipulated stimuli relations in an overt and explicit way (e.g., Moran & Bar-Anan, 2013; Zanon, De Houwer, & Gast, 2012; Zanon, De Houwer, Gast, & Smith, 2014), we used the (in)congruency of targets domain as a subtle cue to manipulate the relation between stimuli involved in the intersections chain. First, we tested the indirect SR effect within a same domain of objects, directed to groups via other groups presented in the SR (study 1) or to brands via other brands (study 2). Then, in study 3, 4, and 5, we tested the same effect when the domain of target objects varied across the two learning phases: brands were presented in the SR task and then categorized with groups in the additional task in Study 3, while the opposite flow was used in Study 4 and 5. We discuss the results in terms of intersecting regularities as a powerful mechanism through which liking can transfer towards multiple objects in series. Moreover, we discuss the importance of stimuli domain as a cue to infer a relationship between them.

Attitude change is any change in the evaluation of an object of thought, which includes the formation of new evaluations toward unfamiliar objects or the changes in a pre-existing attitude towards known objects. There is a general agreement in defining any attitude change as the affective or cognitive meaning that attitudinal objects may acquire within the context of either pleasant or unpleasant events. In persuasion studies, participants are typically exposed to a message containing attributes about an object and different type of peripheral cues, like the expertise of a source (see Chen & Chaicken, 1999; Petty & Cacioppo, 1986). In a classic Evaluative Conditioning (EC) study, liking or disliking an object (conditioned stimulus, CS) results from its repeated presentation in contingency of space and time with another object (unconditioned stimulus, US, see Hoffman, De Houwer, Perugini, Baeyens, & Crombez, 2010). Although they are often seen as distinct ways to affect attitudinal dispositions, both persuasion and EC rely on direct relationships established between valenced stimuli and the attitudinal target. Specifically, in persuasion there is a direct link between the positive/negative information and the target object (e.g., "Drink X is healthy, energetic and tasty"). In EC, there is a direct co-occurrence of stimuli, with positive/negative stimuli presented in spatio-temporal contiguity with the target object.

**Changing attitudes through the indirect route**

However, in many situations the formation of an attitude does not result from a direct relation between stimuli. Prejudice, for instance, is a form of negative attitude often directed to groups of people never encountered. Traits known about an individual are integrated into the group stereotype with subsequent application of that stereotype to unknown group members (Crawford, Sherman, & Hamilton, 2002). Attitude generalization through detection of similarities has been investigated through the "BeanFest" paradigm (Fazio, Eiser, & Shook, 2004) as well as in the context of faces evaluation (Verosky & Todorov, 2010; 2013). This indirect form of attitude change also applies to EC: Conditioned attitudes transfer to objects that are pre- or post-associated with the CS involved into a direct pairing with a US (Walther, 2002; Hammerl & Grabitz, 1996). Throughout five studies, Walther (2002) demonstrated a *spreading attitude effect* on visual stimuli. She adopted a two-steps

learning procedure: The repeated pairings of valenced faces (USs) with neutral faces (CS1) was either preceded or followed by a further phase in which pairings between CS1 and other neutral faces (CS2) were experienced. Depending on whether a CS1-CS2 pairing occurs before or after the actual US-CS1 pairing, the learning mechanism that mediates this indirect evaluative learning has been defined as either sensory preconditioning or second-order conditioning, respectively. As a result, CS2 acquired US valence even though the two classes of stimuli were never paired. Gast and De Houwer (2012) extended previous findings on measures of implicit attitudes by adopting a slightly different procedure where stimuli pairings are not directly experienced but merely instructed. After a US-CS pairing phase, participants were told that the CS presented actually covered another stimulus. Gast and De Houwer's procedure introduced the importance of learning stimuli relations to observe indirect EC effects. Both Walther (2002) and Gast and De Houwer (2012) showed that in either experienced or instructed combinations of stimuli pairings the recursive presentation of one stimulus serves as a cue to infer a link between objects, which ultimately results in indirect EC effects.

**Indirect attitude change and intersecting regularities**

Taken together, these forms of indirect attitude change seem to rely on people's ability to learn that objects are linked because they share characteristics (e.g., group identity, facial features, a CS). These intersections between stimuli work as the necessary precondition for the transfer of valence from the positive/negative object to the neutral one. This reasoning finds evidence in a novel way to operationalize relationship between stimuli, namely, Intersecting Regularities (IR, Hughes, De Houwer, & Perugini, 2016). IR is a pathway of evaluative learning that is not based on the spatio-temporal contingency between neutral and valenced stimuli which characterizes EC. Instead, the transfer of liking from a valenced object to a neutral one occurs thanks to a characteristic shared between the two, even though the objects are never paired. For instance, Hughes et al. (2016, study 1) asked their participants to perform a simple categorization task in which they used the same key (e.g., "F") when presented, with equal probability, a neutral brand name or positively valenced stimuli and a different key (e.g., "J") when presented another neutral brand name or negative valenced

stimuli. The regularities (i.e., pressing key upon stimulus presentation) regarding the first neutral brand and positive stimuli thus intersected in terms of response as well as the ones regarding the second neutral brand and negative stimuli. In other words, an intersection was established through a common response between the valence and the neutral stimulus in either of the two sets of stimuli (neutral brand and positively valenced vs. alternative neutral brand and negatively valenced). The results demonstrated increased implicit and explicit liking towards the brand sharing the response with the positive stimuli relative to the alternative brand. This mechanism also applies to many real-life situations. For instance, imagine Bob gets to know Mike at a party and finds out he is from a small town (first regularity) where Bob used to go on holiday with his family (second regularity). Due to this positive commonality between them, Bob will probably like Mike. The small town represents the intersection or the commonality between a positive event (e.g., the good time Bob has had in the past) and a neutral one (e.g., Mike) and allows a transfer of liking from the former to the latter.

Learning via intersecting regularities can also occur in a more complex and indirect fashion. Going back to the previous example, at the same party Bob also meets another person, John, and while chatting with him he realizes that John and Mike play in the same football team (third regularity). Now the whole picture involves a) a first commonality between Bob and Mike (i.e., the small town), b) another commonality between Mike and John (i.e., the football team) and connects them and c) a resulting commonality between these two chains of connections (i.e., Mike) that in turns links Bob to John. Hughes et al. (2016, study 4) tested this idea by designing another similar learning task after the one described above, where objects previously categorized through the same action as valenced objects shared a new response with a novel target. The authors observed a first evaluative learning effect due to IR within a single learning phase (e.g., pressing the same key in a task that requires to categorize a valenced object and target 1) and a second evaluative learning effect that arose as a function of IR within the second phase (e.g., doing an unrelated action to categorize target 1 and another target). Furthermore, the combination of the two phases led to an indirect

intersection of regularities across the two learning phases (i.e., target 1 is the commonality between the valenced stimulus and a new target).

**The self and associative attitude change**

Previous research on both ownership and EC showed that objects that relate to the self are better liked (Gawronski, Bodenhausen, & Becker, 2007; Gawronski & LeBel, 2008, study 3; Walther & Trasselli, 2003; Zhang & Chan, 2009). Greenwald, Banaji, et al. (2002) posited that self-object relationships may occur through the development of balanced triads of associations. According to their *balance-congruity principle*, two unlinked objects in memory (e.g., valence and brand) that share "first-order" links with a third object (e.g., self-concept) should develop a mutual association. Assuming an a priori positive link between self and valence (Yamaguchi et al., 2007) conceptualized as implicit self-esteem (Greenwald & Banaji, 1995), an association created between the self-concept and an object in the environment should produce a new link between the object and that valence. To the extent that implicit self-esteem is positive for most people (Yamaguchi et al., 2007), it is thought that this newly formed self-object associations should be on average positive in nature. It remains however unexplored whether the positivity can spread to objects that do not share any first order link with the self. In other words, we tested whether a balance congruity principle also applies to quartets of associations where an additional object is directly linked to the one linked to the self.

We used an evaluative learning task based on the IR mechanism to transfer the positivity carried by the self, namely, the Self Referencing task (SR, Prestwich, Perugini, Hurling, & Richetin, 2010). The SR task is a recently developed evaluative learning paradigm that a) uses the self as a positive source to generate attitude change towards objects and b) works via IR. Throughout this categorization task, participants are required to categorize with the same action (e.g., pressing the 'E' key) self-related words and stimuli belonging to a first target category, while a second action (e.g., pressing the 'I' key) is to categorize other-related words and a second target category. The task is then based on two pairs of operant contingencies (Pair 1: *"If self, press 'E'"* and *"If Target A, press 'E'"*; Pair 2: *"If other, press 'I'"* and *"If Target B, press 'I'"*) that intersect at the level of the

performed action. The mere act of pressing the same key for categorizing self-stimuli and a target category leads to a transfer of positivity. A meta-analysis conducted on 53 studies (Mattavelli, Richetin, Gallucci & Perugini, 2016) showed that the paradigm leads to genuine attitude change at both implicit and explicit level (Cohen's $d = .69$ and Cohen's $d = .43$, respectively). The effect of the SR task on implicit and explicit attitude change has been shown on fictitious social groups (Perugini, Richetin, & Zogmaister, 2012) and brands (Perugini, Zogmaister, Richetin, Prestwich, & Hurling, 2013; Richetin, Mattavelli, & Perugini, 2016), among others. However, whether this positivity can transfer to objects that are not directly categorized in the SR task, but presented to participants in a later phase, remains an unexplored empirical question. In the present chapter we combined previous initial evidence of indirect attitude change via multiple intersecting regularities (Hughes et al., 2016, study 4) with the SR effect to investigate the indirect transfer of liking carried by the self on both implicit and explicit attitudes.

**Domain congruency as a stimuli relational cue**

Besides testing indirect associative learning via multiple intersecting regularities, we also assessed the role played by specific characteristics of the stimuli involved in the intersection chain in qualifying the effect. Specifically, this contribution focused on the congruency in stimuli domain. We developed in a novel and indirect way the recent idea that information about the relation between stimuli moderates the effect of evaluative learning through associative procedures. Prior EC research showed that information about the relation between USs and CSs (e.g., informing participants that a US and a CS are either friends or enemies) can alter the impact of stimuli pairings on explicit evaluations (e.g., Fiedler & Unkelbach, 2011; Förderer & Unkelbach, 2012; Gawronski, Walther, & Blank, 2005; Moran & Bar-Anan, 2013; Zanon, De Houwer, & Gast, 2012; Zanon, De Houwer, Gast, & Smith, 2014) with mixed evidence on implicit attitude (Hu, Gawronski, & Balas, 2016). However, in all these studies stimuli relations were overtly presented to participants by explicit informing them about the nature of the relation. No studies have tested whether stimuli relationships can also impact these effect when they are merely inferred through specific characteristics of the stimuli. We used

congruency in stimuli domain as a cue to infer a given relation between stimuli. Specifically, we tested how indirect evaluative learning via IR varies according to the (in)congruency in the domain of stimuli involved in the transfer of valence. Thus, in two initial studies we tested the indirect effect directed either to groups via other groups presented in the SR (study 1) or to brands via other brands (study 2). Target objects differed between the two studies but remained constant across the two learning phases of each study. Conversely, study 3, 4 and 5 tested the same effect when the target objects varied across the two learning phase: brands were presented in the SR task and then categorized with groups in the additional task in Study 3, while the opposite flow was used in Study 4 and 5. The underlying idea is that manipulating stimuli domain could determine variations at the level of the inferred relations between stimuli themselves. In essence, if IRs are meant to form relationships between objects, these relationships may acquire distinct meanings depending on a) the type of stimuli involved and b) whether these relationships are between stimuli of either the same or distinct domains.

**The present research.**

Five experiments were designed to test indirect evaluative learning driven by the self by using complex intersections between regularities by adding another learning procedure after the SR task. During this additional task, objects previously related with self (Target 1) were categorized through the same action as a new neutral target (Target 3), while objects previously paired with the category other (Target 2) shared a common response with a contrast target (Target 4). In other words, the two pairs of operant contingencies in the SR task (Pair 1 *"If Self, press 'I'"* and *"If Target 1, press 'I'"*; Pair 2: *"If Others, press 'E'* and *"If Target 2, press 'E'"*) and the two pairs of operant contingencies in the new IR-based learning procedure, (Pair 3: *"If Target 1, press 'I'"* and *"If Target 3, press 'I'"*; Pair 4: *"If Target 2, press 'E'* and *"If Target 4, press 'E'"*) intersect at the level of the stimuli class (i.e., Target 1 is the commonality between the self and Target 3, Target 2 is the commonality between other and Target 4). The sequential intersections established across the two learning phase are meant to spread the positivity of the self to the novel object (Target 3).

First, we tested the effect between stimuli of the same domain. Then we varied stimuli domain within the same study, also changing the order in which stimuli intersect each other throughout the experimental procedures. According to a propositional account of associative evaluative learning, we expected indirect SR to be weaker in the condition where the transfer occurred between stimuli of distinct domains, given the higher complexity in the propositional reasoning required to qualify and validate their relation.

If one assumes that relationships are important to qualify evaluative learning effect, focusing on participants' ability to detect and qualify relational cues is a central aspect. In our specific case, associative learning via IR can be affected by one's ability a) to recall in memory these regularities between stimuli and b) to qualify these intersections in relational terms. Related to the first point, a recent meta-analysis of the SR effect (Mattavelli et al., 2016) showed that participants' ability to learn and recollect IR in the SR task is of crucial importance for the effect to occur. Whether the occurrence of the hypothesized indirect SR effect is conditional to participants' ability to learn and recollect that two stimuli share a common element is verified in the present work by comparing indirect SR effect for participants who were either correct or incorrect in remembering intersection regularities between stimuli. Concerning the second issue, for one of the studies we tested whether this ability to notice and remember intersecting regularities between stimuli is sufficient for the occurrence of the effect, or whether some higher-order reasoning is required. To this aim, we administered Matching To Sample procedure (Dymond & Whelan, 2010) that measured one's ability to detect equivalences between the stimuli categorized in the two learning phases.

**Targets selection**

Four fictitious social groups and their members (i.e., Lerriani [Target1], Dattiani [Target2], Craviani [Target3] and Zimmiani [Target4]) where used in study 1. All the groups have been used in previous SR studies (e.g., Perugini et al., 2012; Richetin et al., 2016). Four fictitious brand logos with an embedded brand name (i.e., Rama [Target1], Moniso [Target2], Lestea [Target3] and Sabea [Target4]) were used as targets in Study 2. Brands were pre-selected out of a list of twelve candidates.

From two pretests, we selected the most neutral pairs in term of valence, at both implicit (IAT) and explicit (semantic differential on name, logo, products, and brand in general) level. There was no difference (all $p$s > .184) on explicit preference for both pairs. At the implicit level, we observed a significant preference for brand1 over brand2 ($p$ = .001)[11] and no significant difference for brand3 and brand4 ($p$ = .155). In study 3, 4 and 5 we used Lerriani and Dattiani as groups and we created two new brands by slightly modifying Lestea and Sabea in terms of their logos and brand names. The two new brands were named Gester and Vabel (this was done to avoid any overlap in target names: Lerriani could have been perceived as more similar to Lestea than Dattiani). These two novel brands were pretested in term of valence on a 9-point scale, among a set of other thirteen exemplars. Neither Gester nor Vabel were evaluated as significantly different from Lestea and Sabea, $t(24)$ = .12, $p$ = .902 and $t(25)$ = -.29, $p$ = .774, respectively. Moreover, Gester and Vabel were not different from each other in their evaluation, $t(27)$ = .62, $p$ = .538 (see Supplementary Materials Chapter 3 – Appendix G for stimuli used in the studies).

**Study 1**

**Method**

**Participants and procedure**

One hundred and twenty-one participants (University students, 92 women and 29 men, $M$age = 23.12, $SD$ = 3.98) took part to this study. We adopted a 2 (SR task condition: Target1 paired with the self vs. Target2 paired with the self) x 2 (Target Categorization task condition: Target3 paired with Target1 vs Target4 paired with Target1) between subjects. Participants were to imagine that four fictitious target groups actually existed. The experimenter informed participants they would be later asked questions about them. They then completed a simple instruction check task and the SR task.

---

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

Half of participants used the same key to classify members of Target1 and words related to the self, whereas the other half used the same key for Target2 and self. After the SR task, participants answered an Intersecting Regularities memory question and subsequently completed an Implicit Association Test (IAT, Greenwald, McGhee, & Schwartz, 1998) and an explicit attitude measure to assess respectively their implicit and explicit attitudes towards Target1 and Target2. Participants then completed a second learning task. For half of them, exemplars of Target1 were to be categorized with the same key as those of Target3, while exemplars of Target2 were categorized with the same key as those belonging to Target4. For the other half, the established relationships were Target1–Target4 and Target2–Target3. A second Intersecting Regularities memory question was answered after the task, followed by another IAT and an explicit evaluation to assess the implicit and explicit attitudes towards Target3 and Target4. Participants were thanked, debriefed, and paid or given course credit.

**Materials**

*Instructions check task.* To make sure participants read and understood the description of the four target categories, we introduced a manipulation check immediately after the presentation of the targets. It was a twelve-trials categorization task in which targets' exemplars (three for each target) were presented individually in the middle of the screen. The four targets category labels remained visible aligned on the top part of the screen. The inter-trial interval was 400ms. Participants' task was to correctly assign each exemplar to the corresponding target category by clicking with the mouse on the target label. An 80% of correct responses rate was considered as threshold to move to the SR task. Otherwise participants had to read again the description of the four groups.

*SR task.* Participants completed two blocks of 40 trials that required them to categorize, as quickly and accurately as possible, 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"). Participants then 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 case of incorrect classification, a red-X appeared on 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 (see Supplementary Materials Chapter 3 – Appendix G).

*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".

*Target Categorization task.* The task mirrored the general structure of the SR task. In one of the two conditions, participants had to press the same key to categorize stimuli belonging to either the group previously related to the self (i.e., Target1 vs. Target2 depending on the SR condition) or to Target3 and another key for members belonging to either the group previously related to others or to Target4. In the other condition the task was the opposite: Target4 stimuli were categorized through the same action as the stimuli belonging to the category previously related to the self, while Target3 stimuli went with the object prior related with others. The task consisted of 160 trials, divided in four blocks of 40 (2 practice blocks and 2 for testing). Group exemplars appeared randomly and individually in the middle of the screen. Participants indicated, as quickly and accurately as possible, the group to which a member belonged by pressing one of two colored keys. The key used to categorize members was reversed between the two blocks. In case of incorrect classification, a red-X appeared on screen and remained until correction. The inter-trial interval was 400ms. For each target, 5 stimuli names were used (see Supplementary Materials Chapter 3 – Appendix G).

*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 3 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. Five positive (e.g., *positive, joy, happy, smile, nice*) and five negative (e.g., *negative, pain, bad, sad, cry*) stimuli were used. 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 of 500ms. For each attribute and target category, we used five stimuli (see Supplementary Materials Chapter 3 – Appendix G). We calculated the D score following the procedure suggested by Greenwald, Nosek, & Banaji (2003), such as higher scores indicated a preference for Target1 over Target2.

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

*Explicit evaluation Target1/Target2.* Participants evaluated each target separately on 4 dimensions (negative/positive, annoying/friendly, bad/good, unpleasant/pleasant) on a 7-point Likert-type scale. The order of the evaluation of the two targets was counterbalanced and nested with the order in which the critical blocks were administered in the IAT (Target1 first when first presented with positive in the IAT vs. Target2 first when first presented with positive). The average score across the four dimensions was calculated such that higher scores indicated a preference for Target1.

*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 Target3.

### Analysis plan and data preparation

For the standard SR effect, we excluded data from participants who made over 25% of errors in at least one of the classification tasks (SR Task and IAT1), leading to a final sample of 118 participants. We followed a similar reasoning for the analysis of the indirect SR effect. At this level, no participants made over 25% of errors in either the Target Categorization task or IAT2. As introduced in the method section, we assessed participants' ability to recollect the regularities that linked the target and the self through the IR memory question. Given the particular design adopted in the present research, we reasoned that an additional IR memory question should be also administered at the end of the Target Categorization task. The analysis of the standard, as well as the indirect, SR effect are reported by considering participants who answered correctly IR memory question 1 (105, 89%) and both IR memory 1 and 2 (101, 85.6%) (see Supplementary Materials Chapter 3 – Appendix F for the analyses on the full samples).

## Results and discussion

To test the effect of SR manipulation on IAT group1/group2, we ran a one-way ANOVA, with SR condition (group1-Self vs. group2-Self). The reliability of the IAT was good ($\alpha = .86$). Means and standard deviations are reported in Table 6. The SR manipulation significantly affected implicit attitude towards the two groups, $F(1,104) = 8.32$, $p = .005$, $\eta_p^2 = .07$. Reliability was also good for the explicit measure ($\alpha = .96$). A one-way ANOVA revealed a significant effect of SR manipulation on the explicit attitude, $F(1,103) = 13.98$, $p < .000$, $\eta_p^2 = .12$. The group paired with the self resulted more positive than the group paired with others, both at the explicit and implicit level. Both the IAT and the explicit measure administered to test indirect SR effect were highly reliable ($\alpha = .80$ and $\alpha = .98$, respectively).

Table 6. *Direct effect of SR condition on implicit and explicit attitude for each study*

| | Target 1 | | Target 2 | | | | Cohen's *d* |
|---|---|---|---|---|---|---|---|
| | Self | | Self | | *F* | *p* | [95% CI] |
| | *M* | *SD* | *M* | *SD* | | | |
| Implicit attitude | | | | | | | |
| Study 1 | 0.07 | 0.45 | -0.20 | 0.51 | 8.32 | .005 | 0.56 [0.17, 0.96] |
| Study 2 | 0.38 | 0.48 | 0.01 | 0.46 | 16.65 | <.001 | 0.79 [0.40, 1.18] |
| Study 3 | 0.01 | 0.47 | -0.47 | 0.42 | 29.59 | <.001 | 1.07 [0.65, 1.49] |
| Study 4 | 0.27 | 0.52 | -0.15 | 0.48 | 17.11 | <.001 | 0.84 [0.42, 1.25] |
| Study 5 | 0.21 | 0.45 | -0.17 | 0.42 | 16.91 | <.001 | 0.85 [0.45, 1.23] |
| Explicit attitude | | | | | | | |
| Study 1 | 0.60 | 2.29 | -1.32 | 2.91 | 13.98 | <.001 | 0.73 [0.33, 1.13] |
| Study 2 | 0.53 | 1.08 | -0.01 | 1.25 | 5.95 | .016 | 0.46 [0.08, 0.85] |
| Study 3 | 0.11 | 1.64 | -0.83 | 1.42 | 9.43 | .003 | 0.61 [0.21, 1.01] |
| Study 4 | 0.99 | 2.15 | -0.17 | 2.23 | 7.19 | .009 | 0.53 [0.13, 0.93] |
| Study 5 | 0.79 | 2.16 | -0.71 | 1.44 | 14.44 | <.001 | 0.78 [0.36, 1.16] |

To test the indirect effect, the four conditions were recoded depending on the pairing between targets in the two tasks (targets paired with targets paired with the self vs. targets paired with targets paired with others in the SR task). For both implicit and explicit attitudes, higher scores indicated a preference for Target3. Means and standard deviations are reported in Table 7. A one-way-ANOVA showed a significant effect on IAT score, $F(1,98) = 5.34$, $p = .023$, $\eta_p^2 = .05$ and on explicit attitude, $F(1,98) = 4.52$, $p = .036$, $\eta_p^2 = .04$, demonstrating an indirect SR effect.

Results from study 1 showed that the positivity of the self can transfer across groups that relate with it via either direct or indirect intersecting regularities. The indirect SR effect was clear at both the implicit and the explicit attitude levels.

Table 7. *Indirect effect of SR condition on implicit and explicit attitude for each study*

| | Target 3 Self-Target | | Target 4 Self-Target | | *F* | *p* | Cohen's *d* [95% CI] |
|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | | | |
| Implicit attitude | | | | | | | |
| Study 1 | 0.29 | 0.44 | 0.09 | 0.44 | 5.63 | .020 | 0.46 [0.06, 0.86] |
| Study 2 | -0.16 | 0.47 | -0.37 | 0.47 | 5.43 | .022 | 0.45 [0.05, 0.85] |
| Study 3 | 0.24 | 0.48 | -0.02 | 0.53 | 6.13 | .015 | 0.51 [0.08, 0.95] |
| Study 4 | -0.26 | 0.45 | -0.33 | 0.44 | 0.51 | .477 | 0.16 [-0.25, 0.56] |
| Study 5 | -0.28 | 0.44 | -0.39 | 0.44 | 1.18 | .280 | 0.23 [-0.18, 0,64] |
| Explicit attitude | | | | | | | |
| Study 1 | 0.40 | 2.81 | -0.80 | 2.74 | 4.71 | .032 | 0.43 [0.03, 0.83] |
| Study 2 | -0.29 | 1.55 | -0.84 | 1.74 | 2.80 | .097 | 0.33 [-0.07, 0.73] |
| Study 3 | 0.40 | 1.96 | -0.27 | 2.05 | 2.36 | .128 | 0.33 [-0.10, 0.76] |
| Study 4 | 0.20 | 2.21 | -1.01 | 1.97 | 7.84 | .006 | 0.58 [0.18, 0.98] |
| Study 5 | -0.86 | 1.96 | -0.99 | 2.03 | 0.09 | .770 | 0.06 [-0.35, 0,46] |

This study stands for the first empirical demonstration that we can spread the positivity of the self towards target stimuli that have never been either presented or categorized with self-stimuli. Rather, we argue that the presence of intersecting regularities between the two, represented by an object belonging to the same domain as the final target of change, is responsible for this transfer of valence.

**Study 2**

In this second study, we aimed at testing whether indirect SR can also be demonstrated on a different domain (i.e., brands rather than group). Replicating prior findings on a different class of stimuli is important for two main reasons. First, it allows us to test whether indirect SR effects are

qualified by the domain of the stimuli used across studies. Second, it provides a clear picture of indirect SR on both groups and brands when the experimental procedure involves only stimuli of the same domain, to be compared later on with indirect effects on the same two targets objects when stimuli domain is mixed within the same study.

**Method**

**Participants and procedure**

One hundred and twenty University students (76 women and 44 men, $M$age = 22.69, $SD$ = 3.69) took part to this second study. The procedure, the method, and the materials were identical to those used in Study 1. Instead of groups, four fictitious brands (Rama, Moniso, Lestea, and Sabea) were presented as target categories.

**Results and discussion**

For the standard SR effect, we excluded data from participants who made over 25% of errors in at least one of the classification tasks (SR Task and IAT1), leading to a final sample of 120 participants. For the indirect effect of SR, no participants made over 25% of errors in either the Target Categorization task or IAT2. One hundred and eleven (92.5%) participants answered correctly IR memory question 1 and one hundred and one (85.6%) did so for both IR memory 1 and 2 (see Supplementary Materials Chapter 3 – Appendix F for the analyses on the full samples).

Both the IAT and the explicit measure were reliable (α = .86 and α = .82, respectively). Means and standard deviations are reported in Table 6. The SR manipulation significantly affected implicit attitude towards the brand1 opposed to brand2, $F(1,109) = 16.65$, $p < .001$, $\eta_p^2 = .13$ as well as the explicit attitude , $F(1,109) = 5.95$, $p = .016$, $\eta_p^2 = .05$. Similar to what was found in Study1, there was an implicit and explicit preference for the brand paired with the self over the brand paired with others.

Both the IAT and the explicit measure administered to test indirect SR effect were highly reliable (α = .88 and α = .91, respectively). Means and standard deviations are reported in Table 7. The indirect effect of the SR task was significant on implicit attitude, $F(1,99) = 5.43$, $p = .022$, $\eta_p^2 = $

.05, and with a tendency towards a significant effect on explicit attitude, $F(1,99) = 2.80$, $p = .097$, $\eta_p^2$ = .03.

Taken together, study 1 and study 2 both show that the SR can spread towards novel targets that are paired with those presented in the task. This transfer of liking occurs via multiple intersecting regularities. The fact that the indirect effect on the explicit attitude was only a tendency towards significance in Study 2 might be due to the generally weaker SR effect on explicit attitude change towards brands (see Mattavelli et al., 2016). In light of the observed results, we ran a micro meta-analysis of the two studies, focusing on the indirect SR effect on both implicit and explicit attitude. On the implicit attitude we found a mean effect size ($d$) of 0.447, $SE = 0.141$, 95% CI [0.170, 0.724], significantly different from zero, $Z = 3.16$, $p = .002$, and non heterogeneous, $Q$ (1) = .001, $p = .977$. On the explicit attitude, the overall effect ($d$) was 0.379, $SE = 0.141$, 95% CI [0.102, 0.656], significantly different from zero, $Z = 2.68$, $p = .007$, and non heterogeneous, $Q$ (1) = .125, $p = .724$. The non-significant heterogeneity is important in that reveals that the indirect SR effect occurred irrespectively of the type of target object used in the studies.

According to an associative perspective, one might hypothesize that once relationships among stimuli are established, the positivity carried by the self automatically spreads over the nodes of the newly created network. If this reasoning is true, these relationships should suffice to spread the positivity of the self over novel nodes in the network regardless of their nature. From a propositional account however, different stimuli might imply different relationships among them and, in this specific case, among them and the self. For instance, when a group [brand] is the intersection between the self and another group [brand], this IR might serve as a mere similarity cue: the two groups [brands] are similar as they have something in common. However, this similarity cue could make less sense when the positivity of the self spreads across stimuli belonging to distinct domains. If a group [brand] is the intersection between the self and a brand [group], inferring a similarity between them can be meaningless (e.g., why should a group be similar to a brand?). It is therefore plausible that when it comes to transfer the positivity of the self across stimuli belonging to different domains,

people might need to translate the new links into different propositional relationships. Therefore, higher-order propositions that qualify stimuli relationships as well as the validation of these propositions are crucial for indirect SR to occur. The next studies address this point by testing indirect SR generated across stimuli domain. Besides, they introduce a potentially important element, testing whether the order in which the relations between the self and the two classes of stimuli (i.e., groups and brands) affects the spreading across domain. In fact, if learned relationships between stimuli are processed in a propositional way, relating the self with either groups or brands first might result in distinct propositions, which might in turn having differential impacts in qualifying the indirect SR.

**Study 3**

In this study, we introduced stimuli domain incongruence as a potential boundary condition of the indirect effect of the SR. We tested indirect SR effect from a self-paired brand to a group. In the SR task the self was to be categorized with the same action as a brand (and the category others with another brand). In the Target Categorization task these brands (i.e., Gester [Target1] and Vabel [Target2]) were categorized through the same action as two groups (i.e., Lerriani [Target3] and Dattiani [Target4]).

**Method**

**Participants and procedure**

One hundred and twenty-one University students (91 women, 29 men, one missing, $M$age = 24.27, $SD$ = 8.36) took part to the study. The experimental design was identical to that used in previous studies, as well as the flow of the experimental procedure, the only variation being the use of brands as target objects in the SR task, subsequently paired with groups in the TC task.

**Results and discussion**

For the direct SR effect, the exclusion of participants who made over 25% of errors in at least one of the classification tasks (SR Task and IAT1), led to a final sample of 121 participants. The further screening after the TC task and IAT2 led to a final sample of 118 participants. The analysis of the direct as well as the indirect SR effect are reported by considering participants who answered

correctly IR memory question 1 (Direct SR effect: 102, 84.3%) and both IR memory 1 and 2 (Indirect SR effect: 87, 73.7%) (see Supplementary Materials Chapter 3 – Appendix F for the analyses on the full samples).

Both the IAT and the explicit measure used to assess attitudes towards brands were reliable ($\alpha = .88$ and $\alpha = .90$, respectively). The SR manipulation significantly affected the implicit attitude towards the two brands, $F(1,100) = 29.59$, $p < .001$, $\eta_p^2 = .23$ as well as the explicit attitude, $F(1,100) = 9.43$, $p = .003$, $\eta_p^2 = .08$. Means and standard deviations are reported in Table 6.

Both the IAT and the explicit measure were reliable ($\alpha = .90$ and $\alpha = .95$, respectively). For the indirect SR effect, a one-way-ANOVA showed a significant effect on the implicit attitude, $F(1,85) = 6.13$, $p = .015$, $\eta_p^2 = .07$. Conversely, and different to what observed in study 2, the effect was not significant on the explicit attitude, $F(1,85) = 2.36$, $p = .128$. This indicates that people like implicitly, but not explicitly, the group related to the brand previously related with the self. Means and standard deviations are reported in Table 7.

In this study, we found a significant indirect SR effect on implicit attitude change when the positivity of the self transfers from brands to groups. At the explicit level, indirect SR failed to reach the standard level of significance, although the means were in the predicted direction. As far as we are aware, this study stands for the first empirical evidence of indirect evaluative learning effect across stimuli belonging to distinct domains. If we restrict the focus of the analysis to the indirect effect, this study 3 and study 1 share the same target objects (groups). It seems that, at least at the explicit level, participants can transfer the valence of the self to groups more easily when another group, rather than a brand, works as driver for this transfer. To see whether there is any impact of the order in which stimuli relations are learned by individuals, the following study mirrored Study 3 except for the fact that groups were first presented with self-stimuli in the SR task, and then paired with brands in the TC task.

**Study 4**

We tested the indirect SR effect from a self-paired group to a brand. Participants completed a SR where the self was to be categorized with the same action as a group (and the category others with another group). In the Target Categorization task these groups (i.e., Lerriani [Target1] and Dattiani [Target2]) were categorized through the same action as two brands (i.e., Gester [Target3] and Vabel [Target4]).

**Method.**

### Participants and procedure.

One hundred and twenty University students (70 women, 49 men, one missing, $M$age = 23.60, $SD$ = 6.18) took part to this second study. The study followed the same procedure as Study 3, except for the fact the groups were used in the SR task and then paired with brands in the TC task.

**Results and discussion**

For the standard SR effect, we excluded data from participants who made over 25% of errors in at least one of the classification tasks (SR Task and IAT1), leading to a final sample of 119 participants. For the indirect effect of SR, one participants made over 25% of errors in either the Target Categorization task or IAT2. One hundred and three (86.6%) participants answered correctly IR memory question 1 and ninety-three (78.8%) did so for both IR memory 1 and 2 (see Supplementary Materials Chapter 3 – Appendix F for the analyses on the full samples).

Both the IAT and the explicit measure used to assess attitudes towards groups were reliable ($\alpha$ = .86 and $\alpha$ = .93, respectively). The effect of SR manipulation on IAT group1/group2 was basically a replication of what we observed in study 1. Means and standard deviations are reported in Table 6. The SR manipulation significantly affected the implicit, $F(1,104) = 17.11$, $p < .001$, $\eta_p^2 =$ .15, and the explicit attitude, $F(1,103) = 7.19$, $p < .009$, $\eta_p^2 = .07$.

Both the IAT and the explicit measure toward brands were reliable ($\alpha$ = .77 and $\alpha$ = .93, respectively). A one-way-ANOVA showed a significant effect on explicit attitude, $F(1,91) = 7.84$, $p$ = .006, $\eta_p^2 = .08$. However, the effect was not significant at the implicit level, $F(1,91) = .51$, $p = .477$,

$\eta_p^2 = .05$ (means and standard deviations are reported in Table 7). Therefore, participants liked explicitly, but not implicitly, the brand related with the group previously related to the self.

We showed that the evaluation of two brands related with two alternative social groups depended on whether the groups were previously related to the self, but this indirect SR effect occurred only at the explicit level. Interestingly, if we focus our attention on the outcome measures, this pattern of results was reversed compared to what we observed in Study 3. Moreover, a comparison with study 2, in which we tested indirect SR towards brands as we did here, also revealed a different pattern: the effect was significant on implicit attitude and only marginally significant on explicit attitude in study 2, while non-significant on implicit attitude and fully significant on explicit in study 4. Taken together, these findings seem to show that the order in which stimuli of different domains are presented to individuals (and, assuming a propositional perspective, the order in which relations between pairs of stimuli are acquired) matter in qualifying indirect SR effects. Furthermore, this study is the first to show a non-significant indirect effect on implicit attitude. Clearly, one cannot exclude the possibility that it is due to random sample fluctuations. However, this result might support the idea that one's ability to form and validate propositional relations between stimuli in an associative learning paradigm determines effects not only on explicit attitude, as it is assumed by the dual-model account (Gawronski & Bodenhausen, 2011), but also on implicit attitude, in line with a propositional view (De Houwer, 2009; 2014). We investigated this hypothesis in Study 5, in which we repeated the design of Study 4 in order to see whether the observed findings replicate on a different sample and also included a measure that assesses participants' ability to form relationships between stimuli that were both directly and indirectly related throughout the study.

**Study 5**

The main purpose of this study was to understand better the pattern of results observed in Study 4. In particular, we reasoned that the lack of effect on implicit attitude was worthy of further examination as it might entail important theoretical implications. Therefore, Study 5 was meant a) to replicate Study 4 and see whether the lack of effect on implicit attitude was confirmed and b) to

understand whether this possible lack of effect could have been explained by participants' ability to form relationships between stimuli. For the second aim, the study added a Matching To Sample test (MTS, see Dymond & Whelan, 2010), developed in the context of Relational Frame Theory (RFT, Hayes, Barnes-Holmes, & Roche, 2001) and used here to measure participants' ability of relationally framing stimuli either presented together in the same task, or indirectly related across tasks (e.g., self-stimuli and brands). Relational framing is defined as a generalized operant response class that is established through a history of reinforcement across exemplars, and once established, any stimulus event may participate in a relational frame, given the relevant contextual cues. The various patterns of derived relational responding possess three properties: mutual entailment, combinatorial entailment, and transformation of stimulus function (see Hayes & Barnes, 1997). Mutual entailment involves deriving a B-A relation from an explicitly reinforced AB relation. The simplest example of combinatorial entailment involves deriving A-C and C-A relations from explicitly reinforced A-B and B-C relations. Transformation of function involves a derived relation between A and B, and a transformation of function in A based on this relation. For the purpose of this study, the crucial property we are interested in is combinatorial entailment, which we articulated in two different levels (see the method section). If implicit and explicit attitude change is a function of one's ability to respond to pairs of stimuli in a relational way, then one's relational responding tested through the MTS task should moderate the indirect SR effect.

**Method.**

**Participants and procedure.**

One hundred and twenty-two (95 women, 25 men, two missing, *M*age = 23.37, *SD* = 3.47) University students took part to this second study. The study followed the same procedure as Study 4, except for the inclusion of a MTS test at the end of the experimental session.

*MTS test.* This task was administered to test whether the performance of each of the previous learning tasks resulted in the formation of stimuli equivalences at different levels. Prior to the task participants were informed that "*in the next part of the experiment you should look at the item at the*

*top of the screen, and then choose one of the items at the bottom of the screen by clicking on it with the mouse*" and that they should "*try to respond as accurately as possible*". During each test trial, one sample stimulus was presented at the top of the screen and two "comparison" stimuli were presented at the bottom of the screen. Selecting either of the comparison stimuli (by clicking on them with a mouse) removed all stimuli from the screen for 750ms followed by the subsequent trial. No feedback was provided for any response emitted during this task. We tested the formation of equivalence between a) Lerriani group members and group names Lerriani/Dattiani, b) Dattiani group members and group names Lerriani/Dattiani, c) Gester products and the brand names Gester/Vabel, d) Vabel products and the brand names Gester/Vabel (*identity equivalence*: mainly used to ease the task for the participants), e) Lerriani group members and self/other words, f) Dattiani group members and self/other words,  g) Lerriani group members and Gester/Vabel products, h) Dattiani group members and Gester/Vabel products (*directly trained equivalence*), i) Self words and Gester/Vabel products, l) Other words and Gester/Vabel products (*indirect equivalence*). The task involved a total of 64 trials, from which we calculated three distinct MTS scores based on each distinct relationship tested. 16 trials (MTS score 1) tested mutual entailment (4 for each group (brand) exemplars – group(brand) names relationship). 32 trials were to test level1 combinatorial entailment (MTS score 2): 16 trials tested level1 combinatorial entailment between self/other-stimuli and groups, while 16 trials tested level1 combinatorial entailment between groups and brands. Finally, 16 trials tested level2 combinatorial entailment between self/other-stimuli and brands (MTS score 3).

**Results and discussion**

For the standard SR effect, no one made over 25% of errors in at least one of the classification tasks (SR Task and IAT1). For the indirect effect of SR, three participants made over 25% of errors in either the Target Categorization task or IAT2. Ninety-six (78.2%) participants answered correctly IR memory question 1 and eighty-nine (73%) did so for both IR memory 1 and 2 (see Supplementary Materials Chapter 3 – Appendix F for the analyses on the full samples).

To test the direct effect of the SR task we ran a one-way ANOVA, with SR condition as factor. The IAT score was calculated such as higher score indicated a preference for Target1. The SR manipulation significantly affected implicit attitude towards the two groups, $F(1,94) = 16.91$, $p < .001$, $\eta_p^2 = .15$. Also on explicit attitude score we found a significant effect of SR manipulation, $F(1,94) = 14.44$, $p < .001$, $\eta_p^2 = .13$ (see Table 6 for descriptives).

For the indirect effect of the SR, a one-way-ANOVA showed no significant effect on implicit attitude, $F(1,87) = 1.18$, $p = .280$, $\eta_p^2 = .01$. The effect was not significant at the explicit level either, $F(1,87) = .09$, $p = .770$, $\eta_p^2 = .001$ (see Table 7 for descriptives).

We tested a series of moderation analysis to see whether participants' ability to identify equivalence between stimuli did affect the indirect effect of the self-referencing manipulation. As a first moderator, we considered the number of correct responses on the sixteen MTS trials that measured indirect equivalences (MTS score 3) - the ability to relate stimuli never encountered together in the same task throughout the experimental session (i.e., self/other words and brand stimuli). All variables were centered (standardized). A first moderation analysis was conducted on implicit attitude. The model accounted for 12.38% of variance. We did not observe a main effect of either the SR manipulation ($\beta = .14$, $p = .196$) or the MTS score ($\beta = .10$, $p = .390$). However, the interaction term was significant ($\beta = .33$, $p = .004$). We decomposed the effect with a simple slope analysis (see Figure 3, left panel) and observed that for participants whose score on the MTS was low or medium, the manipulation was not significant ($B = -.19$, $p = .255$ and $B = .14$, $p = .196$, respectively) while it was significant for those scoring high on the MTS ($B = .47$, $p = .002$). The same analysis was conducted on explicit attitude change. The model accounted for 12.34% of variance. We did not observe a main effect of either the SR manipulation ($\beta = .08$, $p = .462$) or the MTS score ($\beta = .16$, $p = .162$). However, the interaction term was significant ($\beta = .32$, $p = .006$). We decomposed the effect with a simple slope analysis (see Figure 3, right panel) and observed that for participants whose score on the MTS was low or medium the manipulation was not significant ($B = -.24$, $p = .152$ and $B = .048$, $p = .462$, respectively) but it was significant for those scoring high ($B = .40$, $p = .009$).

Figure 3. *Simple slope analysis of the interaction between indirect SR condition and indirect MTS score on implicit and explicit attitude.*

On the same dependent variable we ran a moderation analysis by considering MTS score 2, calculated on the thirty-two trials that measured directly trained equivalences - the ability to identify equivalence between object categorized together in the same task (i.e., self/other stimuli and groups as well as groups and brands). The model accounted for 14.29% of variance on implicit attitude change. We did not observe a main effect of either the SR manipulation ($\beta$ = .13, $p$ = .220) or this MTS score ($\beta$ = .08, $p$ = .452). However, the interaction term was significant ($\beta$ = .37, $p$ = .001). We decomposed the effect with a simple slope analysis (see Figure 4, left panel) and observed that for participants whose score on the MTS was low or medium the manipulation was not significant ($B$ = -.24, $p$ = .098 and $B$ = .13, $p$ = .220, respectively) while it was significant for those scoring high on the MTS ($B$ = .50, $p$ = .001). On explicit attitude, the model explained 10.74% of variance on explicit attitude. We did not observe a main effect of either the SR manipulation ($\beta$ = .06, $p$ = .576) or this MTS score ($\beta$ = .15, $p$ = .150). However, the interaction term was significant ($\beta$ = .32, $p$ = .004). We decomposed the effect with a simple slope analysis (see Figure 4, right panel) and observed that for participants with low and medium MTS score the manipulation was not significant ($B$ = -.26, $p$ = .087 and $B$ = .06, $p$ = .576, respectively) while it was significant for those scoring high ($B$ = .37, $p$ = .014).
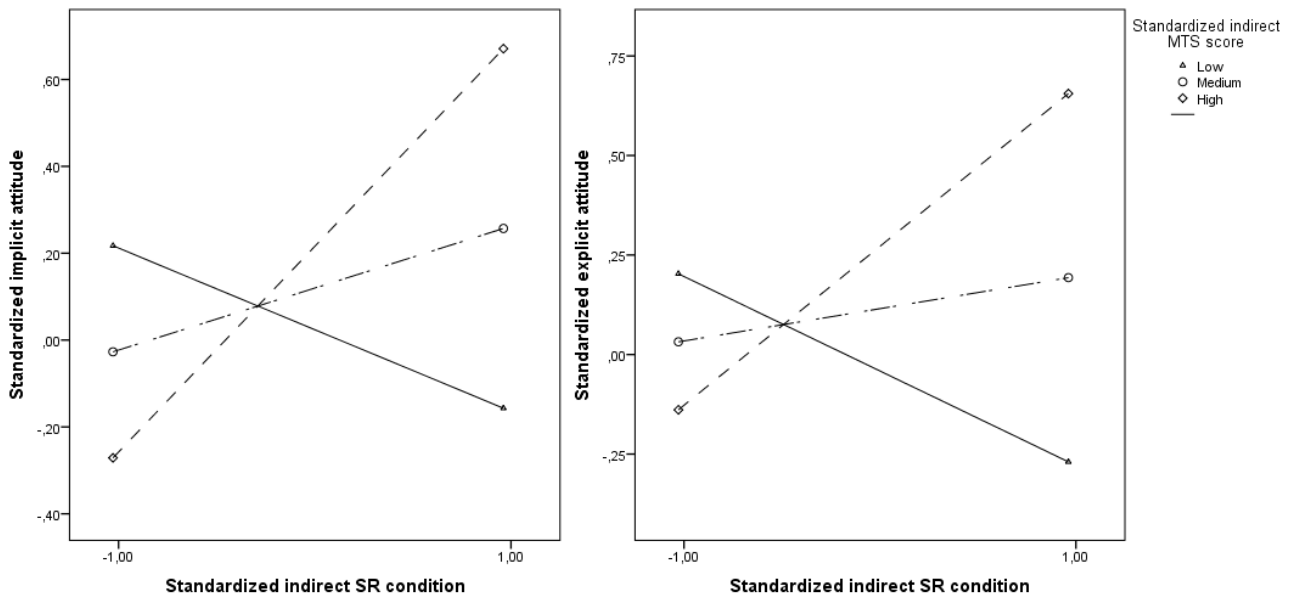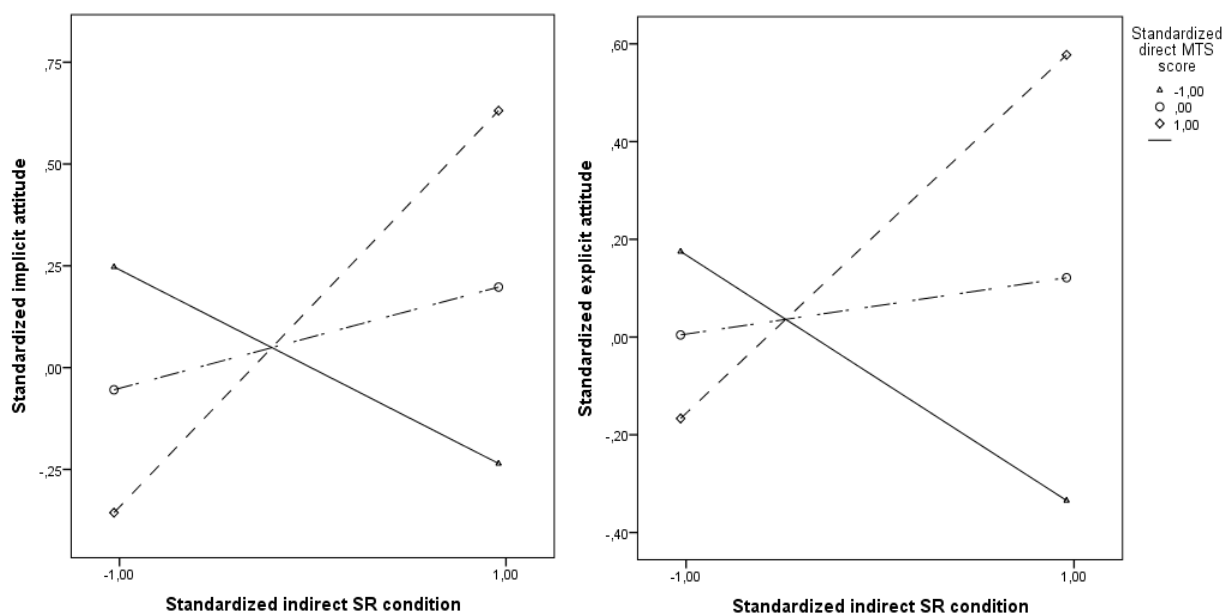
Figure 4. *Simple slope analysis of the interaction between indirect SR condition and direct MTS score on implicit and explicit attitude.*

Results from Study 5 confirmed the lack of an indirect SR effect on implicit attitude when groups convey the valence carried by the self to brands. Contrary to Study 4, we also observed a non-significant indirect effect on explicit attitude. However, the results were importantly qualified. In fact, the key findings of the present study refers to the role of the MTS scores (testing either directly trained or indirect equivalences) in qualifying the effect of indirect SR. We showed that, on both implicit and explicit attitude, the effect of the manipulation was qualified by participants' ability to establish relationships between stimuli. The moderation was significant when considering both types of MTS score (i.e., MTS 2 and MTS 3). Taken together, these moderation analyses suggested that when participants are good in inferring equivalence relations between stimuli, indirect SR occurs at both implicit and explicit level. This is true when looking at the relation between stimuli that were presented together in either the SR task or the TC task as well as when the focus was on the equivalence between stimuli never encountered together.

**The effect of IR memory on indirect SR**

EC studies use participants' awareness of the contingency between stimuli as an indicator of the underlying nature of the observed EC effect. Up to present, there is no agreement on whether

indirect evaluative learning effect are qualified by participants' awareness of the contingency between stimuli. Walther (2002) showed that the effect does not require participants' awareness of the actual pairings between stimuli in the acquisition phases. Gast & De Houwer (2012) measured the awareness of the untrained (indirectly established) contingency and found that at least for implicit evaluation, it amplified the effect. In the SR studies, IR memory can be conceptualized as the equivalent of contingency awareness in EC. A recent meta-analysis of the SR effect (Mattavelli et al., 2016) showed that participants' ability to learn and recall the IR that links the self to a target object is of crucial importance for the effect to occur. In this contribution, we inspected the role of IR memory for the indirect SR effect by comparing at a meta-analytical level the overall effect among participants with either correct or incorrect IR memory. This was done also because otherwise the latter group would have had too few participants, and hence results would have been fragile, if we would have analyzed the incorrect IR memory group separately for each study. Participants who provided the correct response at both IR memory questions constitute the correct IR memory group; participants with either incorrect or no memory in at least one of the two questions form the incorrect memory group. For participants with correct memory ($N = 471$), the overall effect size ($d$) on implicit attitude was 0.362, $SE = 0.093$, 95% CI [0.181, 0.544] and significantly different from zero, $Z = 3.92$, $p < .001$. Similarly, on explicit attitude the overall effect ($d$) was 0.322, $SE = 0.104$, 95% CI [0.117, 0.526] and significant, $Z = 3.09$, $p = .002$. This pattern was not confirmed on participants with incorrect IR memory ($N = 123$). The mean effect ($d$) on implicit attitude was 0.169, $SE = 0.179$, 95% CI [-0.182, 0.519], not significantly different from zero, $Z = .94$, $p = .346$. On explicit attitude the overall effect ($d$) was -0.291, $SE = 0.179$, 95% CI [-0.642, 0.060], also not significantly different from zero, $Z = -1.62$, $p = .105$. These findings are important in that they show that participants' IR memory qualifies the effectiveness of the SR manipulation. Therefore, the ability to remember that a common action is shared between stimuli is a necessary pre-requisite for the indirect SR effect.

**General discussion**

Across five studies we tested the effectiveness of indirect SR over and across different classes of target objects. Previous EC studies showed that attitude can spread towards novel stimuli without requiring any direct pairing with a valenced object (Walther, 2002; Gast & De Houwer, 2012). In all these contributions, a spreading attitude effect occurred thanks to direct pairings between the US and CS1 first and (either experienced or instructed) between the CS1 and a CS2. Therefore, a double direct pairing between stimuli resulted in an indirect relationship between US and CS2, which acts as a precondition for the evaluations to spread over the novel stimulus. The present contribution used a different approach to engineer indirect attitude change through the associative route. We start from the assumption that IR can be conceptualized as a general and abstract mechanism through which people learn that stimuli are somehow related to each other. In its simplest form, this is demonstrated by extant research on the SR effect (Mattavelli et al., 2016): a common behavioral response suffices to transfer the positivity of the self towards a related object. By relying on an experimental procedure of two distinct learning phases, the present studies apply the IR principle in three different ways, which could result in as many stimuli relationships. The common action in the SR task links the self to a first target (direct IR 1), this target is linked to a second one through another common action in the Target Categorization Task (direct IR 2) and, across these two tasks, there is a further intersection between the self and the second target (indirect IR) because they are both linked to the first target. The five studies demonstrate that the positivity of the self can spread towards fictitious brands and groups that have never been encountered in the SR task by exploiting the IR principle.

Moreover, as showed in our micro-meta-analysis, the effect of the indirect SR was qualified by participants' ability to remember the correct intersecting regularities between objects. We demonstrated that the effect was reliable on both implicit and explicit attitudes only for participants with correct IR memory. These findings are consistent with previous results on the standard SR effect and confirmed the importance of detecting a correct shared element between stimuli to transfer liking from one stimulus to another one. Nevertheless, the fact that participants can correctly remember the intersection between stimuli does not tell much about their ability to qualify this intersection in

relational terms. In line with a symbolic interpretation of evaluative learning (De Houwer & Hughes, 2016), we believe that IR works as a proximal cue that people can use to infer relationships between two or more stimuli. This symbolic approach is based on the idea that throughout their history of life, people learn that objects that share something are also related in some way. What are the conditions under which learned IR between stimuli result into actual relations was one of the core questions we tried to answer in this contribution.

Besides testing indirect attitude change by spreading the positivity of the self via IR, we focused on whether the (in)congruency in stimuli domain could have any impact on the transfer of valence from one object to the other one. In Study 1 and 2, where indirect SR was tested across stimuli of a common domain, we found a reliable effect on both implicit and explicit indirect attitude change. Results were more complex in the three studies where indirect SR was transferred across stimuli of distinct domains. Study 3 showed that indirect SR towards groups, when brands were to convey the valence of the self, was significant on implicit attitude and only marginally significant on explicit attitude. In light with dual-model accounts of associative learning (Gawronski & Bodenhausen, 2006; 2011), explicit attitude should be more likely affected by people's ability to relate stimuli in a propositional fashion. While propositions that qualify the relation between stimuli of the same domain can be plausible, this is less likely when stimuli belong to different domains. The lack of effect on explicit attitude might reflect the fact that meaningful relations between stimuli of different domain are more difficult to be inferred through IR. Critically however, the pattern of results observed in Study 4 and Study 5 indicated that when the stimuli related in the other way around (i.e., groups relate with self/others in the SR task and then with brands in the TC task), indirect SR was not significant at the implicit level, while results were mixed on explicit attitudes. This finding paves the way for a series of important considerations concerning the role of inferred relations on implicit attitude change. One could assume that the difficulty in generating propositional relations among stimuli of different domains could also affect implicit attitude change, in line with a propositional account of associative learning (De Houwer 2009; 2014). Taken together, Study 4 and 5 confirmed that also on implicit

attitude, the indirect effect of the SR manipulation is less likely to occur when the stimuli involved in the transfer of liking belong to distinct domains. Study 5 importantly suggests that indirect SR, on both implicit and explicit attitude, is qualified by one's ability to form equivalences between stimuli that are either categorized through the same key in one of the two learning phase (directly trained equivalences) or never presented together in the same task (indirect equivalences). In fact, the two MTS scores considered in the analyses both moderated the impact of the indirect manipulation on implicit and explicit attitudes. We want to highlight the fact that these significant moderations were tested on a sample of participants who correctly remembered the intersecting regularities established in both the SR task and the TC task, which represents the optimal conditions for the SR effect to occur. Therefore, it seems like learning that two stimuli have been categorized through the same action is not sufficient for indirect SR to occur in these specific cases. Conversely, people need to qualify this intersection in terms of stimuli relations. The congruency between stimuli domain seems to be an important qualifier for people's ability to qualify the relation between stimuli created from either a common action (within the SR and the TC tasks) or a common stimulus (across the SR and the TC tasks). We acknowledge that administering the MTS test after the indirect SR manipulation might create ambiguity in the interpretation of these results. In fact, one could hypothesize that stimuli are treated as equivalent in the MTS because participants use their valence as a cue to infer a relation between them.

To summarize, the present work offers a first empirical demonstration of an indirect SR effect. From the perspective of its underlying principle (i.e., IR), the importance of these findings refers to the possibility to exploit the principle to create stimuli relationship in an increasingly indirect fashion. These findings are also promising if we focus on the role of the self. As we know from previous studies, the self is not merely characterized by its positive valence (e.g., Rogers, Kuiper, & Kirker, 1977) and that some of the properties of the self can be acquired by new object related to the self via IR (Richetin et al., 2016). Future research can test whether these cognitive properties (e.g., memory, accessibility, identification) might also transfer to novel stimuli through indirect intersections of

regularities. Furthermore, the five studies offered important indications in terms of the importance of using intersecting regularities as relational cues. We demonstrated that intersections are more effective when they occur through stimuli of the same domain. Therefore, besides acquiring through the procedure of the task that two objects share a common feature, people need to qualify the newly created link in relational terms.

# Chapter 4.

# Increasing implicit and explicit attitudes towards an organic food brand by referencing to oneself.[12]

---

[12] This chapter is based on 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.

**Overview**

Across the previous chapters, we analyzed the SR task from the perspective of its two main features, that is, the use of the self as a special source of attitudinal change and intersecting regularities as the underlying evaluative learning principle. Chapter 4 opens to a series of lines of research in which the SR task has been used as a mean to test the power of the self to change attitude, identification, and behavioral intention towards economic goods. As said in the general introduction, studying attitudes and understanding the condition under which attitudes change is important as to predict and influence people's behavior. Therefore, focusing on a paradigm that is meant to affect attitude and identification is important *per-se*, but it is even more valuable when such changes on attitude and identification results in intentions to behave accordingly. In this chapter, we investigated in two studies ($N = 324$) the possibility of changing hypothetical shopping behavior, brand identification, and attitude change persistence towards two alternative labels of green food. Despite the increasing development of green labels in the food market, to our knowledge no study has investigated implicit attitudes toward organic food nor has tried to change them. Capitalizing on the important role of the self in consumer or pro-environmental behavior, we aimed to change implicit and explicit attitudes toward organic food brands using the self. Targeting two fictitious brands of organic food in two studies, we assessed the SR effect on implicit and explicit attitudes. We showed that the SR manipulation results in more positive implicit attitudes, explicit attitudes, and to more frequent hypothetical choice of products for the eco-brand paired with the self compared to the other eco-brand. SR manipulation also resulted in higher level of brand identification. Moreover, changes in implicit attitude mediated changes in explicit attitude, identification, and hypothetical choice. Finally, we provided evidence the persistence of the effect: Participants liked and identified more with the brand originally paired with the self even after removing the pairing with the self. The discussion is organized around the importance of taking into account implicit attitudes toward organic food brands and the usefulness of the self in changing these attitudes.

**Introduction**

When confronted to different food items at the supermarket, consumers' choices may be determined by different processes. Individuals can choose a product after considering specific goals, values, and beliefs. Alternatively, they might grab items from the shelves simply on an impulse because they like the package. In other words, processes leading to consumers' food choice can be controlled, rational, or automatic, impulsive (Shiffrin & Schneider, 1977; Lieberman, 2003; Epstein, 1994). In psychology, dual-process models have provided a theoretical framework for the investigation of impulsive and controlled behavior defining two systems with different processes and capacities (e.g., Fazio & Towles-Schwen, 1999; Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004). The Impulsive system (Strack & Deutsch, 2004) or System 1 (Evans, 2003, 2008; Kahneman, 2011) integrates information in an automatic and cognitively efficient way based on associative processes whereas the Reflective system or System 2 involves slow and controlled processes that require time and cognitive resources. In the last few years, consumer researchers, economic psychologists, and economists have started to adopt the perspective of dual-process models (Alós-Ferrer & Strack, 2014; Samson & Voyer, 2012). For instance, Fudenberg & Levine's Dual Selves Model (2006) identifies different processes as reflecting the image of either a long-run or a short-run player, where a rational/controlled long-run self controls the impulses of a short-run self tempted by immediate rewards. In terms of attitude, these dual-process models have resulted in the differentiation between explicit and implicit attitudes, in reference to the deliberative and impulsive system, respectively. An implicit attitude reflects an automatic affective reaction resulting from associations activated when a person encounters an object (Gawronski & Bodenhausen, 2006). Moreover, implicit attitudes are supposed to "mediate favorable or unfavorable feeling, thought, or action toward social objects" (Greenwald & Banaji, 1995).

**Implicit attitudes: The specific case of food**

The general idea that human behavior can be explained by different processes has influenced attitude research and assessment (see Gawronski & Bodenhausen, 2011). In other words, the increased focus on implicit consumer cognition (Brunel, Tietje, & Greenwald, 2004) has led research to use indirect measures designed to tap into automatic processes (Friese, Hofmann, & Wänke, 2009). Moreover, the validity of direct measures is limited by the fact that people are not always able to retrieve their attitudes in memory prior to measurement (Dholakia & Morwitz, 2002) and not necessarily willing to reveal them. Among the indirect measures of attitude, the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) is the most used and reliable paradigm. In its most common form, the IAT consists of a computerized classification task of two target categories (e.g., fruit versus snack) and two attribute categories (e.g., positive versus negative words) with a 7- or 5-block structure. The underlying logic of the IAT refers to response interference or compatibility. If one has an implicit preference for snacks over fruit, it should be easier to classify positive words and snacks with a single key and negative words and fruit with another key than to classify negative words and snacks with the same key and positive words and fruit. The easiness of the task is evaluated through reaction times and error rates and the IAT effect is defined as the difference between these two response conditions.

In the food domain, researchers started adopting indirect measures such as the IAT to tap into the automatic processes involved in consumers' behavior towards food-related targets (Friese, Hofmann, & Wänke, 2009). For example, Maison, Greenwald, & Bruin (2004) demonstrated that implicit preferences towards competitor brands measured through an IAT predicted consumers' behavior over and above explicit preferences. Moreover, using a low/high calorie IAT, Maison, Greenwald, & Bruin (2001, Study 2) reported that women who prefer high-calorie over low-calorie products in terms of taste have nevertheless implicit preferences for low-calorie products. This dissociation between implicit and explicit food preferences underscores the importance of assessing both. Furthermore, empirical evidence demonstrates the validity of the IAT to predict individuals'

food choices between healthy versus unhealthy goods (Conner, Perugini, O'Gorman, Ayres, & Prestwich, 2007; Friese, Hofmann, & Wänke, 2008; Perugini, 2005, study2; Richetin, Perugini, Prestwich, & O'Gorman, 2007; for a review see Greenwald, Poehlman, Uhlmann, & Banaji, 2009), although some studies failed to do so (Karpinski & Hilton, 2001; Roefs & Jansen, 2002). In sum, because theoretical and empirical work support the importance of implicit attitudes in food related behaviors and cognitions, it seems key to consider them when examining food related cognitions. Moreover, if implicit attitudes are significant predictors of food choices, it becomes important to find procedures through which they can be changed.

**Changing implicit attitudes using the self**

As we mentioned earlier, an implicit attitude results from associations activated upon encounter of an object (Gawronski & Bodenhausen, 2006). Therefore, an associative-based procedure which involves the learning of a new evaluation toward the attitude object (e.g., pairing food with positive stimuli) appears to be an ideal candidate to induce implicit attitude change (Bodenhausen & Gawronski, 2013). Moreover, according to Gawronski and Bodenhausen's (2006, 2011) Associative-Propositional Evaluation (APE) model after a procedure expected to change evaluative associations (i.e., implicit attitudes), these latter serve as an input for deliberative evaluations (i.e., explicit attitudes) through a validation process (see also MODE model for similar flow, Fazio, 1990; Olson & Fazio, 2009). Therefore, such associative-based procedure could lead to implicit attitude as well as to explicit attitude change. In the general food domain, pairing information such as sensory information or valenced images with food items through co-occurrence in space and time (i.e., evaluative conditioning procedure, see Hofmann et al., 2010 for a review) has been shown to induce implicit attitude change toward such food items (Hollands, Prestwich, & Marteau, 2011; Lebens et al., 2011; Verhulst, Hermans, Baeyens, Spruyt, & Eelen, 2006).

Another important determinant of preferences and a main focus of research is the self. Starting from the fact that the majority of people has a high self-esteem and a general positive view of

themselves (e.g., Yamaguchi et al., 2007), the self has been demonstrated to affect liking towards self-related objects (Hoorens & Nuttin, 1993; Nuttin, 1985). In economic psychology literature, this effect has been defined as the endowment effect (e.g., Kahneman, Knetsch, & Thaler, 1990) or more recently in terms of ownership. For example, ownership is supposed to be the underlying mechanism of people's willingness to accept exceeding willingness to pay for the same good (Morewedge, Shu, Gilbert, & Wilson, 2009). Crucially, Gawronski, Bodenhausen, & Becker (2007) showed that this self-anchoring effect occurs also at the associative level with more positive implicit attitudes towards objects chosen or experimentally assigned, without the necessity to involve higher-order propositional and deliberative reasoning. From balanced triads of associations that include the self, an object, and valence, the association between the self and valence defines the implicit self-esteem, the one between the object and valence the implicit attitude, and the one between the object and the self the implicit self-identity (Greenwald, Banaji et al.,2002).

The role played by the self appears quite relevant for organic food products or brands, whose presence in the food market during the last few decades has had one of the biggest growths in the food industry. Many research demonstrate indeed the great importance of self-identity in determining beliefs and behaviors toward organic food (e.g., Sparks & Shepherd, 1992; Aertsens, Verbeke, Mondelaers, & Huylenbroeck, 2009; Grunert & Juhl, 1995). Even though people have also been shown to be willing to pay a price premium on eco-labelled food products (Roheim, Asche, & Santos, 2011; Zhang, Epperson, Huang, & Houston, 2009), many consumers are still reluctant to give up their usual products in favor of alternative ones produced by organic brands (e.g., Young, Hwang, Mcdonald, & Oates, 2010). Presumably, one main reason is their unwillingness to pay a price premium for organic products but, given the ample empirical evidence that food choices can be affected by uncontrolled associative processes, it would seem that an important role could be played also by implicit attitudes. Yet, very few studies have considered implicit attitudes as a predictor of

eco-brand choice (e.g., Vantomme, Geuens, de Houwer, & de Pelsmacker, 2005), none for organic or eco-brand food choice and, as far as we are aware, no study has tried to change them.

**Research aims**

Literature demonstrates the important role played by implicit attitudes in predicting food choices on the one hand and by the self on brands and in organic food related cognitions on the other hand. We thus argue that focusing on implicit attitudes and considering the properties of the self could constitute an important avenue for changing attitudes toward organic food. In other words, we wish to capitalize on previous findings to change attitudes in the domain of organic food products using the self. One procedure that might be effective in determining changes at the level of the associative structure is the self-referencing (SR) task (Perkins & Forehand, 2012; Prestwich, Perugini, Hurling, & Richetin, 2010). The SR task is an associative attitude change paradigm that uses the self as a source to induce positive attitude towards a certain target. In particular, the SR task requires participants to perform a common action (i.e., pressing the same key) for categorizing stimuli related to the self and to a target, and an alternative common action for the categorization of stimuli belonging to the category 'Others' and to another target. This simple commonality of actions provides the context within which the positivity of the self can be transferred to the target that has been paired with it, hence resulting in more positive implicit and explicit attitudes towards that target object (e.g., Perugini, Zogmaister, Richetin, Prestwich, & Hurling, 2013). In this context, our aims are fourfold. First, we aim to increase implicit and explicit preferences towards an organic food brand ("eco-brand" for simplicity's sake) by pairing it with the self through the SR task (Study 1 & 2). We use the Implicit Association Test (Greenwald et al., 1998) for assessing implicit preferences and semantic differential for assessing explicit ones. We also aim at testing whether the SR manipulation affects hypothetical shopping behavior (Study 1). Moreover, because brand identification or perceived closeness is important in brand related cognitions and purchase behavior (e.g., Kuenzel & Halliday, 2008), we investigate the SR effect on identification with an eco-brand (Study 2). Furthermore, because

theoretical models (Gawronski & Bodenhausen, 2006, 2011; Fazio, 1990; Olson & Fazio, 2009) postulate that implicit attitude change could serve as input for changes at the explicit level, we test the mediation of the SR effects on explicit attitude, identification, and hypothetical shopping behavior by the implicit attitude. Finally, we aim to examine the consequences in terms of evaluation and identification toward the brand if the self, as a source of change, is no longer connected with that product. As a final note, in both studies we measured participant's recollection of the same response action (i.e., IR) to classify stimuli related to the self and to a specific eco-brand (target) and ran the main analyses considering only participants who correctly remembered it. There are two main reasons for this strategy. First, theoretically the SR task relies on the principle that the self and a target share the same response to allow the transfer of properties from the self to the target: If people do not notice this regularity, it is unlikely that it can affect their attitudes. Second, empirically previous studies using the SR paradigm have demonstrated that when such memory is incorrect there are no noticeable effects (Perugini, Richetin, & Zogmaister, 2014). Having said that, the analyses performed on the full samples show a very similar pattern of results, albeit obviously weaker (see Supplementary Material Chapter 4).

**Pilot study**

Given the format of the SR task and of the IAT, two categories of organic food products were required. We chose to illustrate these two categories using two fictitious eco-brands to avoid potential effects of previous knowledge or preference and to avoid any influence of familiarity, an element that has been hypothesized as potentially relevant for organic food products (Wheeler, Sharp, & Nenycz-Thiel, 2013). We thus conducted a preliminary study to identify two alternative eco-brands (i.e., logo and name) and two sets of products similar in valence and ecological meaning.

**Method**

**Participants and design**

We preselected fourteen logos and names as potential eco-brands and nine pairs of pictures of similar food products (e.g., two milk bottle pictures). Twenty-seven students (18 women, 9 men, $M$age = 23.21, $SD$ = 4.21) took part in a rating task. First, participants rated the 14 pictures of logos and 18 pictures of food products in a random order on 9-point scales from 1 (*I do not like at all*) to 9 (*I like very much*). Participants had to answer within 4 seconds after which they were prompted to reply more quickly. Second, participants indicated the extent to which each of the 14 logos represented an ecological/organic dimension on 7-point scales from 1 (*not at all*) to 7 (*very much*).

**Results**

We chose two logos that were neutral in valence ($M$ = 5.33, $SD$ = 2.04 and $M$ = 5.37, $SD$ = 2.02) $t(26)$ = .85, $p$ = .403 and $t(26)$ = .95, $p$ = .350, respectively and not different from each other, $t(26)$ = .09, $p$ = .932. They were also well representing the ecological dimension ($M$ = 5.59, $SD$ = 1.18 and $M$ = 5.44, $SD$ = 1.16), $t(26)$ = 6.98, $p$ < .001 and $t(26)$ = 6.50, $p$ < .001) to a similar extent, $t(26)$ = .66, $p$ = .515. We named these two logos "Ecove" and "Ambio" as in Italian both names evoked an ecological dimension. Concerning the pictures of products, we elaborated two similar groups of products that did not differ in valence to be used later for the SR task ($M$ = 5.66, $SD$ = 1.22 and $M$ = 5.68, $SD$ = 1.25, respectively), $t(26)$ = -.11, $p$ = .910, and for the IAT ($M$ = 5.67, $SD$ = 1.19 and $M$ = 5.62, $SD$ = 1.61, respectively), $t(26)$ = .31, $p$ = .756.

**Study 1**

Study 1 investigates the self-referencing effect on cognitions related to organic food products. We test whether pairing self-related stimuli and a neutral eco-brand through a classification task leads to increased implicit and explicit liking towards this brand and its products as well as increased choice in a hypothetical shopping behavior. For this purpose, we use the SR paradigm, the Implicit Association Test and semantic differentials to assess its effect on implicit and explicit attitude respectively. Finally as target objects, we use the two fictitious eco-brands (i.e., Ambio & Ecove) selected in the preliminary study.

**Method**

**Participants and procedure**

One hundred and ninety-six students (120 women, 76 men, *M* age = 22.60, *SD* = 2.83) took part to a one-session study. The study was presented as a research assessing the opinion of people on two potential lines of organic food products with lower environmental impact compared to classic food brands (eco-brand) in order to choose the best one to be launched on the market. After reading a brief description of the two lines, participants completed a short learning task and the self-referencing task. For half of them, the manipulation consisted of pairing the eco-brand "Ambio" with the self, and for the other half it consisted of pairing the eco-brand "Ecove" with the self. Then, half of participants completed first an IAT followed by a hypothetical behavioral choice in a shopping task and the other half completed first the behavioral measure followed by the IAT. Finally, all participants indicated their explicit evaluation of the eco-brands and products and completed a self eco-brand memory test. After that, the experimenter thanked, debriefed, and gave course credit to the participants. Note that the use of two brands each paired with the self in one of the two SR conditions provides a distinctive methodological and theoretical advantage and allows for a strong experimental control. First, participants in both conditions performed the exact same task and therefore the two conditions are equalized over all relevant details (e.g., the cognitive activities involved are the same). Second, given that both eco-brands are paired with the self, the results can be generalized to both; therefore one can exclude potential theoretically irrelevant effects due to some idiosyncratic features of a brand.

**Materials**

*Familiarization task*. On each trial, a food product picture (e.g., box of cookies, milk) appeared in the center of the screen. Participants indicated, as quickly as possible, the eco-brand to which the product belonged by pressing one of two keys on the keyboard (i.e., 'E' and 'I'). The eco-brands labels and corresponding keys remained on the left or right upper portion of the screen

throughout the task. The familiarization consisted of 20 trials divided into two blocks of 10. In each block, the trials featured products of each eco-brand (five pictures for each), intermixed at random. The response keys for the two eco-brands were switched after one block. This prevents the keys from being paired with specific responses. The order in which participants completed these blocks was counterbalanced. In case of incorrect classification, a red-X appeared on screen and remained until correction. The inter-trial interval was 400ms. Each set of five products was "Ambio" for half participants and "Ecove" for the other half to prevent any effect of the set of products. We proceeded in the same way for the other tasks (i.e., SR, IAT, and shopping task, see Supplementary Materials Chapter 4 – Appendix I).

*Self-referencing task*. Participants categorized in two blocks of 40 trials, as quickly as possible, "Ambio" ["Ecove"] pictures and words related to self (self, me, my, mine, I) to one response key (e.g., 'E') and "Ecove" ["Ambio"] pictures and words relating to others (they, them, their, his, her) to a different response key (e.g., 'I'). Participants then repeated the two blocks of 40 trials but with switched keys, i.e., "Ambio" ["Ecove"] and self assigned to the 'I' key, and "Ambio" ["Ecove"] pictures and other-related words to the 'E' key. The order in which participants completed these two sets of blocks was counterbalanced. In case of incorrect classification, a red-X appeared on screen and remained until correction. The inter-trial interval was 400ms.

*IAT*. Participants classified words and pictures individually presented in a random order in the middle of the screen, using two keys (i.e., 'E' and 'I'). The target concept was "Ambio" and its contrast was "Ecove", 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 "Ambio" and "Positive" being presented first and the other half having the combination "Ecove" and "Positive" being presented first. 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 200ms in case of incorrect response but without requiring correction (no

built-in penalty). The inter-trial interval was of 500ms, and the category labels stayed on the upper part of the screen throughout the task. For each attribute category, we used five stimuli. The D score with 600ms of penalties for errors was calculated (Greenwald, Nosek, & Banaji, 2003) such that positive values indicated an automatic positive preference for "Ambio" over "Ecove". The reliability of the IAT score was good ($\alpha$ = .90).

*Hypothetical shopping task.* Participants had to imagine they needed to shop for a series of 6 food items (e.g., rice, coffee). For each item, they had to choose between two products (of equivalent price) the one they would purchase. The two products of the same food type were accompanied with a short description (for example for the rice of one eco-brand: "*a rice that never sticks, ideal for all recipes*" and for the rice of the other eco-brand: "*a rice always al dente to please every cook*", see Supplementary Materials Chapter 4 – Appendix I). None of the products was presented in the previous tasks. The order of presentation (left vs. right) of the two products was fixed random in a way such that 4 "Ambio" products were presented on the left side of the screen and 4 "Ecove" products were presented on the left side of the screen. We computed a shopping task score by summing the number of times Ambio was chose over Ecove.

*Explicit Evaluation of the brands and products.* There were two types of explicit evaluation for both eco-brands and their set of products, one single and one relative. First, participants rated each brand and each set of products separately ("Ambio" first when first presented with positive in the IAT vs. "Ecove" first when first presented with positive in the IAT) on four bipolar dimensions (ugly-nice, unpleasant-pleasant, worthless-valuable, useless-useful) on 7-point scales from 1 to 7. Then, participants rated one brand relative to the other and one set of products relative to the other on 4 dimensions (interesting, pleasant, attractive, beneficial) on 7-point scales from 1 (e.g., "Ecove" more interesting) to 7 (e.g., "Ambio" more interesting, see Supplementary Materials Chapter 4 – Appendix I). For brands and sets of products taken together, an overall relative explicit attitude score was calculated. First, for brands and products separately we computed four difference scores for each

bipolar dimension then we included these scores with the relative evaluations (sixteen scores in total) in a Principal Components Analysis that revealed one factor accounting for 54.91 % (factor loadings ranging from .50 to .84) ($\alpha$ = .94). Higher factor scores indicated a more positive evaluation for "Ambio" over "Ecove".

*IR memory test.* Participants indicated which eco-brand of the two was paired with the words linked to the self by responding to the following question: "One of the tasks that you have done consisted in classifying with the same key words related to the self and pictures related to one eco-brand. Do you remember which eco-brand?" Participants indicated one of the two eco-brands or the option "I don't know". Participants were categorized as having a correct (correct response) or incorrect (no recollection or incorrect response) memory.

**Results and discussion**

The data of six participants were excluded because of a large proportion of errors (over 25%) in either the SR task or the IAT, indicating random responding. Of the remaining 190 participants, 128 (67.4%) correctly recollected the eco-brand for which the response action for classifying its stimuli was the same than for self-related stimuli (self eco-brand test). We only present the analyses relative to these participants (see Supplementary Materials Chapter 4 – Appendix H for the analyses on the full sample).

The main analyses involved a 2 (SR Condition: "Ambio" + "Self" vs. "Ecove" + "Self") x 2 (Order of measures: IAT first vs. Shopping task first) analysis of variance (ANOVA). There was a significant main effect of the SR manipulation on the implicit attitude (IAT), $F(1, 124) = 21.15$, $p <$ .001, $\eta_p^2 = .14$, and on explicit attitude, $F(1, 124) = 7.16$, $p = .008$, $\eta_p^2 = .06$ (this effect was also present when considering the explicit attitude toward the brands and the products separately, $p = .002$ and $p = .045$, respectively). Moreover, this SR manipulation also resulted in a significant effect on the shopping task score, $F(1, 124) = 4.09$, $p = .045$, $\eta^2 p = .03$.

Table 8. *Means and standard deviations for Study 1 and Study 2 for each SR condition*

| | Study 1 | | Study 2 | |
|---|---|---|---|---|
| | Ambio+Sel | Ecove+Self | Ambio+Self | Ecove+Self |
| | *n* = 68 | *n* = 60 | *n* = 51 | *n* = 49 |
| IAT score | .26 (.49) | -.16 (.55) | .32 (.53) | -.19 (.51) |
| Explicit Attitude score | .22 (.88) | -.25 (1.07) | .30 (.88) | -.31 (1.03) |
| Identification score | NA | NA | .45 (1.99) | -.47 (1.87) |
| Shopping Task score | 3.12 (1.20) | 2.67 (1.31) | NA | NA |

*Note.* Higher scores indicate a preference towards Ambio.

The eco-brand and its products paired with the self were evaluated more positively at both implicit and explicit levels and chosen more frequently in the shopping task compared to the eco-brand paired with others (see Table 8 for means and standard deviations for each SR condition). There was no main effect of Order of measures on the implicit attitude nor on the shopping task score, $F(1, 124) = .37$, $p = .547$ and $F(1, 124) = .66$, $p = .419$, respectively. We found no interaction effect between the manipulation and the order of measures on the implicit attitude, $F(1, 124) = .17$, $p = .679$ nor on the shopping task score, $F(1, 124) = .46$, $p = .500$.

Finally, the three dependent variables were correlated (see Table 9). These significant correlations, as well as the fact that the SR manipulation affected significantly all three criteria, allowed us testing two potential mediation effects. We tested the two hypotheses with mediation analyses using the Process macro (Hayes, 2012). First, because of the associative nature of the paradigm we used for changing attitude and following theoretical assumptions (e.g., Gawronski & Bodenhausen, 2006, 2011; Olson & Fazio, 2009), we investigated whether the SR effect on explicit attitude was mediated by the implicit attitude change (see Figure5, left panel).

Table 9. *Correlations for Study 1 (N = 128)*

|  | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. SR manipulation | 1 |  |  |  |
| 2. Implicit Attitude score | .38** | 1 |  |  |
| 3. Explicit Attitude score | .23** | .46** | 1 |  |
| 4. Shopping Task score | .18* | .43** | .41** | 1 |

*Note. SR manipulation is coded 1: Ambio+Self, 0: Ecove+Self. ** p < .01. * p < .05.*
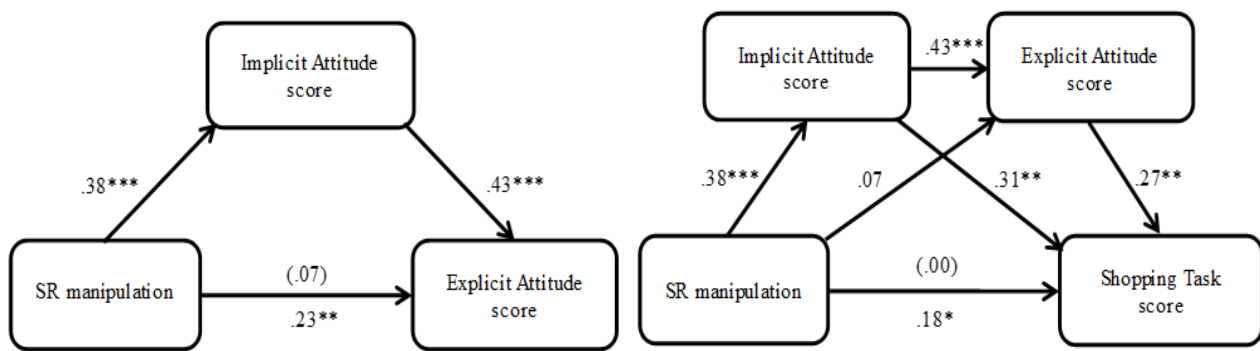


Figure 5. *Mediation of SR manipulation effects (Study 1). The upper panel illustrates the mediation of the SR effect on the explicit attitude by the implicit attitude. The lower panel illustrates the serial mediation of the SR effect on the shopping task score through the implicit attitude and the explicit attitude.*

The analysis revealed a full mediation with a non significant effect of the SR manipulation on the explicit attitude score when controlling for the effect of the implicit attitude score (direct effect), $\beta$ = .07, $p$ = .438, 95% CI [-.10, .24] with a significant indirect or mediated effect, $M$ = .16, 95% CI [.08, .27]. The effect of the SR manipulation on the explicit attitude score was thus fully mediated by the implicit attitude score. Second, we tested a serial mediation hypothesis in which both implicit and explicit attitude mediated serially the SR effect on the choice in the shopping task (see Figure 5, right panel). Because the effect of the SR manipulation on the explicit attitude was mediated by the implicit attitude, an indirect or mediated effect of the SR manipulation on the shopping task through implicit and explicit attitude in serial is more likely than two mediated effects operating in parallel. Note that

the shopping task was taken prior to the assessment of implicit attitudes for half of the sample. We thus acknowledge this double mediation model, while theoretically sound, is hypothetical. However, the lack of interaction between the SR manipulation and the order of tasks (IAT first vs. shopping task first) on the IAT score and on the shopping score provides a rationale for considering the full sample, hence achieving greater statistical power, regardless of the order of tasks. The analysis revealed a non significant direct effect of the SR manipulation on the shopping task score, $\beta = -.004$, $p = .968$, 95% CI [-.17, .16], indicating a full mediation. There was a significant indirect effect of the SR manipulation on the shopping task score through the IAT score, $M = .12$, 95% CI [.05, .21], indicating the mediating role of the IAT. The indirect effect of the SR manipulation on the shopping task score through the explicit attitude score was not significant, $M = .02$, 95% CI [-.04, .08], indicating the lack of mediation by the explicit attitude. Moreover, there was a significant indirect effect of the SR manipulation on the shopping task score through the implicit attitude score and the explicit attitude score in sequence, $M = .04$, 95% CI [.01, .11], indicating a serial mediation. In other words, the effect of SR manipulation on the shopping task score was mediated by the IAT and by the explicit attitude in a serial manner.

This study showed that the self-referencing manipulation influences both implicit and explicit attitudes toward an eco-brand and its products. These findings confirm the usefulness of the self for changing attitudes toward food products (Perugini et al., 2013; Prestwich et al., 2010). Moreover, this study provides preliminary direct evidence that the SR effects might extend on buying organic food products, although hypothetically. In other words, performing a simple common action (i.e., pressing the same key) for categorizing stimuli related to the self and to a brand of organic food products can be sufficient to lead not only to an implicit and explicit preference toward this brand but also to choose its products more often. The results from the mediation analyses showed that changing implicit attitude toward a brand of organic food mediated a change in explicit attitude. Moreover, the serial mediation of the SR effect on the choice in the shopping task through the implicit and the explicit

attitude, together with the lack of mediation by the explicit attitude, support the idea that, once implicit attitudes toward organic food products are changed, explicit attitudes and behavior can follow suit.

**Study 2**

To explore further the SR effects obtained in Study 1, the aims of Study 2 are twofold. First, we focus on brand identification, which plays a role in determining brand pleasantness (Burnkrant & Unnava, 1995), consumer loyalty (Bhattacharya & Sen, 2003; He, Li, & Harris, 2012), consumer satisfaction, and likelihood of repurchase (Kuenzel & Halliday, 2008), and might also facilitate storage and later retrieval of brand-relevant information (Rogers, Kuiper, & Kirker, 1977; Symons & Johnson, 1997). It seems therefore important to investigate means of increasing the identification toward a brand. We thus test whether participants' identification (or closeness) would be stronger for an eco-brand paired with the self than with an eco-brand paired with others. Second, this study addresses the issue of the persistence of the attitude change driven by the self. If the self is a source of change for the attitude towards a product, what happens if this source is no longer connected with that product? For example, what would have happened to Nespresso® if George Clooney had disappeared from the ads after a few days? To test this hypothesis experimentally, we investigate whether after the SR manipulation, a subsequent simple classification of the eco-brand products without being paired anymore with the self would lead to a decline in the liking of and the identification with the brand initially paired with the self. Finally, we introduce a minor methodological modification with the aim of increasing the proportion of participants who correctly recall the commonality of response between the self and a specific eco-brand. If participants made a percentage of errors superior to 20% in the SR task, they were required to complete 2 additional blocks of classification before passing to the next task. Moreover, we administered the IR memory test immediately after the SR task rather than at the end of the experimental procedure.

**Method**

**Participants and procedure**

One hundred and twenty-eight participants took part to this study (40 men and 88 women, *Mage* = 22.17, *SD* = 2.46). We used the same cover story as for Study 1. Different from Study 1 though, participants completed the tasks by providing their responses through a response box instead of using the keyboard. Participants first read a description of the two eco-brands and then completed the Self Referencing task. The first manipulation consisted of pairing stimuli belonging to the eco-brand "Ambio" ["Ecove"] with the self and the eco-brand "Ecove" ["Ambio"] with stimuli belonging to the category "other". Participants then indicated their recognition of the source-target pairing. The second manipulation consisted of the removal of the pairing between the self and the eco-brands. Half of participants completed an "Ambio"/"Ecove" categorization task (simple brand categorization task group), whereas the other half did not perform such task (control group). After that, all participants performed an IAT and filled in the explicit evaluations of the brands and products as well as a measure of inclusion of the eco-brands in the self. Finally, participants were thanked, debriefed, and given course credit.

**Materials**

*Familiarization task*. Participants completed the same learning task used in Study 1.

*Self-referencing task*. The task was identical to the one used in Study 1 with one exception. In the last block of 40 trials, the number of correct responses was calculated. If this proportion of errors was above 20%, participants were informed of their poor performance and completed two additional blocks of 20 trials (one block for each key assignment).

*IR memory test*. The measure was the same than the one used in Study 1.

*Simple brand categorization task.* Participants completed a task that consisted of classifying products into the two eco-brand categories. The task had the same structure and the main features than the learning task (i.e., same response keys, same number of blocks, same number of trials), but with different products pictures for the two brands used in the learning and SR tasks.

*IAT.* The procedure was identical to the one used in Study 1. The IAT score was transformed into the $D_6$ as it was done in Study 1 ($\alpha = .89$).

*Explicit evaluation.* The procedure was identical to the one used in Study 1. As done in Study 1, we used the factor scores from a Principal Component Analysis that extracted one factor accounting for 57.95 % of variance (factor loadings ranging from .60 to .88) ($\alpha = .95$).

*Brand Identification.* We adapted the Inclusion of other in the self scale (Aron, Aron, & Smollan, 1992) to assess brand identification with the eco-brand. Participants evaluated their level of identification with each eco-brand indicating a pair of circles representing the self and the eco-brand on a scale ranging from 1 (very distant circles) to 7 (overlapping circles). An identification score was calculated for each participant by subtracting the score for the brand "Ecove" to the one for "Ambio", with positive score indicating an advantage for "Ambio".

**Results and discussion**

The data from one participant was excluded from further analyses because of high number of errors in the SR task ($> 25\%$). We also eliminated the data from one participant because of extreme values on both attitudes ($z > |3|$) and identification measures ($z > |2.89|$). Finally, we ran the analyses by considering only participants who correctly remembered the eco-brand for which the response action was the same than for the self ($N = 100$, 78.1%, see Supplementary Materials Chapter 4 – Appendix H for analyses on the full sample).

To test the SR effects and their strength on implicit and explicit attitudes and identification, we ran a 2 (SR condition: Ambio+Self vs. Ecove+Self) x 2 (Pairing removal: brand categorization task vs. control) analysis of variance (ANOVA) on each dependent variable. For the implicit attitude score, a main effect of SR manipulation was observed, $F(1, 96) = 24.09$, $p < .001$, $\eta_p^2 = .20$. Like in Study 1, the eco-brand paired with the self in the SR task resulted in a more positive implicit attitude compared to the eco-brand paired with others (see Table 8 for descriptive statistics in each SR condition). This effect was not qualified by an interaction with the Pairing removal condition, $F(1,$

96) = .10, *p* = .757 and this latter factor was not significant, $F(1, 96) = 1.10$, *p* = .296. For the explicit attitude, the pattern was the same. There was a main effect of the SR manipulation, $F(1, 96) = 10.05$, *p* = .002, $\eta_p^2 = .10$ indicating an explicit preference for the eco-brand paired with the self (this effect was also present when considering the explicit attitude toward the brands and the products separately, *p* = .002 and *p* = .005, respectively). Again, there was no significant interaction between SR manipulation and the Pairing removal manipulation, $F(1, 96) = 1.16$, *p* = .284, and no effect of the Pairing removal, $F(1, 96) = .84$, *p* = .361. For participants' relative identification, the pattern was also identical. There was a significant effect of the SR manipulation, $F(1, 96) = 5.60$, *p* = .020, $\eta_p^2 = .05$ with no interaction with the Pairing removal, $F(1, 96) = .37$, *p* = .546, nor a main effect of the latter, $F(1, 96) = .62$, *p* = .434. Participants identified themselves more with the eco-brand that was paired with the self than with the other eco-brand.

Like in Study 1, the correlations between the three dependent variables were significant (see Table 10) and the SR manipulation affected significantly all three. We thus tested two mediation hypotheses. First, we investigated whether the SR effect on explicit attitude was mediated by the implicit attitude (see Figure 6, left panel). The analysis revealed a full mediation with a non significant effect of the SR manipulation on the explicit attitude score when controlling for the effect of the implicit attitude score (direct effect), β = .08, *p* = .424, 95% CI [-.11, .26] and a significant indirect effect, *M* = .23, 95% CI [.13, .39]. The results corroborated the ones obtained in Study 1. Second, we tested a serial mediation hypothesis in which both implicit and explicit attitudes mediated the SR effect on the identification score (see Figure 6, right panel). The analysis revealed a non significant direct effect of the SR manipulation of the identification score, β = -.05, *p* = .424, 95% CI [-.19, .08] indicating a full mediation. There was a significant indirect effect of the SR manipulation on the identification score through the IAT score, *M* = .06, 95% CI [.003, .13], indicating a mediation effect by the implicit attitude.

Table 10. *Correlations for Study 2 (N = 100)*

|  | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. SR manipulation | 1 | | | |
| 2. Implicit Attitude score | .45** | 1 | | |
| 3. Explicit Attitude Score | .31** | .56** | 1 | |
| 4. Identification score | .23* | .52** | .80** | 1 |

*Note. SR manipulation is coded 1: Ambio+Self, 0: Ecove+Self. ** p < .01. * p < .05.*
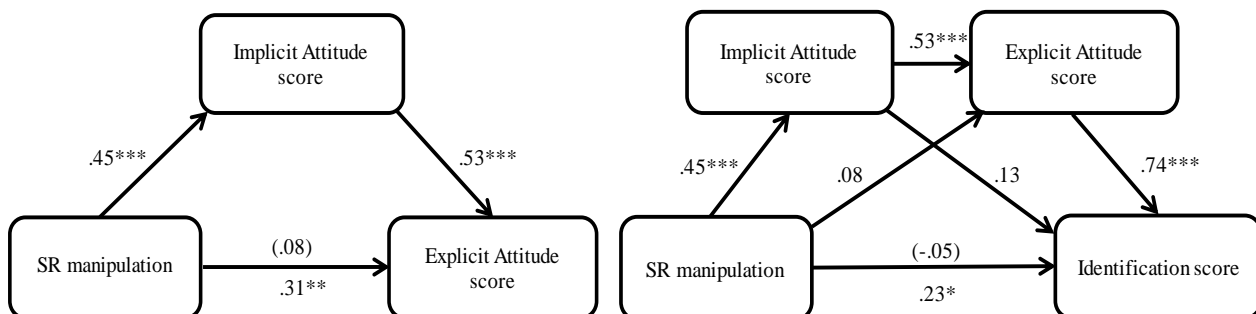


Figure 6. *Mediation of SR manipulation effects (Study 2). The upper panel illustrates the mediation of the SR effect on the explicit attitude by the implicit attitude. The lower panel illustrates the serial mediation of the SR effect on the identification score through the implicit attitude and the explicit attitude.*

A non-significant indirect effect of the SR manipulation on the identification score through the explicit attitude score emerged, $M = .06$, 95% CI [-.08, .19], indicating a lack of mediation by the explicit attitude. Finally, there was a significant indirect effect of the SR manipulation on the identification score through the implicit attitude score and the explicit attitude score, $M = .17$, 95% CI [.11, .29]. Like in Study1 for the shopping task choice, the SR effects on identification were mediated serially, first by the implicit attitude and then by the explicit attitude.

Study 2 offers further evidence to the effectiveness of the SR paradigm in increasing eco-brands positive evaluations at both the implicit and explicit levels and extends it to the identification with the brand. Moreover, this study shows that the SR has a strong effect on both attitude change and

identification with the brand once paired with the self. The lack of interaction observed between SR condition and Pairing removal suggests that the positivity of and the identification to the brand paired with the self is unaffected by a subsequent removal of that pairing. Finally, as shown in Study 1, the SR effect on explicit attitude was mediated by the implicit attitude and the SR effect on the identification was mediated by the implicit and the explicit attitude serially and not in parallel. This shows once again the possibility of a domino effect (i.e., a change in implicit attitude produces a change in explicit attitudes, which in turn produces a change in brand identification).

**General discussion**

The advent of organic food in the market has received great attention in the last decades. However, few if any study on organic food have taken into consideration the important role of implicit preferences that already have been demonstrated to predict food-related behavioral choices (e.g., Conner et al., 2007; Richetin et al., 2007). In the present contribution, capitalizing on the positive valence of the self and its role in brands and pro-environmental related cognitions, we tested the possibility to induce change in implicit and explicit preferences, as well as affect choice in a shopping task and identification toward organic food using the self.

The results from two studies indicate that the SR is effective in generating both implicit and explicit attitude change towards organic food brands: Both the brand paired with the self and the products belonging to that specific brand were more liked than those paired with the category 'others'. These results are in line with previous studies on other food items such as drinks (Prestwich et al., 2010) and crisps (Perugini et al., 2013). To determine the consistency of the different effects of the SR manipulation on the implicit and explicit attitudes across the two studies, we meta-analyzed the results using the software Comprehensive Meta Analysis (Borenstein, Hedges, Higgins, & Rothstein, 2005). For the implicit attitude, Cochran's $Q$ statistic yielded a non significant effect, $Q(1) = 0.37$, $p = .543$, indicating homogeneity and therefore suggesting to apply a fixed-effect model. The overall effect was significant ($z = 6.35$, $p < .001$) with an average effect size $d = 0.88$ (95% CI: 0.61, 1.16).

For the explicit attitude, Cochran's $Q$ statistic yielded also a non significant effect, $Q(1) = 0.32$, $p = .570$, indicating homogeneity. The overall effect was significant ($z = 4.07$, $p < .001$) with an average effect size $d = 0.55$, 95% CI: [0.29, 0.82]. These meta-analysis results revealed substantial effect sizes for the implicit attitude and for the explicit attitude, providing therefore robust empirical evidence of the SR effect on implicit and explicit preferences toward brands of organic food.

Moreover, we demonstrated that the SR effect can be extended to hypothetical behavioral choices and that a brand can become incorporated in the extended concept of the self (Belk, 1988) after a SR task. This identification can lead to the enactment of a series of mechanisms typical of the self, such as for example, defending the brand when it is threatened (Lisjak, Lee, & Gardner, 2012). Moreover, given that brands seen as closer to the individuals are more likely purchased than other brands (Malhotra, 1988), the effect on identification might be an additional indicator for potential influence on behavior.

The consistent mediation of the SR effect on explicit attitude by the implicit one together with the single mediations by the implicit attitude of the same effects on both choice and identification strongly underscores the necessity of taking into account implicit preferences when studying issues related to organic food. In short, the results suggest that, once an implicit attitude towards an organic food brand is changed, this effect can ramify further it in a domino like manner. These results also support the sequential flow from associative evaluations to deliberative evaluations suggested by theoretical models such as the MODE (Fazio, 1990; Olson & Fazio, 2009) and the APE (Gawronski & Bodenhausen, 2006, 2011).

Finally, this work provides initial evidence that the positive nudge given by pairing the self with an organic food brand persists after the pairing is removed. This removal has been done on a short time scale, hence illustrating the resistance of the attitude change but arguably limiting its implications to a relatively short span. Future studies could investigate whether a more substantial duration of the removal of the pairing (e.g., multiple simple brand categorization tasks performed

during some days) would lead to a decline in the positivity of the attitude or it will instead persist. One could even imagine studies where this issue is explored systematically to find approximate temporal thresholds of decline.

The significant results on all criteria in the two studies indicated the breadth of the effects of the self-referencing paradigm. It might be worth noting two features. First, the food items used in the manipulation task were different to the ones used for the assessment of implicit and explicit attitudes and for the shopping task. Second, the SR effects on the explicit preferences were observed considering both the brand and the products. These procedural and measurement details suggest that the self-referencing effects could in principle generalize to a full range of products belonging to one brand. In other words, once a brand is paired with the self, its positivity might spread to all newly encountered products of the same brand. Future studies specifically focused on this issue might be interesting.

Two methodological details might be worth mentioning. In the procedure we have used in both studies, participants were exposed an equal number of times to both eco-brands. Hence, familiarity, which is known to be a main factor affecting consumer's choices of eco-products (Wheeler et al., 2013), cannot explain the differential preference for the eco-brand paired with self. We have considered here two organic food brands, each in turn paired with the self. The advantage is that we can exclude that the effects are due to some unwanted subtle difference between the brands and therefore allow us generalizing. The results also provides some insights in what one should do to successfully launch a new brand of organic food products, that is, to foster some associations with the self such that it might be imbued by its positivity. Future research should extend to the effects of pairing a brand of organic food with the self when contrasted with a brand of non-organic food, given that one of the main challenges for the organic food market is to compete against non-organic food products.

Taken together, these findings suggest that the self serves to increase liking, choice, and identification towards organic food. A limitation of this contribution is that we have used a hypothetical, rather than real, choice with no economic implications. Future research should focus on real choices with monetary consequences. Considering the role of ownership on both attitudes (Gawronski et al., 2007) and economic decisions (Morewedge et al., 2009), it is possible, if not quite likely, that the self-referencing task can have effects also on real-economic behavior.

To conclude, our work suggests that in order to foster positivity towards organic food, one should consider also automatic processes and thus focus on changing implicit attitudes. We have provided all-round evidence that the self can be used successfully to change implicit and explicit cognitions toward organic food brands by creating some commonality through, for instance, simple actions. This minimal manipulation can have profound effects.

# Chapter 5.

**How can implicit and explicit attitudes both be changed? Testing two interventions to promote consumption of green vegetables.[13]**

---

[13] This chapter is based on Mattavelli, S., Avishai, A., Perugini, M., Richetin, J., & Sheeran, P. (*accepted for publication, Annals of Behavioral Medicine*). How can implicit and explicit attitudes both be changed? Testing two interventions to promote consumption of green vegetables.

**Overview**

In chapter 4, we demonstrated that the SR is a reliable paradigm to change implicit attitude, brand identification and behavioral intention towards organic food labels. Critically however, rather than change, the two studies provided evidence of the ability of the SR task to form attitudes towards newly encountered food brands. As a matter of fact, an attitudinal change requires a shift from an existing to a new degree of favor or disfavor towards an object (Petty & Brinol, 2010). Chapter 5 addresses this issue in the context of both the SR task and persuasive communication. We tested the impact of (a) a SR intervention and (b) a persuasive communication in modifying implicit and explicit attitudes towards green vegetables and promoting readiness to change. To test a real attitudinal change, the study targeted individuals who explicitly reported they did not like or only moderately liked green vegetables. We assigned participants ($N = 273$) to a 2 (self-referencing: present vs. absent) x 2 (persuasive message: present vs. absent) factorial design. The outcomes were implicit and explicit attitudes as well as readiness to increase consumption of green vegetables. We found implicit attitudes to increase after repeatedly pairing green vegetables stimuli with the self in the self-referencing task but not in response to the persuasive communication. The persuasive message increased explicit attitudes and readiness to change, but did not alter implicit attitudes. A three-way interaction with pre-existing explicit attitudes was also observed. In the absence of a persuasive message, the self-referencing task increased readiness to change among participants with more negative pre-existing explicit attitudes. This study is the first to demonstrate that a self-referencing task is effective in changing both implicit attitudes and readiness to change eating behavior. Findings indicate that distinct intervention strategies are needed to change implicit and explicit attitudes towards green vegetables.

**Introduction**

Most theories of health behavior assume that changing a person's conscious cognitions (e.g., attitudes) will change behavior (Ajzen, 1991; Bandura, 1998). However, dual process theories point out that behavior is a function not only of the reflective system (conscious cognitions) but also the associative (automatic) system (Strack & Deutsch, 2004). Accordingly, attitudes towards a behavior can be measured both explicitly (via self-reports) and implicitly (by measuring underlying associations between mental representations of the behavior and its evaluation) and the valence of these two types of attitude may diverge (Greenwald, McGhee, & Schwartz, 1998; Nosek, 2007). Research indicates that explicit (McEachan, Conner, Taylor, & Lawton, 2011) and implicit attitudes (Prestwich, Hurling, & Baker, 2011) both predict dietary behavior. However, only a handful of studies have tested how to modify implicit attitudes towards health behaviors (e.g., Hollands, Prestwich, & Marteau, 2011). It is also unclear whether different intervention strategies are needed to change implicit and explicit attitudes (i.e., a reflective intervention to target explicit attitude, and an associative intervention to target implicit attitude).

The present research tests the impact of two interventions, a persuasive communication that targets explicit attitudes (reflective intervention), and a novel *self-referencing* technique that targets implicit attitudes (associative intervention). To test the effect of the two interventions on attitude *change*, we assessed participants' pre-existing explicit attitude towards eating green vegetables. Based on this assessment, we recruited only participants whose liking of green vegetables ranged from very unfavorable to moderately favorable (i.e., scores from 1 to 5 on a 7-point scale), and tested whether the effectiveness of the interventions depends upon the valence of participants' pre-existing explicit attitudes. We deliberately excluded participants with 'very positive' or 'extremely positive' pre-existing attitude towards the target behavior from the study as (a) there is little scope for attitudinal change in this group, and (b) people with less favorable attitudes towards eating green vegetables are of greater public health concern.

Implicit measures of attitude typically consist of reaction-time tasks that capture memory-based associations. For instance, the Implicit Association Test (IAT, Greenwald et al., 1998) measures attitudes toward target objects by inviting participants to classify stimuli belonging to one category (e.g., fruit) versus those belonging to an alternative category (e.g., candy) together with positive versus negative words (e.g., sunshine vs. rain). The faster participants classify high-fat foods together with positive words (as compared with low-fat foods with negative words), the more favorable is their implicit attitude toward high-fat foods compared to low-fat foods. Based on a similar logic is the single target IAT (Karpinski & Steinman, 2006), which differs from the standard IAT in that it does not require a contrast target category, and thus represents a non-relative measure of implicit attitude. Although some studies failed to observe a significant relation between implicit attitudes and behavior (Karpinski & Hilton, 2001, Study 2), there is now reliable evidence to show that implicit attitudes predict choices among a large variety of food items, such that the chosen items are those with greater implicit liking (Friese, Hofmann, & Wänke, 2008; Perugini, 2005; Richetin, Perugini, Prestwich, & O'Gorman, 2007). Implicit attitudes have also been shown to correlate positively with self-reported consumption of snacks (Conner, Perugini, O'Gorman, Ayres, & Prestwich, 2007) as well as the intake of low-calorie food (Maison, Greenwald, & Bruin, 2001).

However, most research investigating the relationship between implicit attitudes and dietary behavior is correlational. Intervention strategies that aim to change implicit attitudes have received less consideration. As implicit attitudes are automatic affective reactions resulting from the particular associations activated automatically upon encounter with a relevant stimulus (Gawronski & Bodenhausen, 2006), associative learning procedures are prime candidates as change techniques. Evaluative Conditioning is the best known of these procedures and involves repeated pairing a valenced stimulus with a neutral stimulus, which causes the latter to acquire the valence carried by the former (Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). So far as we are aware, only a handful of studies have used evaluative conditioning to induce implicit attitude change in the context of health behavior, and all of these studies involved repeatedly pairing unhealthy target

stimuli (e.g., snacks) with negative stimuli (Hollands et al., 2011; Houben, Havermans, & Wiers, 2010; Lebens, Roefs, Martijn, Houben, Nederkoorn, & Jansen, 2011).

However, changing attitude in a positive direction is important in order to promote health-protective behaviors such as consuming green vegetables. Here, evaluative conditioning procedures need to use positive stimuli to promote favorable attitudes towards the target. The self constitutes an important positive stimulus as most people have positive feelings about the self (i.e., moderate to high self-esteem) (Yamaguchi et al., 2007). The Self-Referencing Task (Prestwich, Perugini, Hurling, & Richetin, 2010) is an associative learning procedure that exploits this finding and uses the self to generate positive attitude change. The self-referencing task differs from the common procedures used in evaluative conditioning in that it relies on the principle of Intersecting Regularities (Hughes, De Houwer, & Perugini, 2016). Intersecting regularities refers to a class of associative learning procedures where the transfer of valence from one stimulus to another occurs is generated by establishing common features between valenced and target stimuli. In the case of the self-referencing task, this common feature is that the same action is used to categorize self-related stimuli and the target stimuli.

The self-referencing technique has proven effective in changing attitudes and behavioral intentions towards food brands (Perugini, Zogmaister, Richetin, Prestwich, & Hurling, 2013). However, all previous studies focused on either fictitious or unknown food brands, thus demonstrating that self-referencing is a powerful strategy to form, rather than change, attitudes towards food. No studies have tested the possibility that self-referencing could modify pre-existing attitudes towards food, which represents the greatest challenge in health behavior interventions. Moreover, as previous evaluative conditioning studies involved pairing an undesirable behavior and negatively valenced stimuli, there is as yet no evidence that associative learning interventions can change implicit food associations by pairing *positive* stimuli with the target behavior.

Both the Associative and Propositional Evaluation (APE) model (Gawronski & Bodenhausen, 2006) and the systems of reasoning analysis of attitude change propose that implicit and explicit

attitudes are formed or changed via different processes (Rydell & McConnel, 2006). For instance, Rydell and McConnell showed that explicit attitudes were sensitive to propositional information (i.e., a communication about negative characteristics of the attitude object) but implicit attitudes were not (Study 1). Conversely, implicit attitudes were sensitive to a subliminal priming procedure that altered the underlying associations, whereas explicit attitudes were not (Study 5). This analysis implies that different interventions are required to change implicit and explicit attitudes – the intervention should modify associations to change implicit attitudes, and offer persuasive arguments to change explicit attitudes. However, as these studies involved newly-formed attitudes concerning a fictitious individual (Bob), it is not yet clear whether the systems of reasoning analysis hold for behaviors that are familiar to, and consequential for, participants. The present research takes these considerations seriously, and also measures people's pre-existing attitudes towards the behavior at issue (Wilson, Lindsey, & Schooler, 2000), as empirical evidence (Krosnick & Petty, 1995) indicates that people's pre-existing attitudes can undermine attempts to change (explicit) attitudes.

**The present study**

Based on the foregoing discussion, the present study adopted a 2 (Self-referencing: present vs. absent) x 2 (Persuasive message: present vs. absent) factorial design. Participants' pre-existing attitudes were measured at explicit level. The study focused on implicit and explicit attitudes towards, and readiness to increase consumption of, green vegetables. In the USA, people eat 1.5 cups of vegetables daily, or approximately 50-60 percent of the 2-3 cups recommended for adults. More than half of vegetable intake comes from potatoes and tomatoes, whereas only 10% comes from dark green (and orange) vegetables (Lin, Wendt, & Guthrie, 2013). Green vegetables consumption is important for physical well-being (Di Noia, 2014), and is associated with reduced rates of cancer and other chronic diseases (Higdon, Delage, Williams, & Dashwood, 2007). Increasing consumption of green vegetables thus has considerable public health importance, especially among people who are more reluctant to eat them. For this reason, only participants who do not explicitly like or only moderately liked eating green vegetables were targeted in the interventions. Our research questions derived from

the APE model, which assumes that associative interventions affect primarily implicit attitudes, whereas persuasive communications affect mainly explicit attitudes and readiness to change. Three key questions were formulated. (1) Is a self-referencing intervention effective in changing implicit attitudes towards green vegetables?, (2) Does the self-referencing intervention only affect implicit attitudes, and does the persuasive message intervention only affect explicit attitudes and intentions?, and (3) Do pre-existing explicit attitudes towards green vegetables moderate intervention effects such that the self-referencing or persuasive message interventions are less effective among participants with more negative initial attitudes?

**Method**

### Participants and procedure

Participants ($N = 273$; 159 females, $M$-age = 36.12, $SD = 11.82$) took part in a single-session experiment via Amazon Mechanical Turk. The procedure involved an initial screening questionnaire first, followed by the main study. The screening phase asked questions about participants' liking of three different targets (smoking, African Americans, and eating green vegetables) to conceal the purpose of the research. Participants who scored 1-5 on a 7-point scale (not at all-very much) in response to the question, "How much do you like to eat green vegetables?" were allowed to take part in the main study. The screening session served to recruit approximately 60 participants to each of the five groups based on pre-existing explicit attitude scores. An algorithm ensured that approximately an equal number of participants scoring 1, 2, 3, 4, or 5 on the measure of pre-existing explicit attitudes to green vegetables entered the study. The algorithm was set such that once the quota was met in each score group, additional participants who fell in that group were no longer allowed to access the main study. The final sample consists of two hundred and seventy-three participants because for some of the baseline scores it was difficult to meet the expected subsample of 60 participants (score 1 = 46 participants; score 2 = 59 participants; score 3 = 60 participants; score 4 = 49 participants; score 5 = 59 participants). Thereafter, participants were randomly assigned to intervention conditions in the 2 (Self-referencing: present vs. absent) x 2 (Persuasive message: present

vs. absent) design. All participants then completed a single-target IAT to measure implicit attitudes, as well as measures of explicit attitude towards, and readiness to increase consumption of, green vegetables.

**Interventions**

*Self-referencing intervention.* The self-referencing intervention consisted of four blocks of 35 trials. In the first two blocks participants categorized, as quickly as possible, pictures of green vegetables and words related to "Self" (e.g., *I, me*) using one response key (*E*) on the keyboard and words relating to "Other" (e.g., *they, them*) to a different response key (*I*). Participants then repeated the two blocks of 35 trials by switching the keys assigned to the categories (i.e. 'green vegetables' and "Self" assigned to the 'I' key, and "Others" to the 'E' key). In this way, participants learned to associate green vegetables with the self. The order in which participants completed these two pairs of blocks, and the proportion of left- vs. right-hand responses was counterbalanced. In the case of incorrect classification, a red X appeared on screen and remained until participants corrected their response. Stimuli comprised 5 "Self" words (*I, me, my, mine, myself*), 5 "Other" words (*they, them, their, other, theirs*), and 5 pictures of green vegetables selected from *food.pics* database (see www.food-pics.sgb.ac.at, picture numbers: 232, 250, 274, 432, 455, see Supplementary Materials Chapter 5 – Appendix L for stimuli used in the study). Participants' contingency memory (their recollection of source-target pairings) was tested by answering the question *"The tasks you have just done consisted in classifying with the same key green vegetables and a category of words. What category?"*. Participants could answer "Self," "Other," or *"I don't know."*

*Control task.* The control condition for the self-referencing task had an identical structure but participants did not learn to associate green vegetables with the self via sharing the same response key. Instead, participants categorized "Self" with one key *(E),* "Other" with an alternative key (*I*), and pictures of green vegetables with another key (*B*). The stimuli were identical to the self-referencing intervention.

*Persuasive message.* The persuasive message (see text below) was based on the communication developed by Harris et al. (2014) that proved effective in changing attitudes, intentions, and behavior. The message emphasized advantages pertaining to both health conditions and physical appearance (in bold). It also stressed the fact that green vegetables can be tasty and easy to prepare by offering two alternative ways to enjoy them in a meal.

"*Dark green leafy vegetables are nutritional powerhouses filled with vitamins, minerals and nutrients. They are rich in chlorophyll and fiber, which keep the blood and colon healthy. Because of their high content of antioxidants, **green leafy vegetables may be one of the best cancer-preventing foods**. Studies have shown that eating 2 to 3 servings of green leafy vegetables per week may lower the risk of stomach, breast and skin cancer. **Leafy greens can also make you more attractive**. Pigments that give vegetables their color accumulate in the skin and give it color, too. They enhance our natural coloring and when we see someone whose skin reflects these enhanced pigments, they appear healthier and more attractive to us. The Dietary Guidelines for Americans recommends increasing average intakes of fruits and vegetables, particularly those that provide more vitamins, minerals and fiber. Dark leafy greens fulfill this need. Many varieties of greens are available in the American markets and there are many easy and tasty ways to enjoy a meal with leafy greens:*

***Wrap it up:** Make a wrap with tuna, chicken or turkey and add romaine lettuce, spinach, arugula, and other veggies for some extra flavor.*

***Add to soup:** Add greens with larger, tougher leaves such as collard greens, kale or mustard greens into your favorite soup.*"

Participants in the control condition did not receive a message.

**Measures**

*Implicit attitude*. We used a single target IAT to measure how strongly participants associated green vegetables with positive and negative words. Participants were asked to classify 4 green vegetables pictures and positive (*good, pleasant, nice, love*) and negative words (*bad, unpleasant, nasty, hate*) with two response keys (left and right) in a congruent and an incongruent block. The

pictures used in the single target IAT were different to those used in the self-referencing intervention (Picture numbers in the food.pics database: 273, 395, 430, 508, Supplementary Materials Chapter 5 – Appendix L). A practice block of 24 trials preceded the two critical blocks. The two critical blocks consisted of 84 trials each and order was counterbalanced. Within the blocks, all stimuli were presented in random order. In cases of incorrect classification, a red X appeared on screen and remained until correction. Shorter response latencies on the congruent block than on the incongruent block indicated more positive implicit attitudes towards green vegetables. Reliability was high ($\alpha$ = .83).

*Explicit attitude*. Four items on 9-point scales were used to measure explicit attitudes to green vegetables: "My feelings about eating green vegetables are…' (negative - positive, unfavorable - favorable), and "The consequences of my eating green vegetables are…" (negative - positive, unfavorable - favorable). Principal Components Analyses revealed a single factor that explained 70.94% of the variance. Reliability was high ($\alpha$ = .86).

*Readiness to change*. Readiness to change was measured by two items using 9-point scales: "How willing are you to try to increase the amount of green vegetables you eat in future?' (not at all - very much) and "Are you prepared to increase the amount of green vegetables you eat in future?" (not at all prepared - definitely prepared). Reliability was high ($r$ = .92).

**Analysis Strategy**

We first checked data quality. Next, we examined the correlations between the manipulations and measures, and among the measures for each condition. Finally, hierarchical multiple regression analyses were used to test the impact of the reflective and impulsive interventions and pre-existing explicit attitude, and their interactions on implicit attitudes, explicit attitudes, and readiness to change. Interventions were dummy coded (0, 1) and pre-existing explicit attitude scores were standardized prior to computing interaction terms.

**Results**

**Checks for Data Quality**

Four participants exhibited an exceptionally high error rate (> 25%) during the self-referencing or IAT tasks, and were excluded from further analyses. Data from the remaining 269 participants were analyzed. All but three participants (2%) showed correct contingency memory.

**Correlations**

Table 11 shows the correlations between manipulations and measures. As expected, the self-referencing intervention was significantly related to implicit attitude but not explicit attitude, whereas the message intervention was significantly associated with explicit but not implicit attitude. Pre-existing and post-intervention explicit attitudes correlated significantly ($p < .001$), but neither variable was significantly correlated with implicit attitude. Both pre-existing and post-intervention explicit attitudes were associated with readiness to change but implicit attitude was not. This pattern of correlations was also observed when the sample was split by experimental condition (see Table 12). These findings suggest that the interventions did not affect the strength of relationships among the measures.

**Regression Analyses**

A two-step hierarchical multiple regression was conducted for all three dependent variables. Pre-existing explicit attitude was entered at step 1, and the dummy coded intervention conditions (i.e., self-referencing and message conditions) and the interaction terms entered at step 2 (see Table 13). For implicit attitude, the regression analysis revealed that pre-existing explicit attitude did not contribute significantly to the regression model at step 1 ($p = .452$). At step 2, the overall equation predicting implicit attitude from self-referencing condition, message condition, pre-existing explicit attitude, and their interactions did not account for a significant proportion of variance either ($p = .270$), but the beta for self-referencing condition was significant ($p = .036$). We therefore reran the regression with self-referencing condition as the sole predictor. Consistent with the correlation observed in Table 11, this effect of the self-referencing intervention on implicit attitude remained significant, $F(1,267) = 4.47$, $p = .035$ .

Table 11. *Correlations between Manipulations and Measures*

|  | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. SR condition | 1.00 | | | | | |
| 2. Message condition | .00 | 1.00 | | | | |
| 3. Pre-existing explicit attitude | -.02 | -.01 | 1.00 | | | |
| 4. Implicit attitude | .13* | .02 | .05 | 1.00 | | |
| 5. Explicit attitude | -.01 | .12* | .63** | .07 | 1.00 | |
| 6. Readiness to change | .00 | .09 | .60** | .04 | .67** | 1.00 |

*Note.* * $p < .05$, ** $p < .01$.

Table 12. *Correlations between measures based on participants' assigned condition.*

|  | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| **SR = absent; Message = absent** | | | | |
| 1. Pre-existing explicit attitude | 1.00 | | | |
| 2. Implicit attitude | .13 | 1.00 | | |
| 3. Explicit attitude | .68** | .11 | 1.00 | |
| 4. Readiness to change | .75** | .09 | .72** | 1.00 |
| **SR = absent; Message = present** | | | | |
| 1. Pre-existing explicit attitude | 1.00 | | | |
| 2. Implicit attitude | -.11 | 1.00 | | |
| 3. Explicit attitude | .70** | .07 | 1.00 | |
| 4. Readiness to change | .62** | -.04 | .68** | 1.00 |
| **SR = present; Message = absent** | | | | |
| 1. Pre-existing explicit attitude | 1.00 | | | |
| 2. Implicit attitude | .08 | 1.00 | | |
| 3. Explicit attitude | .61** | .04 | 1.00 | |
| 4. Readiness to change | .43** | -.10 | .65** | 1.00 |
| **SR = present; Message = present** | | | | |
| 1. Pre-existing explicit attitude | 1.00 | | | |
| 2. Implicit attitude | .11 | 1.00 | | |
| 3. Explicit attitude | .60** | .05 | 1.00 | |
| 4. Readiness to change | .63** | .18 | .63** | 1.00 |

*Note.* ** $p < .01$

Table 13. *Hierarchical Regression of Implicit Attitude, Explicit Attitude, and Readiness to Change on Pre-existing Attitude and Experimental Conditions.*

| Variable | Implicit atttitude | | | Explicit attitude | | | Readiness to Change | | |
|---|---|---|---|---|---|---|---|---|---|
| | B | SE | β | B | SE | β | B | SE | β |
| Step 1 | | | | | | | | | |
| Pre-existing attitude | .02 | .02 | .05 | 1.16 | .09 | .63*** | 1.36 | .11 | .60*** |
| | | | | | .40*** | | | .36*** | |
| | | | | | 176.44*** | | | 150.90*** | |
| Step 2 | | | | | | | | | |
| Pre-existing attitude | .02 | .02 | .05 | 1.17 | .09 | .63*** | 1.38 | .11 | .61*** |
| SR condition | .04 | .02 | .13* | .01 | .09 | .01 | .04 | .11 | .02 |
| Message condition | .01 | .02 | .03 | .23 | .09 | .13** | .22 | .11 | .10* |
| SR*Message | -.02 | .02 | -.07 | -.04 | .09 | -.02 | -.12 | .11 | -.05 |
| SR*Baseline | .02 | .02 | .04 | -.04 | .09 | -.02 | -.18 | .11 | -.08 |
| Message*Baseline | -.02 | .02 | -.05 | -.02 | .09 | -.01 | .05 | .11 | .02 |
| SR*Message*Baseline | .02 | .02 | .07 | .03 | .09 | .02 | .25 | .11 | .11* |
| | | .03 | | | .42*** | | | .39*** | |
| | | 1.26 | | | 26.47*** | | | 24.11*** | |

*Note.* SR = Self-referencing, B = unstandardized regression coefficient, *SE* = standard error, β = standardized regression coefficient.

* *p* < .05, ** *p* < .01, *** *p* < .001.

Participants who completed the self-referencing task showed greater implicit liking of green vegetables ($M$ = .18, $SD$ = .35) than participants in the control condition ($M$ = .10, $SD$ = .33). Thus, only the associative (self-referencing) intervention increased the favorability of implicit attitudes towards green vegetables. For explicit attitude, pre-existing explicit attitude was a significant predictor at step 1 ($p < .001$). At step 2, the equation including all the variables and their interactions accounted for a significant amount of variance on explicit attitude change ($p < .001$). Message condition and participants' pre-existing attitudes were the only significant predictors ($p$s $< .05$). Participants who received the persuasive message had more positive explicit attitudes towards green vegetables ($M$ = 6.90, $SD$ = 1.84) than participants who received no message ($M$ = 6.44, $SD$ = 1.82). As for implicit attitude, there were no significant interactions among the intervention conditions and pre-existing attitudes in predicting explicit attitudes ($p = .704$).

Pre-existing explicit attitude also predicted readiness to change at step 1 ($p < .001$). The inclusion of all the other explanatory variables in the equation at stage 2 accounted for a significant amount of variance on readiness to change ($p < .001$). Message condition had a significant main effect on readiness to change ($p = .045$), showing that participants who receive the persuasive message were more inclined to change their behavior than those in the control condition ($M$ = 6.10, $SD$ = 2.32 and $M$ = 5.67, $SD$ = 2.20, respectively). Pre-existing explicit attitudes were also positively associated with readiness ($p < .001$). These main effects were qualified by a significant three-way interaction between self-referencing condition, message condition, and pre-existing attitudes ($p = .022$). We decomposed the interaction by looking at the effects of self-referencing condition and message condition at less negative ($M$ + 1$SD$) and more negative ($M$ – 1$SD$) pre-existing explicit attitudes towards green vegetables (see Figure 7). For participants with less negative pre-existing attitudes towards green vegetables, there was no significant effect of the self-referencing condition ($\beta$ = -.07, $p = .340$), no main effect of the message condition ($\beta$ = .12, $p = .087$) and no significant interaction ($\beta$ = .06, $p = .372$).
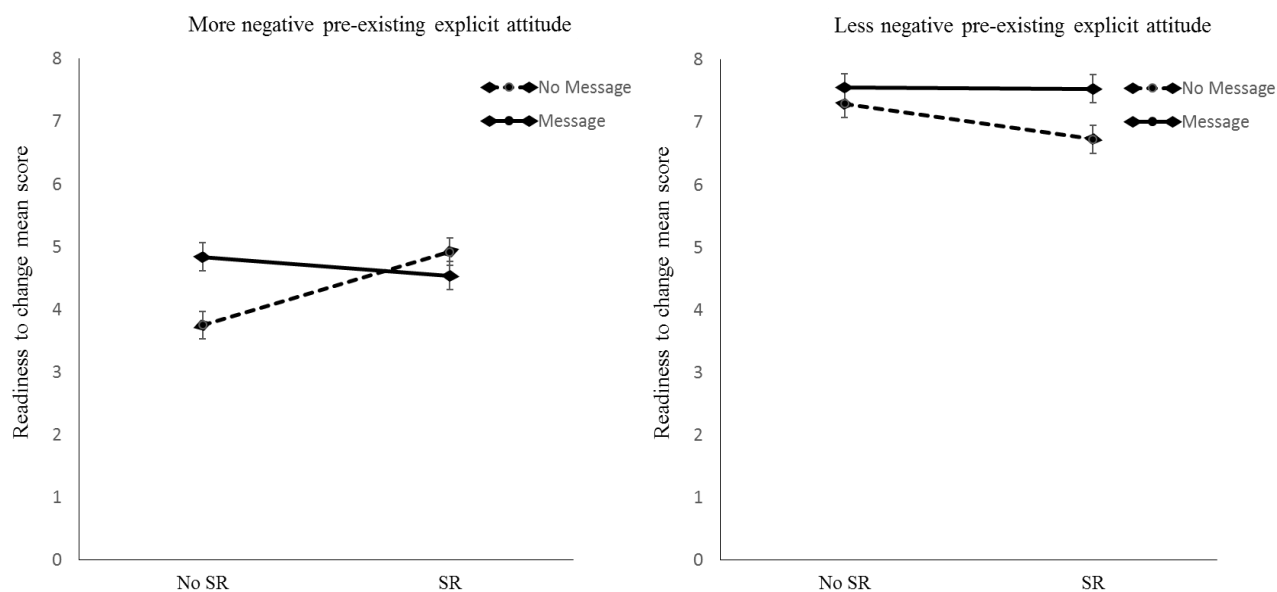
Figure 7. *Effect of Self-Referencing and Persuasive Message Conditions on Readiness to Change by Pre-Existing Attitude.*

The same regression for participants with more negative pre-existing attitudes revealed that neither self-referencing condition ($\beta = .10$, $p = .157$) nor message condition ($\beta = .08$, $p = .262$) predicted readiness to change. However, we observed a significant interaction ($\beta = -.16$, $p = .018$). Simple slopes analysis showed a significant effect of the self-referencing intervention when participants were not exposed to the persuasive message ($B = .59$, $p = .008$) whereas no significant effect was found in the message condition ($B = -.15$, $p = .491$).

**Discussion**

The present study tested the effectiveness of both a reflective and an associative intervention in increasing implicit and explicit attitudes towards green vegetables among a sample with negative to moderately positive pre-existing explicit attitudes. The self-referencing task, in which participants learned to pair green vegetables with the self, increased implicitly measured attitudes, but not self-reported attitudes. On the other hand, a persuasive communication that emphasized the advantages of consuming green vegetables led to changes in explicit, but not implicit, attitudes. This double dissociation pattern whereby the reflective and associative interventions uniquely affected explicit and implicit attitude, respectively, is consistent with both the systems of reasoning analysis of attitude

change (Rydell & McConnell, 2006) and the Associative-Propositional Evaluation model (Gawronski & Bodenhausen, 2006). For instance, according to the Associative-Propositional Evaluation model, persuasive messages add new propositions to the set of propositions that are considered relevant for an evaluative judgment; when the new set of propositions imply a different evaluation, this leads to a change in explicit attitude. Changes in implicit attitude, on the other hand, are driven by changes in the associative structure and require learning a new evaluative association. We observed that a categorization task that required participants to press the same key to categorize self-related words and pictures of green vegetables may lead to such learning and generated more positive evaluations of green vegetables, as revealed by a single-target IAT.

The use of the self to change attitudes is a key innovation of the present research. Previous evaluative learning interventions for food attitudes all consisted in making undesirable food choices (e.g., unhealthy snacks) less attractive after being paired with negative stimuli (e.g., overweight bodies) (Hollands et al., 2011; Houben et al., 2010; Lebens et al., 2011). It was therefore unclear whether attitudes towards food could change due to learning associations with *positive* stimuli. Our study suggests that the self may be an ideal positive stimulus to obtain evaluative learning effects on implicit attitudes when the aim is to promote the consumption of healthy food. We note that the absence of (a) any target category other than green vegetables and (b) a neutral contrast category (i.e., "Other") in the self-referencing task indicate that the changes in implicit attitudes were likely driven by the positivity of the self-concept. It is also unlikely that the results could be due to priming effects as participants in the control condition were exposed to the exact same stimuli as participants who undertook the self-referencing task.

The present study extends previous findings on the self-referencing effect (Prestwich et al., 2010; Perugini et al., 2013) showing that a brief procedure can engender changes in the associative system even for consequential food targets, and for participants whose self-reported attitudes were negative or only moderately positive to begin with. The significant three-way interaction observed for readiness to change is important as it suggests that self-referencing influences not only evaluative,

but also motivational, processes. Moreover, self-referencing influences motivation (readiness to change) when influence is most needed, that is, when participants receive no persuasive message and hold more negative pre-existing attitudes towards green vegetables. In sum, the present study offers new evidence about how to change implicit and explicit attitudes. Our findings indicate that reflective and associative interventions are both needed to successfully change both types of attitude, and demonstrate that negative pre-existing attitudes are no barrier to implicit or explicit attitude change.

As with any new program of study, the present research has limitations. First, the present study used a sample that was recruited online. Additional tests using other recruitment strategies are needed to determine generality. Second, the present study tested changes in implicit and explicit attitudes and readiness to change, but did not test behavior change. It will be important in future studies to examine consumption of green vegetables in the wake of the interventions used here. Consumption could be assessed either straight after the interventions (e.g., via taste test) or in the long term (e.g., via completion of dietary diaries). So doing would provide important evidence concerning both the effectiveness of the interventions deployed here on behavioral change, as well as on the durability of such change. Notwithstanding these limitations, the present research offers new insights about how implicit and explicit attitudes can both be changed, and lays crucial groundwork for future studies that combine interventions targeting different systems to promote health behavior change.

**Chapter 6.**

**I do not like what makes me fat. Testing the Fat-Self-Referencing task on soft drinks' implicit and explicit evaluations and attributes.[14]**

---

[14] This chapter is based on Mattavelli, S., Perugini, M., Richetin, J., & Sheeran, P. (*manuscript in preparation*). I do not like what makes me fat. Testing the Fat-Self-Referencing task on soft drinks' implicit and explicit evaluations and attributes.

**Overview**

Across previous chapters, we showed that relating one object to the self via intersecting regularities has the power to change implicit and explicit liking, one's level of identification with the target and the intention to change behavior according to the acquired preferences. This is possible because of the general positivity of the self-concept (Yamaguchi et al., 2007) that transfer from stimuli related to the self to either novel or existing objects. However, positivity is not the sole characteristic of the self. Instead, given that it represents a central cognitive structure for each individual, information that involve the self can have high impact. We developed this idea in the context of health behavior. For instance, protection motivation theory claims that behavior change is more likely when one feels at risk of a health threat (Rippetoe and Rogers, 1987). One particular method that may increase people's perceptions of personal vulnerability to negative impacts is to show them how health misconducts could affect their own physical appearance. Across two studies ($N = 192$), we engineered a Fat Self Referencing task. Participants paired two alternative soft drinks with either unaltered of fattened pictures of their face. We aimed at changing implicit and explicit liking towards the drink by relying on the salience of the negative information (e.g., the drink relates with a fat version of one's face) learned throughout the completion of the task. Moreover, we tested whether this relation also led to changes a) in implicit and explicit attribution of health-related properties to the soft drinks themselves (i.e., was the drink paired with the Fat-Self perceived as fatter?) and b) in the objective behavior (i.e., the amount of drink taken in a taste test). The paradigm showed reliable effects on implicit and explicit attitude change, on implicit and (partly) on explicit attribution of properties, while it did not affect the subsequent amount of drink tasted by participants. We discussed these findings from a practical perspective in light of the power of the self to increase the accessibility of health related messages in an associative-based procedure. Moreover, we discussed the fact that this study represents the first attempt to test the transfer of extra-evaluative properties via intersecting regularities.

**Introduction**

In the context of health behavior, intervention strategies that operate at the level of one's conscious thoughts have shown little effect in changing behavior (Sheeran, Harris, & Epton, 2014). Because our behaviors can reflect the activation of a pattern of automatic associations in mind, it seems therefore important to use strategies that change associations between a target object and its evaluation and perception to change behavior. Associative learning procedures have been recently proposed as an effective way to change implicit attitudes towards objects that can cause aversive health consequences (e.g., unhealthy snacks, beer, alcoholic drinks). Several studies in this field have relied on Evaluative Conditioning (EC) (Hollands, Prestwich, & Marteau, 2011; Houben, Havermans, & Wiers, 2010; Lebens et al., 2011) procedures that consisted in the repeated pairing between a target object (e.g., pictures of unhealthy snacks) and negative health-related stimuli (e.g., pictures of an obese man). However, more recent contributions have shown that the transfer of liking from valenced stimuli to target object can also occur without experiencing any direct pairing between them. The principle of Intersecting Regularities (Hughes, De Houwer, & Perugini, 2016) is based on the idea that people can infer relations between stimuli by simply learning that those stimuli share a given feature. This principle is operationalized in the Self-Referencing (Prestwich, Perugini, Hurling, & Richetin, 2010) task, a categorization task through which people develop positive attitudes towards a target object as a function of the performance of a common action required to categorize that object and the self. Besides changing its general level of liking, relating concepts with a given good might also lead people to infer a causal relationship (e.g., "this drink makes people fat"). This might reflect in the attribution of specific nutritional properties to the good itself. However, the inferential transfer of extra-evaluative properties has never been tested in health psychology using associative learning interventions.

The present contribution used an associative learning manipulation that capitalizes on a) the IR principle and b) the centrality of the self in human cognition to affect the processing of relations between negative physical characteristics and target goods. We presented participants to either

unaltered (fat) or altered self-images and require them to perform a categorization task where the same action is required for either version of the self-image and one of two alternative soft drinks. We tested the impact of such an associative procedure on implicit motivation to approach or avoid either product as well as on explicit liking. Moreover, we investigated implicit and explicit attribute change towards the products, measuring whether the drink related to the fat version of one's self is also perceived as fatty.

**The importance of automatic associations in health behavior**

Likes and dislikes govern approach and avoidance behavior (Martin & Levey, 1978). Hence, positive attitudes towards unhealthy goods are likely to increase fat intake whereas positive attitudes towards healthy goods should decrease fat intake. Recent literature in social psychology has made a distinction between explicit and implicit attitudes. Explicit attitudes are measured through self-report and are deliberative and propositional in nature (Gawronski & Bodenhausen, 2006). Implicit attitudes, by contrast, are more automatic, impulsive, and primarily based on associations in memory, and are assessed through indirect methods such as reaction-time based tasks. Both theoretical and empirical work support the importance of implicit attitudes in food related behaviors and cognitions. Implicit attitudes positively correlate with self-reported consumption of snacks (Conner, Perugini, O'Gorman, Ayres, & Prestwich, 2007) as well as low-calorie food intake (Maison, Greenwald, & Bruin, 2001). Moreover, implicit attitudes are significant predictors of one's choices among a large variety of food items, such that items receiving greater implicit liking are more likely to be chosen (Perugini, 2005; Richetin, Perugini, Prestwich, & O'Gorman, 2007). If implicit attitudes are significant predictors of food choices, it becomes important to find procedures through which they can be changed. Associative-based procedures appear to be an ideal candidate to modify the automatic associations between concepts that determine implicit attitudes.

**Associative-based interventions in the health domain**

In the general food domain, one of the associative routes through which people learn to like and dislike objects is Evaluative Conditioning (EC, Hofmann, De Houwer, Perugini, Baeyens, &

Crombez, 2010). EC refers to a change in liking of an object caused by its frequent pairing with another stimulus that either holds a positive or negative valence (De Houwer, Baeyens, & Eelen, 1994). In this procedure, the conditioned stimulus has a neutral valence at first and acquires the valence carried by the unconditioned stimulus. Research has shown that EC procedures based on the co-occurrence of food items and images representing negative health-related consequences can change the explicit and implicit evaluations of these food items as indexed by questionnaire measures (Lascelles, Field, & Davey, 2003) and by measures of implicit associations (e.g., Hollands et al., 2011; Houben et al., 2010), respectively. For instance, Hollands et al. (2011) used pictures of negative consequences of unhealthy diet, such as heart disease and obesity, presented in co-occurrence with snacks. They observed that participants in the EC condition were more likely to choose a fruit over a snack at a behavioral choice task administered at the end of the session, and this effect on behavior was mediated by implicit attitude change. Taken together, EC studies suggest that food items paired with negative health-related consequences can change valence. These studies also underline the need to tackle both implicit and explicit attitudes.

**A novel learning mechanism: Intersecting regularities**

However, recent studies proposed that valence can also transfer from one object to the other without experiencing any direct pairing between them. Hughes et al. (2016) argued that people may transfer the valence of one stimulus to another one by merely learning that they share something, following a learning principle defined as Intersecting Regularities (IR hereafter). This IR principle may explain a wide range of phenomena. Imagine you are at the supermarket to buy some cookies and you are very hesitant in deciding to buy cookies of the Brand A rather than of the Brand B. While trying to figure out which one would be a better choice, you realize that Brand B happens to be very similar to the brand depicted on the tee shirt an obese man is wearing at the entrance of the shop. Therefore, the Brand B is likely to acquire negative valence and you may eventually opt for the Brand A cookies, put it in your basket, and checkout. According to the IR perspective, Brand B represents the commonality between a negative source (i.e., being obese in the context of food choice) and a

neutral target of evaluation (i.e., cookies): the obese man and the cookies have never been paired directly, yet they come to be perceived as sharing a (superficial) similarity. This perceived similarity is the commonalitythat allows for the transfer of valence. The IR principle is applied in the Self-Referencing  task (Prestwich et al., 2010), an associative learning task through which people learn to perform a common action to categorize Self-related stimuli and one target, while another action links stimuli related to the category 'Others' and a contrast target. Due to the IR principle and the positivity of the self (Yamaguchi et al., 2007), the target object paired with the self is liked better. Therefore, the SR task exploits the IR principle and the self to create positive attitude towards desirable good. As far as we are aware, there is only one study in literature that used an associative procedure which can be described as based on the IR principle to establish relationship between negative attitudinal sources and health (food) related targets (Haynes, Kemps, & Moffitt, 2015). Haynes and colleagues used one critical block of the IAT to create links between a target object and valenced attributes. They found that implicit evaluations of unhealthy food became more negative in the condition where these food where paired with negative stimuli. The effect on a snack consumption measure was moderated by self-control, with only participants low in inhibitory self-control showing lower snack intake following the food negative training. Although the authors did not explicitly refer to the IR principle, such procedure can be considered as based on that principle because participants need to learn that one common key serves as link between the target and the negative attribute.

**Beyond valence: The transfer of properties**

One way to construe valence is based on the attributes that stimuli possess. In a complex social and physical world, people constantly ascribe attributes to stimuli. For example, people see a person as unfit, taste a peach as juicy, or feel a tissue as soft. However, in many daily life situations people are required to make decisions meant to be driven by attributes even without having the chance to have any sensory experiences. Research on attribute conditioning has shown that individuals can learn that an unknown object possesses a certain attribute through the regular co-occurrence of the object and stimuli possessing the specific attribute (Kim, Allen, & Kardes, 1996; Förderer & Unkelbach,

2011; 2014; 2015; Glaser & Walther, 2013) in an EC-like procedural fashion. For instance, the repeated presentation of a pizza delivery brand with a race car promoted inferential belief information about the speed of the delivery service (Kim et al., 1996). This seems to suggest that once relationships between stimuli are created, also extra-evaluative properties can transfer from one stimulus to another. However, research on this transfer of extra-evaluative properties is confined to a unique type of learning mechanism, that is, the direct pairing between stimuli. No studies have tested whether this could also be the case of learning via IR.

Going back to our example, one could also hypothesize that the IR principle can lead to a transfer of the non-evaluative characteristics of the obese man (i.e., fatness) to the characteristics of the cookies. Following an information processing approach, complex metaphors of human cognition are meant to determine the way in which we assign attribute to objects (Neisser, 1976): cookies belongs to Brand B; Brand B belongs to a fat man; therefore, Brand B is fat. Such complex metaphor may be interpreted as the result of an intersection of regularities concerning two objects, so that one acquires the properties of the other. Therefore, we reasoned that under specific circumstances the IR principle could be a way to induce the perception of a causal relationship between two stimuli linked by a commonality. This could have important practical implications in the context of health behavior.

**Using the centrality of the self to tailor associative interventions to the individuals**

Research suggests that the self is a central psychological structure (Greenwald & Pratkanis, 1984). It has a highly developed and differentiated memory structure (Showers & Zeigler-Hill, 2003), it is chronically active and therefore accessible (Markus & Kunda, 1986), and has a motivational relevance. Moreover, the involvement of personally relevant features affects the processing of a vast number of information. Theories of information processing provide valuable insight into the effectiveness of using self-relevant information to increase the effectiveness of persuasive messages. Petty and Cacioppo's (1986) Elaboration Likelihood Model suggests that people are more likely to process information thoughtfully if they perceive it as personally relevant. When the self is involved, they consider messages carefully, relate them to other information they have encountered,

and compare them with their past experiences. Messages "elaborated" in this way tend to be retained longer and are more likely to lead to permanent attitude change (Petty, Cacioppo, Strathman, & Priester, 1994). For example, self-relevant messages are more likely to be read, remembered, and perceived as interesting (Brug, Steenhuis, Van Assema, & De Vries, 1996; Campbell et al., 1994; Skinner, Strecher, & Hospers, 1994).

Protection motivation theory pushes further on this idea by claiming that behavior change is more likely when an individual feels personally vulnerable or at risk of a health threat (Rippetoe and Rogers, 1987). One particular method that may increase people's perceptions of personal vulnerability to negative impacts is to show them how health misconducts could affect their own physical appearance. The idea of tailoring health messages to the individual by means of threatening visual stimuli has been investigated in the context of smoking behavior (e.g., Flett, Grogan, Clark-Carter, Gough, & Conner, 2015). Flett et al. (2015) tested 30 male smokers' experiences of an appearance-focused (facial-ageing) intervention. Male smokers reported that viewing the impacts of smoking on their own faces was the most effective part of the intervention and 22 men (73%) said that they intended quitting smoking or reducing number of cigarettes smoked post-intervention.

As far as we are aware, no studies have tested whether self-tailored information acquired via associative learning interventions can also affect one's attitudes towards target food objects. Moreover, the reference to the self in an associative learning paradigm that aims at establishing relationship between a target and specific physical characteristics can offer important advantages also from the perspective of a transfer of non-attitudinal properties. If valence is a basic and very salient property (Zajonc, 1980), the transfer of other attributes seems more difficult to occur. For instance, EC research on the transfer of certain non-evaluative properties (e.g., size) from unconditioned stimuli to conditioned stimuli showed that this is possible only if a priming procedure was administered to increase the accessibility of that attribute (Förderer & Unkelbach, 2014; Olson, Kendrick, & Fazio, 2009). Given the properties mentioned above, making the attribute relevant for

one's self seems to be a perfect candidate to strengthen its accessibility, which is meant to be acquired by a related target object.

**Overview of the studies**

The present chapter presents two studies testing the effect of modified (i.e., virtually fattened) self-images in changing implicit and explicit attitudes (Study 1) as well as proximal attributional properties (Study 2) of two alternative soft drinks, by using an IR-based learning procedure. The consumption of sugar-sweetened beverages has increased in the United States over the past three decades (Andreyeva, Chaloupka, & Brownell, 2011). Sodas, or carbonated soft drinks, are consumed more than any other sugar-sweetened beverage and remain the largest source of added sugar in the diet (Welsh, Sharma, Grellinger, & Vos, 2011). These beverages have been linked to weight gain, diabetes, hypertension, hyperlipidemia, gout, and coronary artery disease (Malik, Popkin, Bray, Despres, & Hu, 2010; Shulze et al., 2004). Therefore, an increasing demand for intervention strategies that could help in reducing the intake of sweetened beverages is needed. To do so, we used a modified version of the SR task (Prestwich et al., 2010), namely Fat-Self-Referencing. The aims of the present contribution are threefold. First, we extend the findings observed in Haynes et al. (2015) and test whether both implicit motivation and explicit attitudes towards soft drinks can be changed through the performance of a common action (IR) to categorize one's negative self with either the two drinks. Second, we rely on this learning principle to examine if the Fat-Self-Referencing may also lead to a change in the perception of the properties of the two drinks. Finally, we assess objective behavior and test whether the amount of drink consumed by participants depends on its relation with either one's fat or unaltered self.

**Study 1**

In this study, we tested the effect of a Fat-Self-Referencing procedure on implicit and explicit evaluation of two extant soda brands. We adopted a one factor (Experimental manipulation: Fat Self + Drink A vs. Fat Self + Drink B vs. control) between-subject design. Implicit attitude was assessed through an approach/avoidance IAT. To measure explicit attitudes, we used a relative and an absolute

evaluation of the two drinks. Because both explicit attitudes and implicit approach/avoidance motivation are both important predictors of health-related behaviors (e.g., Palfai & Ostafin, 2003), showing an effect of the manipulation on both implicit motivation and explicit attitude would be a first important indicator of the power of the Fat-Self-Referencing to influence behavior. Even though we did not pre-test for their evaluation, two brands, namely Shasta and Faygo, were chosen as target stimuli because of their high similarity in most of their visual features. Furthermore, associative learning effects resulting from IR, like the Self-Referencing effect, have shown to be affected by one's ability to recall in mind the correct relationship between source and target objects (Mattavelli, Richetin, & Perugini, 2016). For this reason, an IR memory question was administered right after the associative manipulation.

**Method**

### Participants and procedure

One hundred and ten university students (71 women and 39 men, *M* age = 19.02, *SD* = 1.44) took part individually in a one-session study presented as a categorization task. Participants were randomly assigned to one of three different conditions. In the first two conditions, they completed a Fat-Self-Referencing task. Before starting the experiment, participants assigned to these two conditions were told that photographs of them would be taken, altered, and used in the computer task they would have lately performed. To take the pictures, the experimenter used the FatBooth application on a smartphone. For each participant, ten pictures were taken, five to be used as the real-self stimuli and the other five as the negative-self stimuli. Depending on the Fat-Self-Referencing condition they were assigned to, participants paired the drink "Shasta" versus the drink "Faygo" with their unaltered self-images and the brand "Faygo" versus "Shasta" with fattened-self images. Then they completed a IR memory assessment right after the Fat-Self-Referencing for half of them or after an approach/avoidance Faygo/Shasta IAT (Greenwald, McGhee, & Schwartz, 1998), two

thermometer measures, and a relative evaluation of drinks[15]. In the third condition, participants skipped this task and completed directly the outcome measures. After that, they completed a measure of familiarity with the drinks and a Self-Affirmation (Reed & Aspinwall, 1998) scale[16]. Then they were thanked and debriefed by the experimenter.

**Materials**

*Fat-Self-Referencing task.* The task was an adapted version of the common Self-Referencing (Perkins & Forehand, 2012; Prestwich et al., 2010) task. Participants first completed one block of 60

---

[15] Participants who responded the memory questions immediately after the task were also asked to respond an Intersecting Regularities memory question as a further indicator of their ability to recall in mind the correct source-target pairing: "*The task you have just completed required you to press the same key to categorize pictures of your altered face and one drink. Do you remember which drink?*" They could select one of the two drinks or the option 'I don't remember'. Responses were classified as correct or incorrect IR memory. This IR memory question is in line with the standard question used in the Self-Referencing task and was administered to see whether it could qualify the effect of the manipulation on the outcome measures. However, the analyses revealed no significant interaction between the experimental manipulation and the time of assessment of the IR memory question on all the three outcomes (all $p$s > .559)

[16] For the Self-Affirmation scale participants indicated whether they had ever performed a series of 10 behaviors (yes/no), and if they had, to provide a brief written example. Sample items are, "*Have you ever been considerate of another person's feelings?*" and "*Have you ever attended to the needs of another person?*" Thus, an affirmative response to the scale items required the endorsement and recall of common and relatively minor acts of kindness, not great feats of altruism or self-sacrifice. This measure was used to rule-out the risk of any emotional distress participants could experience throughout the study.

trials in which they categorized, as quickly and accurately as possible, Shasta versus Faygo stimuli (i.e., pictures of a can) and pictures related to the unaltered self with one response key (e.g., 'E') and Faygo versus Shasta stimuli and pictures related to altered self to a different response key (e.g., 'I'). Participants then repeated the block of 60 trials after switching the keys assigned to each category (i.e., Shasta vs. Faygo stimuli and self-altered pictures assigned to the 'I' key, and Faygo vs. Shasta stimuli and self pictures to the 'E' key). The order in which participants completed these two blocks was counterbalanced between participants. In case of incorrect classification, a red-X appeared on screen and remained until correction. The inter-trial interval was 400ms. For each category, we used a single stimulus (see Supplementary Materials Chapter 6 – Appendix N).

*Approach/avoidance IAT.* Participants categorized words presented individually and in a random order in the middle of the screen using two keys (i.e., 'E' and 'I'). The target concept was Shasta and its contrast was Faygo, whereas the attribute categories were words related to either approach (e.g., *forward, advance, closer*) or avoidance (e.g., *escape, withdraw, away,* see Supplementary Materials Chapter 6 – Appendix N). The order of the two critical blocks was counterbalanced between participants, with half of the participants having the combination Shasta and Approach being presented first and the other half having the combination Faygo and Approach being presented first[17]. We administered a 7-block version of the IAT (Greenwald, McGhee, & Schwartz, 1998). All practice blocks consisted of 20 trials and each critical block consisted of 40 trials. 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

---

[17] The order in which the two critical blocks appeared in the IAT was nested with the position in which Shasta and Faygo appeared in the SR task. The drink that appeared on the left first in the SR was the one approached in the first critical block of the IAT. For instance, if Shasta appeared on the left first in SR, irrespective of whether it was paired with either real or altered self-faces, then that was the target to be paired with approach words in the first critical block of the IAT.

target category, we used five words. We calculated a D600 score following the procedure suggested by Greenwald et al. (2003), such that higher IAT scores indicated a greater approach towards Shasta compared to Faygo.

*Thermometer measure.* Participants indicated their level of liking towards either the two drinks separately entering manually a number below a thermometer ranging from 0 ("*Don't like it*") to 100 ("*Love it*"). We computed the difference between the two thermometer scores, such that higher scores indicate a preference for Shasta over Faygo. The order of administration of the two thermometers was randomized.

*Relative evaluation of drinks.* Participants expressed their explicit preference for one drink over the other on an 11-point slider anchored with the name of the two drinks. For half of participants, 'Faygo' and Shasta were located at the right and left extremes of the slider, respectively while for the other half, the sides were reversed. This method factor was randomly counterbalanced between participants.

*IR memory assessment.* Participants provided their recall of the correct source-target pairing in the SR task by answering four questions that combine the different possibilities of the Fat-Self-Referencing task: "*During this study, what percent of the time was the category Shasta (Faygo) paired with the category self (altered self)*?" For each question, participants indicated the percentage they believed was correct by selecting one of ten percentage intervals (i.e., 0-10%, 11-20%, 21-30%, 31-40%, 41-50%, 51-60%, 61-70%, 71-80%, 81-90%, 91-100%). We computed an index of memory through the following steps. First, we transformed interval responses into 1-10 scores. Then we subtracted the score resulting from the two questions involving the two incorrect pairings (e.g., Faygo + Self and Shasta + Fat Self in the SR condition Faygo + Fat Self) to the score from the two questions about the correct ones (e.g., Faygo + Fat Self and Shasta + Self in the SR condition Faygo + Fat Self). Positive scores indicate that participants remembered the correct source-target pairings they have been exposed to.

*Familiarity measure*. The familiarity measure consisted of four questions. The first two questions stated as follows "*Had you ever heard the soft drink Shasta (Faygo) before?*" The other two questions asked participants whether they had tried the two drinks before the experiment "*Had you ever consumed the soft drink Shasta (Faygo) before?*" Participants indicated their familiarity with the drinks by answering "*Yes*" or *"No"* to the familiarity questions and "*Yes*", *"Yes but not regularly"* or *"No"* to the two direct experience questions . An overall indicator of familiarity was obtained by coding as -1 participants either familiar or with previous experience with the drink Faygo but not Shasta, as 0 in case none of the two drinks was either known or tasted (or if both were to equal extent), while the score 1 was used for participants either familiar or with previous experience with the drink Shasta, but not Faygo.

## Results

We first looked at the effect of the Fat-Self-Referencing on implicit attitude score (see Table 14 for descriptives). We ran a one-way-ANOVA on the IAT score considering the three conditions (Shasta + Fat Self vs. Faygo + Fat Self vs. Control). There was a significant effect, $F(2,107) = 5.40$, $p = .006$, $\eta_P^2 = .09$. As follow-up analyses, we ran three independent samples $t$-tests. There was no significant difference between the Fat-Self-Referencing condition 'Shasta + Fat Self' and the control group, $t(72) = .42$, $p = .674$. The difference between the Faygo + Fat Self and the control condition was significant, $t(60) = -2.87$, $p = .006$. Crucially, we observed a significant effect of condition when comparing the two Fat-Self-Referencing groups, $t(82) = -2.88$, $p = .005$, with participants in the Faygo + Fat Self condition showing higher IAT scores. When Faygo was paired with the fat version of the self, it was implicitly avoided more than Shasta.

Then we tested the Fat-Self-Referencing effect on the first explicit relative preference for one drink over the other one (i.e., slide bar). A one-way-ANOVA revealed a significant main effect, $F(2,107) = 7.30$, $p = .001$, $\eta_P^2 = .12$.

Table 14. *Study 1 means and standard deviations.*

|  | Shasta+Self / Faygo+FatSelf $n = 36$ | Faygo+Self/ Shasta+ FatSelf $n = 48$ | Control $n = 26$ |
|---|---|---|---|
| Approach/avoidance IAT score | .17 (.26) | .00 (.28) | -.03 (.29) |
| Relative Explicit Attitude score | 6.97 (1.77) | 5.46 (1.77) | 6.12 (1.71) |
| Thermometer differential score | 11.89 (24.20) | -4.98 (20.53) | .15 (19.47) |

The condition Shasta + Negative Self and the control group were not significantly different, $t(72) = -1.49$, $p = .141$. The difference between the Faygo + Fat Self and the control condition was marginally significant, $t(60) = 1.91$, $p = .060$. Moreover, the two Fat-Self-Referencing conditions were significantly different, $t(82) = 3.76$, $p < .001$.

For the second explicit attitude score, we again observed a significant main effect, $F(2,107) = 6.38$, $p = .002$, $\eta_P^2 = .11$. The Shasta + Fat Self condition and the control group were not different, $t(72) = -1.05$, $p = .299$ but the difference between the Faygo + Fat Self and the control condition was significant, $t(59,24) = 2.11$, $p = .039$. Again, the two Fat-Self-Referencing conditions were significantly different, $t(82) = 3.45$, $p = .001$. Note that the two explicit scores were highly correlated, while neither of the two significantly correlated with IAT score (see Table 15).

These results show an effect of the manipulation on both implicit approach/avoidance and explicit attitude. Importantly, follow-up analyses showed a genuine difference between the two Fat-Self-Referencing conditions on all outcomes. When Shasta is categorized through the same action as the Fat Self, participants implicitly approach Faygo more than Shasta and show a greater explicit preference for Faygo, while the opposite pattern occurs in the alternative Fat-Self-Referencing condition[18].

---

[18] ANCOVAs were run to see whether the effect of the experimental manipulation in the two SR condition was still there when controlling from previous knowledge. The analysis revealed no main

Table 15. *Study 1 Correlations between measures.*

|  | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1.Experimental condition | 1 | | | |
| 2.Approach/avoidance IAT score | .25** | 1 | | |
| 3.Relative Explicit Attitude score | .32** | .04 | 1 | |
| 4.Thermometer differential score | .35** | .09 | .79** | 1 |

In sum, the study shows that the mere act of categorizing through the same key (IR) one drink and one's own fat face leads to negative attitude towards that drink. One can conceptualize this effect as a function of the perceived relationship between the drink and some negative physical consequences related to it throughout the task. However, this interpretation suffers from one main limitation. The structure of the associative intervention (Fat-Self-Referencing task) does not rule-out the possibility that the preference for one drink over the alternative one is driven by the general positivity of the unaltered self that transfers to the other drink instead of the negativity resulting from the intersection between one drink and the fat self. Therefore, a Fat-Self-Referencing effect resulting from the IR established between one's fat self and a drink should also lead to stronger associations between that drink and the portrayed physical consequences. In study 2, we administered an implicit measure that tests the strength of the association between the two drinks and fat-related words, as well as the

---

effect of previous knowledge on implicit approach avoidance, $F(1,40) = .42$, $p = .520$, while there was a significant effect on both slider and thermometer measure, $F(1,40) = 5.61$, $p = .023$, $\eta_P^2 = .12$ and $F(1,40) = 6.23$, $p = .017$, $\eta_P^2 = .14$, respectively. The effect of group on the two explicit attitude measure was significant when controlling for previous knowledge, $F(1,40) = 13.09$, $p = .001$, $\eta_P^2 = .25$ and $F(1,40) = 4.18$, $p = .048$, $\eta_P^2 = .10$, respectively. No significant interaction between experimental condition and previous knowledge emerged (all $p$'s > .053).

assessment of the perceived amount of calories of the drink and how long it would take to walk those calories off.

**Study 2**

In the previous study, we demonstrated that when a drink is paired with a modified (fat) version of one's face participants tend to implicitly avoid that drink, and exhibit a more negative explicit attitude in comparison to a drink paired with the normal version of the self. The aims of study 2 are twofold. First, by assessing the association between the target drinks and the concept of fatness at both implicit and explicit levels, we want to see whether the paradigm can also lead to a transfer of the non-evaluative properties carried by the fat self to the drink related with it. We tested implicit transfer of attributes through a Brief IAT (Sriram & Greenwald, 2009). As measures of explicit attitudes, we asked participants to estimate the amount of calories of the target drink and the amount of time required to walk them off. An additional measure assessing sensory and extra-sensory perceived characteristics of the drink was administered after a tasting test. With the tasting implicit transfer of attributes through a Brief IAT (Sriram & Greenwald, 2009). As measures of explicit attitudes, we asked participants to estimate the amount of calories of the target drink and the amount of time required to walk them off. An additional measure assessing sensory and extra-sensory perceived characteristics of the drink was administered after a tasting test. With the tasting test, we aimed to extend previous findings on a behavioral measure in which participants taste one of the two drinks previously categorized in the Fat-Self-Referencing. We expect participants to consume less of the drink if it was previously paired with their fat self. In light of the previous results on attitude from Study 1 and the postulated effect on the transfer of attribute properties, we hypothesized that pairing the drink and the negative self would influence participants' consumption of that drink. Two additional method changes were introduced in the study with the aim of increasing participants' ability to learn and recollect the IR principle underlying the task. In study 1, when comparing the two Fat-Self-Referencing conditions, the effect size of the manipulation increased on all outcomes when considering only the subsample ($N = 47$, 56%) of participants who were capable to recollect the type

of face each drink was paired with (IAT: from $d = .09$ to $d = .18$; Thermometer measure: from $d = .15$ to $d = .28$; Relative explicit evaluation: from $d = .13$ to $d = .14$). Hence, whereas in the first study we resorted to an abbreviated version of the SR manipulation, in this study we used the standard version therefore administering a total of 160 trials divided into 80 practice and 80 test trials (Prestwich et al., 2010). Second, we adopted an overall simpler measure of memory of the intersecting regularities, asking participants to indicate which drink was categorized through the same action as the altered version of themselves right after the Fat-Self-Referencing (e.g., Perugini, Richetin, & Zogmaister, 2014).

**Method**

### Participants and procedure

Eighty-two university students (44 women and 38 men, $M$age = 19.00, $SD$ = 1.70) took part individually in a one-session experiment. All participants were instructed not to drink or eat anything during the hour preceding the experimental session. The experimenter presented the study as a marketing research that aimed at assessing participants' opinion towards drinks. Participants were randomly assigned to one out of two versions of a Fat-Self-Referencing task. Depending on the condition they were assigned to, participants paired the drink "Shasta" versus the drink "Faygo" with unaltered self-images and the brand "Faygo" versus "Shasta" with altered self-images. All participants then answered an IR memory question, completed a Brief IAT (Sriram & Greenwald, 2009) assessing the strength of the association between fat related words and the two drinks. Half of the participants then completed the measure of explicit attributes followed by the behavior measure and the post-test evaluation of the drink whereas the other half completed the measures in the opposite order[19]. After that, they were thanked and debriefed by the experimenter.

---

[19] Familiarity with the drinks was also assessed, as well as a self-affirmation questionnaire, like in Study 1.

**Materials**

*Fat-Self-Referencing task*. The structure of the task was identical to that used in study 1 except for the administered number of trials (i.e., 160 trials).

*Intersecting regularities memory*. Participants answered the following question: *"The task you have just completed required you to press the same key to categorize pictures of your altered face and one drink. Do you remember which drink?"* The three answer options were "*Shasta*", "*Faygo*" and "*I don't remember*". Responses were classified as either correct or incorrect IR memory.

*Attribute BIAT*. This BIAT (Sriram & Greenwald, 2009) is composed of two combined-task blocks, each of which characterized by its two focal categories (e.g., *Shasta* and *Fattening* and *Faygo* and *Fattening*). One category is focal in both combined tasks (*Fattening*) and has a contrast non focal category (e.g., *Slimming*). The BIAT is named, with the non-focal category in parentheses. High scores at this *Shasta-Faygo/Fattening-(Slimming)* BIAT indicate greater strength of associations between *Shasta-Fattening* than between *Faygo-Fattening*. The instructions for the BIAT blocks display all exemplars for the upcoming block's two focal categories. After the instruction page, the lists of focal-category exemplars disappeared, but the focal category labels remained in view. Each of the blocks of each BIAT consisted of 40 trials (plus 4 initial dummy trials). On each BIAT trial, an exemplar of one of the four categories appeared in center screen. The inter-trial interval was 400ms. For each category we used four distinct stimuli, with pictures representing the two drinks and word stimuli for Fattening (*Fattening, Bigger, Heavier, Chubby*) and the category Slimming (*Slimming, Smaller, Lighter, Slender*, see Supplementary Materials Chapter 6 – Appendix N). The order in which participants completed the two blocks was counterbalanced across participants.

*Explicit attributes measures*. Participants indicated the amount of calories they thought were contained in the drink Shasta, with the picture of the drink being visible in the center of the screen ("*How many calories do you think are in this Shasta can?*"). A second question asked "*How long (in hours and minutes) do you think it would take to walk off the calories in this Shasta can?*"

*Behavior measure*. In the product-testing phase, a 355 ml Shasta can was placed in front of each participant. The can was weighted before. Participants were told to take all the time they needed to taste the product and to rate it on different dimensions (i.e., tastiness, naturalness, healthiness, sweetness, sourness, bitterness, product look, and overall liking) on seven-point scales from 'Not a lot' to 'Very much' (Taylor, Webb, & Sheeran, 2014). They also completed a Sensory Evaluation Form (See Supplementary Materials Chapter 6 – Appendix O). Once the participants left the lab, the experimenter weighed the can again to have an objective indicator of the total amount of drink consumed by the participants.

**Results.**

The data from six participants were excluded from the analyses because of their high percentage of errors ($> 25\%$) made in the BIAT. We also excluded data from participants who failed in providing the correct response at the IR memory questions ($N = 5$, 6%). Moreover, one participant who took the whole drink during the tasting phase was treated as clear outlier (consumption $z > |5.79|$) and his/her data were excluded from the analyses, leading to a final sample of seventy participants (see Supplementary Materials Chapter 6 – Appendix M for analyses on the whole sample).

We first looked at the effect of the manipulation on the implicit association between fat-related words and the drinks. The D score was calculated such that higher scores indicated stronger associations between Shasta and Fattening. We ran a one-way ANOVA and found a significant main effect of the Fat-Self-Referencing condition on the BIAT score, $F(1,68) = 8.98$, $p = .004$, $\eta_P^2 = .12$[20]. Participants were faster in categorizing fat words with Shasta when it was paired with their fat face, suggesting that the Fat-Self-Referencing leads to stronger implicit associations between the drink and

---

[20] We ran a one-way ANCOVA entering two method factors (Position in which Shasta appeared first in the SR and Order of the critical blocks in the BIAT) as covariates in the model. The main effect of the SR condition on the BIAT score was still significant, $F(1,66) = 9.31$, $p = .003$, $\eta_P^2 = .12$, and there was no significant effect of the two covariates ($p > .082$).

fat related concepts. The means revealed also that in both conditions the BIAT score was negative. Even though there was a main effect of Fat-Self-Referencing, there seems to be a baseline association between Faygo and fat-related words, perhaps due to the similarity in terms of wording.

Then we tested the effect of the Fat-Self-Referencing on the amount of calories attributed to a 355ml Shasta's can and the amount of time required to walk them off. A one-way ANOVA revealed that even though the means followed the predicted pattern (see table 16 for descriptives), with higher amount of calories attributed to Shasta when it was paired with the altered self, the effect of the Fat-Self-Referencing was not significant, $F(1,68) = 1.54$, $p = .220$, $\eta_P^2 = .02$. There was no Fat-Self-Referencing effect on walking time, $F(1,68) = .25$, $p = .621$, $\eta_P^2 < .01$[21]. The difference between the initial and the final weight of the 355ml Shasta's can was our behavior criterion. A one-way ANOVA showed that, against our hypothesis, no effect of the experimental manipulation on the total amount of Shasta drunk, $F(1,68) = .94$, $p = .336$, $\eta_P^2 = .01$.

---

[21] We also ran a one-way ANCOVA entering an additional method variable as covariate in the model (i.e., Order in which the explicit question were administered: before vs after the behavioral measure). The effect of the SR on the amount of calories was not significant, $F(2,65) = 1.20$, $p = .278$, $\eta_P^2 = .02$. Among the covariate variables, we only found a significant effect of the order of administration of the calories question, $F(1,65) = 6.31$, $p = .015$, $\eta_P^2 = .09$. Participants tended to attribute greater amounts of calories to the drink when the question was administered before the behavior test. However, this method variable showed no interaction with the SR condition, $F(1,64) = .06$, $p = .802$, , $\eta_P^2 < .01$. We observed no SR effect on walking time, $F(1,65) = .29$, $p = .593$, $\eta_P^2 < .01$. There was no significant effect of the three method variables ($p$s > .521). We also ran a one-way ANCOVA to test the effect of the main experimental manipulation on the consumption of the drink, with the three method variables mentioned before as covariates. The effect of the SR was not significant, $F(1,65) = .92$, $p = .342$, $\eta_P^2 = .01$. There was no effect of the covariates ($p$s > .635).

Table 16. *Study 2 means and standard deviations on IR memory participants.*

|  | Shasta+FatSelf / Faygo+Self $n = 37$ | Faygo+FatSelf/ Shasta+Self $n = 33$ |
|---|---|---|
| BIAT score | -.11 (.48) | -.49 (.57) |
| Amount of Calories | 174.19 (63.83) | 154.85 (66.67) |
| Walking Time (min.) | 85.14 (54.94) | 91.52 (52.21) |
| Consumed Drink (ml) | 69.70 (38.28) | 61.06 (38.75) |
| Taste pleasantness | -.04 (.90) | .04 (1.11) |
| Extra-Taste properties | -.20 (1.00) | .21 (.98) |

Except for amount of calories and walking time ($r = .39$), there was no significant correlation among the four dependent variables (see table 17).

We also looked at whether pairing Shasta with one's fat face would influence the ratings of the drink after the tasting phase. We ran a Principle Component Analyses on all thirteen items. A two factors solution explained 48.88% of the overall variance. Given the weak correlation between the two factors ($r = .17$), we performed a Varimax rotation. Three items of the scale (i.e., Sourness, Bitterness and Aroma/Smell) showed no clear loadings and were thus excluded. The new model explained 58.68% of variance. The first component explained 39.87% of variance and was interpreted as 'Taste pleasantness' and included item assessing overall liking, overall acceptability, tastiness, sweetness, taste/flavor and texture (factor loadings from .56 to .86). The second component explained 18.80% of variance and pertained to the further properties of the drink related to its healthiness and looking (extra-tasting properties). They included naturalness, product looking, healthiness and appearance (factor loadings from .61 to .76). A first one-way ANOVA showed that despite we observed that the means were in the expected direction, with Shasta judged as less pleasant after its consumption, the effect of the Fat-Self-Referencing condition was not significant, $F(1,67) = .11$, $p = .747$, $\eta_P^2 < .01$.

Table 17. *Study 2 correlations between measures.*

|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1. Experimental condition | 1 | | | | | | |
| 2. BIAT score | .34** | 1 | | | | | |
| 3. Amount of Calories | .14 | .19 | 1 | | | | |
| 4. Walking Time | -.06 | .03 | .39** | 1 | | | |
| 5. Consumed drink | .00 | .15 | -.01 | .03 | 1 | | |
| 6. Taste pleasantness | -.05 | -.11 | .08 | .07 | .23 | 1 | |
| 7. Extra-Taste properties | -.22 | -.09 | -.20 | -.04 | .02 | .00 | 1 |

Similarly, the effect of the Fat-Self-Referencing on the extra-tasting properties of the drink did not reach the conventional level of significance, $F(1,67) = 2.98$, $p = .089$, $\eta_P^2 = .04$, but at the same time showed an interesting trend. The means indicated that for articipants in Shasta – Fat Self condition, more negative extra-tasting properties were attributed to the drink (see Table 16). Even more central to the present work, the effect of the manipulation significantly affected the perceived healthiness of the drink, $F(1,67) = 4.28$, $p = .043$, $\eta_P^2 = .06$. The effect was in the expected direction, such that in the Shasta – Fat Self condition the drink was evaluated as less healthy ($M = 2.08$, $SD = .98$) than in the opposite condition ($M = 2.61$, $SD = 1.22$).

Overall, study 2 provides an additional demonstration that the Negative Self-Referencing represents a novel and effective way to create implicit association change, not merely at the evaluative level, but also with respect to further attribute properties. Our results do not support a transfer of properties at the explicit level, nor a consequent change on actual behavior. Although ancillary to our analyses, results concerning the effect of the manipulation on the subsequent rating of the drink after tasting may suggest that the act of pairing a drink with one's fat image can play a role in altering the judgment

of both taste and extra-taste properties attributed to the product. These altered judgments could be of importance, for instance, in determining one's intention to choose (or to avoid) a certain product.

**General discussion**

The present work tested the impact of an evaluative learning manipulation based on associations with the negative self in affecting evaluation, attribution of properties, and consumption of soft drinks. The first study showed a significant effect of the Fat-Self-Referencing manipulation on all the outcomes assessed. Participants pressed the same key for fattened self-pictures and one soft drink developed a higher tendency to avoid implicitly that drink, compared to those who performed the same action for the same drink and standard pictures of themselves. Moreover, participants disliked the drink paired with the negative self at explicit level, too. These results provide empirical support to previous demonstrations of the effectiveness of associative learning in changing attitude towards food and health related target (Hollands et al., 2011; Houben et al., 2010; Lebens et al., 2011). An important element of novelty is in the learning principle meant to underlie the changes in implicit and explicit liking. Hence, our research provides additional proof to the fact that stimuli pairing (e.g., EC) is not the only pathway we might pursue to change implicit and explicit attitude in a health related context through the associative route. The effect of the manipulation on implicit approach/avoidance is also of importance. In fact, previous research in the context of alcohol use showed that approach/avoidance IAT scores were associated to subjective responses to alcohol cues, thus suggesting the idea that the measure may serve as useful indicator of alcohol use motivation (Palfai & Ostafin, 2003). We can therefore state that our Fat-Self-Referencing manipulation might affect one's motivation to engage in health related behaviors. Finally, there was an effect on explicit preference: to our knowledge, out of the five studies assessing evaluative conditioning effect on explicit attitude toward health related targets, only one of them succeeded in finding a significant effect of the manipulation (Houben et al., 2010).

The second study broadens these findings by showing that when participants learn that a trivial element is shared between one drink and the negative self, the salient attribute portrayed in the

negative self (i.e., fatness) transfers implicitly to the drink itself. The transfer of properties other than valence from sources to the target objects (i.e., attribute conditioning) has been demonstrated in previous studies (e.g., Kim et al., 1996; Förderer & Unkelbach, 2011; 2014; 2015). Moreover, some researchers showed that attribute conditioning effect can be qualified by some boundary conditions, among which, the accessibility of the attribute: attributes need to be salient to transfer from one object to the target (Förderer & Unkelbach, 2014; Olson et al., 2009). We believe that applying the to-be-transferred attribute to one's own self-image might be a way to increase its accessibility, given that the self is a central structure and chronically accessible (Markus & Kunda, 1986). Future research might further investigate this hypothesis by comparing a Fat-Self-Referencing condition with another condition in which, for instance, other's fat faces are paired with the drink at issue.

We demonstrated an effect of the Fat-Self-Referencing manipulation on implicit transfer of attribute, while no effect emerged when participants where explicitly asked to indicate the amount of calories contained in each drink. Following the Associative-Propositional Evaluation model (Gawronski & Bodenhausen, 2006; 2011), this dissociative pattern of findings might suggest that the structure of the learning paradigm is effective in strengthening the mental associations between the negative attribute depicted in the modified picture and the drink categorized through the same action. On the other hand, further propositional validations are required to alter the actual perception of the drink's properties when participants are overtly asked about them. Having said that, it remains worth noticing that the Fat-Self-Referencing effect on the evaluation of the drink after their tasting test was marginally significant when focusing on taste-unrelated properties, and fully significant on healthiness. This finding is encouraging for the possibility of altering the perception of those properties that are made salient through the manipulation. Moreover, there are reasons to believe that by increasing the sample size we could get a significant effect of the manipulation on the evaluation of the drink. More in general, the idea that a Fat-Self-Referencing task based on the IR between a target object and the fat self can alter the perception of the properties of that target is new and potentially interesting.

Finally, our hypothesis to reduce the consumption of the drink as a function of its prior association with negative health related consequence was not confirmed. In trying to understand which factors might have contributed in preventing the effect, one possibility is that our behavioral measure was presented to participants as a tasting test, which does not reflect a standard drinking behavior. In fact, the amount of drink they tasted could be affected by the final goal to provide a proper evaluation of the drink itself. A more realistic environment could help in reducing the impact of external factors. Participants could taste the drink while performing another activity (e.g., watching TV) or alternatively the behavior assessment could be an economic decision (e.g., deciding to buy or not the drink after receiving some money as a prize). Moreover, Haynes et al. (2015) recently failed to show a main effect of an evaluative learning manipulation on the subsequent consumption of snack while revealed that only participants who were low in inhibitory self-control had lower snack intake following the food negative training. Our design did not take into account dispositional factors that may play a role in determining a main effect of the Fat-Self-Referencing manipulation. Future investigations are needed in order to test whether inter-individual characteristics may play a role in determining the impact of the Fat-Self-Referencing manipulation on the subsequent consumption of drinks.

In sum, these studies show that by simply learning that visual stimuli depicting negative health related consequences share a trivial feature with a drink changes the evaluation of the latter, the implicit tendency to approach (or avoid) it. This finding confirms that IR is an effective way through which people can acquire relations between food-related items and certain physical characteristics. In fact, these established relationships not only affected individuals' liking towards the target soft drinks, but were also impactful with respect to the implicit transfer of properties and, partially, the sensory evaluation of the drinks. Importantly, we observed these findings by using visual stimuli that referred to the individuals, hence exploiting the centrality of the self-concept. These results could pave the way for further investigations testing associative learning interventions in the context of health behavior that are tailored to the individual's self.

**General discussion**

Knowing, understanding how attitudes form and change is important to influence and predict human behavior. Besides the role of overt preferences, previous research outlined the importance of implicit attitudes in predicting behavior in different contexts (e.g., Greenwald, Poehlman, Uhlmann, & Banaji, 2009). The present dissertation examined the impact of an associative learning procedure in affecting both evaluative and extra evaluative properties (i.e., behavioral intention, identification, and attribution of properties). Research on associative learning paradigm has been dominated by evaluative conditioning, where a change in liking towards an object is conceived as the effect of repeated pairings between that object and a valenced stimulus (Levey & Martin, 1975; see Hoffman, De Houwer, Perugini, Baeyens, & Crombez, 2010 for a review). In EC, the actual pairing between two stimuli represents the environmental regularity that links two objects (De Houwer, Gawronski, & Barnes Holmes, 2013). More recently, Hughes, De Houwer and Perugini (2016) proposed a novel pathways of associative learning based on environmental regularities that intersect one another. This dissertation focused on associative learning via IR, with a special attention on the SR task. Specifically, the SR task (Prestwich, Perugini, Hurling, & Richetin, 2010) was presented as a prime exemplification of IR-based associative learning paradigm. In the present work, a description of the main characteristics of the task and its implications from both a practical and a more theoretical perspective were presented through a meta-analytical approach. The next parts developed the idea of considering learning via IR as an indirect form of attitude change. This idea of indirect transfer of liking in the SR was tested by focusing on two specific aspects, that is, the centrality of the self and the possibility to exploit the IR principle to build complex chains of intersection between stimuli. If focusing on the self answers the *what* question (i.e., "*What drives the SR effect?*") and testing learning via IR answers the *how* question (i.e., "*How does the SR effect occur?*"), the second and more applied part of the dissertation focused on gathering empirical evidence of the ecological validity of the SR task.

**The SR effect and IR memory as its main boundaries**

The meta-analytical summary of the SR studies conducted in our lab offered us a comprehensive way to introduce the task, defining its main characteristics and identifying the most important peculiarities that were analyzed at the level of the individual studies in the rest of the present work. In trying to summarize the most important points raised by the meta-analysis, the issue concerning participants' ability to remember what target has been paired with the self throughout the task (measured after it with an IR memory question) deserves special attention. Besides its role in qualifying the SR effect, this finding is relevant from the perspective of evaluative learning through associative procedures. Our reasoning is organized around the idea that IR memory in the SR task is the formal equivalent of the most common way to measure contingency awareness in evaluative conditioning. In fact, in both cases, participants' responses provide indications on whether they noticed that one object related with another one in the learning phase. Whether associative evaluative learning does or does not require this ability has been debated for long time, with the purpose to understand what mental mechanism can account for EC effects. Several studies have reported EC even when participants were not aware of the CS–US contingencies (e.g., Baeyens, Eelen, & Van den Bergh, 1990; Dickinson & Brown, 2007; Walther & Nagengast, 2006). Other studies, however, have indicated that EC occurs only after the participants become aware of the contingency between the CS and the US with which it was paired (e.g., Pleyers, Corneille, Luminet, & Yzerbyt, 2007; Stahl, Unkelbach, & Corneille, 2009). A meta-analysis on EC (Hofmann et al., 2010) clarified that EC requires contingency awareness. In meta-analyzing the SR effect on implicit attitude, explicit attitude, and identification, we provided compelling evidence of the crucial role of IR memory in qualifying the effect of the task on all the outcome measures. Taken together, these meta-analytical results seem in line with a propositional interpretation of evaluative learning via the associative route (De Houwer, 2007, 2009, 2014). According to this account, people need to form a conscious proposition about stimuli relation before this relation can affect liking. Hence, as in EC a proposition that defines the relation between stimuli arise from their spatio-temporal contingency, so in the SR the formation of a relational proposition requires the conscious knowledge that the self and one target shared the same

behavioral response. Nonetheless, while IR memory matters, our meta-analyses showed that aspects related to the statistical and procedural properties of the task do not affect the SR effect. For instance, the task was effective regardless of the number of trials involved in the task. Again, this finding seems to support the importance of a propositional reasoning that allows participants to relate objects in a certain way, compared to the automaticity involved in the act of repeating an action for multiple times.

If noticing and learning that two stimuli share something (e.g., a behavioral response) is crucial to infer that they are related, it is less intuitive why people should transfer the properties carried by one of the stimuli to the other one. Related to this aspect, De Houwer and Hughes (2016) recently proposed the idea of a symbolic meaning of evaluative learning[22]. The underlying idea is that once people have learned to use symbols, they can apply this ability to any kind of proximal stimuli. The fact that humans use symbols in such a flexible and ubiquitous manner has led to the proposal that humans are essentially symbolic beings (Deacon, 1997). If a symbol is meant as something that stands for something else, then any proximal event can function as a symbol. Therefore, even a range of proximal events can function as symbols on the basis of which meaning is constructed. In EC, for instance, the proximal stimuli consist in the repeated pairings between conditioned and unconditioned stimuli. In the SR task, the proximal stimuli are in the act of performing the same actions (e.g., pressing a common key) to categorize two pairs of stimuli categories. With their idea of symbolic meaning construction upon proximal stimuli, De Houwer and Hughes (2016) claimed that the individuals tend to respond to these proximal stimuli based on how they defined their environment. At an abstract level of analysis, performing a common action in response to two classes of stimuli (e.g., the self and a target object) could influence liking because it is interpreted as a symbol of the

---

[22] Although the authors presented their idea of symbolic meaning in the context of evaluative conditioning, we believe that this symbolic interpretation can be extended to any form of associative learning effect.

relation between the stimuli. In fact, throughout their experience and exposure to life events, people might have learned that when two stimuli share something, they are also similar. For instance, people tend to infer that two persons who own a Lamborghini are both rich and therefore similar, even without knowing anything about their lives. Similarly, the SR effect should result from humans' ability to treat a target object as similar to the self because of the action they have been categorized with throughout the task. Therefore, the stimulus categorized through the same action as the self should be considered as more similar to the self than the alternative target object. This idea of symbolic meaning construction can be tested in many different ways. In a recent research, Mattavelli, Hughes, Perugini, Richetin, and De Houwer (2016) conducted two studies in which they tested the effect of an evaluative learning paradigm based on intersecting regularities after manipulating the distal meaning attributed to the act of pressing the same action. The task was structurally identical to the SR task, except for the fact that distinct positive stimuli (rather than the self) were used as main source and contrasted by negative stimuli. Crucially, this task was preceded by another task in which people learned to press the same key for words that were either similar or opposed one another. The magnitude of the effect was qualified by the task preceding the evaluative learning manipulation, such that the effect was stronger for participants who previously learned that pressing the same key connected stimuli of similar meaning. These findings support the idea that IR represents a proximal cue that people use to construct relations between stimuli and that the type of relation is crucial in determining the transfer of liking from the source to the target. Future studies should further extend these results to the specific case of the SR task, and might also include different types of manipulation to test the symbolic meaning of the SR effect. One possibility is to see whether intrinsic similarity relations between the self and a target qualifies the SR effect. For instance, the magnitude and perhaps the direction of the SR effect should depend on whether the target object that is categorized through the same action as the self are actually similar or different (e.g., friends or enemies).

In essence, our meta-analysis showed that the memory of the intersection between the self and a target object is crucial in qualifying the SR effect on all the outcomes considered. However, this

does not tell anything about what people infer from this learned intersecting regularities and what kind of relationships is inferred between stimuli. For instance, it is plausible that the same behavioral commonality between the self and one target acquires distinct meaning because of the nature of the target stimulus involved. Thus, after pressing the same key to categorize the self and a brand, people might infer that they own the brand; conversely, categorizing the self with one group might result in the feeling of being part of that group. Discovering whether different stimuli lead to different relationship with the self and how this affect the SR effect is important as it provides indication about learning via intersecting regularities and, specifically for the case of the SR task, about the impact of different types of self-target relationships in determining a transfer of liking.

**Self-driven indirect SR effect**

The SR effect can be conceptualized as an indirect form of associative attitude change. A first and simple reason is in the fact that the SR effect occurs without implying any direct relation between source and target stimuli. Thus, people end up liking a target object when they learn that there is a third element (e.g., an action) that establishes an indirect relation between that target and an attitudinal source (i.e., the self). In chapter 2, the idea of the SR as an indirect way to change attitudes was combined with the other important finding outlined in the meta-analysis, that is, the role of the type of stimuli used to represent the self-concept in the task. Specifically, the effect of the SR showed significantly stronger when idiographic or self-relevant stimuli (e.g., personality traits, individuals' names) were used to represent the self. Participants preferred the brand products paired with self-relevant personality traits when they have the same valence of those selected as not relevant for the self (used in the SR task to represent the category 'others') and, more important, when the former are more negative than the latter. These findings showed that stimuli relevance with respect to the self-concept, rather than stimuli own valence, drives the SR effect on attitude and identification.

If learning via IR represents an indirect pathway that leads to attitude change, the effects observed in chapter 2 strengthen this idea at a higher order level. Let us consider for instance study 4, in which moderately negative self-relevant traits were paired with one brand and moderately

positive self-irrelevant traits were paired with another brand. From a purely associative perspective of learning via IR, the brand paired with the moderately positive stimuli should be preferred over the other one. Instead, adopting a propositional approach, the findings support the idea of an indirect effect of the positivity carried by self (Yamaguchi et al., 2007) in qualifying source-target relations. Thus, the relevance of the IR principle is demonstrated at two distinct levels. At the concrete level, through the same action performed in the categorization task people learn that two stimuli relate each other. At the abstract level, the direction (and the magnitude) of the effect seems to be determined by one's ability to use the personality traits as intersecting regularities between the self and one target object. For instance, consider the adjective *'lazy'* as an example of moderately negative trait chosen by an individual. Once paired with a brand X, people might form a first regularity *"The brand X is for lazy people"*. This newly created relational proposition might be in turn accompanied by a second proposition, *"I am lazy"*. The resulting conclusion of this hypothesized syllogistic reasoning is *"The brand X is like me"*. Under this view, laziness is the intersection between the two regularities concerning the self and the brand. Thus, intersections are important at the associative level to relate stimuli throughout the task and they play a crucial role at the propositional level to qualify the nature of this learned relation.

All this reasoning about the IR principle conceptualized in its abstract form paves the way for a number of empirical questions that remain unanswered at present. For instance, from chapter 2 we learned that self-relevance is more important than valence for the SR effect to occur. Moreover, in their first SR paper, Prestwich et al. (2010) demonstrated that the effect of the SR on implicit attitude change is moderated by one's level of implicit self-esteem. Thus, at least implicitly, the SR is less likely to occur for those participants who possess a 'negative self'. It has been demonstrated that individuals with high self-esteem are most apt to display favoritism when they are directly involved in group processes, whereas low self-esteem subjects are most apt to display favoritism when they are not directly involved in group processes (Brown, Collins, & Schmidt, 1988). Combining these findings together could be of great interest to test the relations between the SR effect, the way in

which the self is depicted, and individual differences in terms of one's self view. In particular, it seems plausible that people might express a certain degree of favoritism for objects linked to the general self-concept (represented by self-related pronouns) when they have a positive view of themselves, while this might not be the case for people with negative view of themselves: for people with low self-esteem a relationship between the general self and the target object might result in a transfer of negative valence, given that negativity is easily accessible when they think about their person. Hence, establishing a less direct relation between one's general self and the object might allow people to simply infer that the object has something in common with him or her, without making the self-related negativity accessible. Conversely, less direct self-object links (e.g., pairing the attitudinal object with a specific feature possessed by the self) could result in perhaps positive idea that the object is somehow similar to them in a specific feature.

In essence, showing that self-relevance is more important than valence in driving the SR effect opens up the possibility that different ways to depict the self might result in different relationships between the individual and the target object and that the nature of the resulting relationship inferred may depend on inter-individual aspects. Future studies should focus on how individual characteristics that affect one's self view impact the SR effect.

**Procedural indirect SR and the role of target congruency**

Individual differences, especially related with self-esteem, can be also related with the findings presented in chapter 3. Specifically, chapter 3 tested indirect SR from the perspective of its procedural mechanism. Five studies demonstrated that the positivity of the self can transfer towards objects that never relate with the self in the SR task. This effect occurred thanks to intersecting regularities across distinct learning phases. In fact, we showed that an object paired with the self in the SR task and then with a focal target in an additional learning task worked as commonality between the self and the focal target, therefore leading to implicit and explicit attitude change towards the latter. Hence, at the procedural level, we demonstrated that intersecting regularities can be operationalized in complex ways to indirectly affect stimuli liking. In some sense, this form of indirect

SR is an attempt to spread the positivity carried by the attitudinal source across multiple links of stimuli chain. While we know that spreading this positivity is possible on target objects that are linked to the self by means of another stimulus, we do not know whether this transfer is affected by the level of positivity of the attitudinal source. Also, none of the previous studies that focused on indirect attitude change in the context of associative evaluative learning have focused on the role played by the source positivity in qualifying the spreading of liking across stimuli (Walther, 2002; Gast & De Houwer, 2012). By relying on the self as an attitudinal source, in the case of the SR the positivity attributed to the source corresponds to one's level of self-esteem. If one's level of implicit self-esteem moderates the SR effect on implicit attitudes (Prestwich et al., 2010), it is plausible that it can also determine the magnitude of an indirect SR effect. Thus, we hypothesize that the higher one's self-esteem, the stronger the indirect SR effect on target objects never related with the self. In essence, when the SR effect is tested on individuals with a highly positive self, liking will less likely fade away in transferring from multiple target objects.

In discussing about individuals' ability to infer a relationship between stimuli that share a commonality, we claimed that the specific characteristics of the stimuli can impact the type of relation inferred between them and that this might impact the subsequent transfer of liking. Focusing on the role of the target objects' category, chapter 3 also showed that the indirect SR effect across multiple target objects was qualified by congruency in targets domain. Liking transfers more easily across target objects of the same, rather than different, domain. This additional demonstration of indirect SR effect could be more easily explained by adopting a propositional approach to evaluative learning (De Houwer, 2009; 2014). In fact, not only participants need to remember the correct IR at the end of the two learning phases. They also have to qualify the acquired relationship between stimuli in a relational way. Such findings seems also in line with the conceptualization of associative evaluative learning proposed by De Houwer and Hughes (2016). In line with this idea, in the indirect SR tested in chapter 3, the transfer of valence might be determined by the extent to which this relationship is meaningful or not. Thus, the congruency in stimuli domain might serve as a subtle relational cue that

participants use to validate the learned relationship between stimuli. Related to this relational issue is the other relevant finding that emerged from chapter 3. We showed that, when the indirect SR was tested with target objects belonging to different domains, people's ability to qualify the relation between stimuli was a crucial moderator of the SR effect. The role of stimuli congruency in affecting the SR is important as it shows that the construction of a relation between stimuli is qualified by the meaningfulness implied in that relationship. This meaningfulness varies as a function of the intrinsic similarity between stimuli involved: the higher the similarity, the higher the likelihood to observe an indirect SR effect.

**Addressing the boundaries of the SR task**

      **Construct validity and attitude/behavior change**

      After focusing on the two main characteristics of the SR task – the self and the IR principle – chapters 4 and 5 addressed two potential limitations that emerged from meta-analytical findings. First, it was shown that the SR effect is stronger on the IAT (Greenwald, McGheee, & Schwartz, 1998) than on the AMP (Payne, Cheng, Govorun, & Stewart, 2005). This finding might raise doubts on the construct validity of the SR task. The structure of the SR task in fact mirrors one block of an IAT. Both tasks have substantial shared-method variance. Thus, the effect of the SR might derive from the fact that participants extensively practice the mapping of one critical IAT block throughout the SR task. The 'shared variance' issue calls into question whether the SR effect observe on implicit attitude change is actually a change in attitude or, instead, an artifact due to the high overlap between the manipulation and the measure. This issue was already addressed in the discussion of the meta-analysis (Chapter 1). We would like to further note that another way to deal with the threat of construct validity would be to provide theory-consistent relations of the dependent variables (IAT) with outside criteria. In the first study of Richetin, Mattavelli, and Perugini (2016) presented in chapter 4, we provided such a type of demonstration of the construct validity of the SR effect. After completing a SR task in which participants paired two alternative fictitious eco-brands with either the self or the category others, they completed an IAT, a semantic differential, and a shopping choice task. The effect of the

SR task was significant on implicit and explicit attitudes as well as on behavioral choice. The three outcome variables were also correlated with each other. Importantly, the effect of the SR on the IAT fully mediated the effect of the SR on explicit attitude (single mediation) and the effect of the SR on the IAT and explicit attitude in series fully mediated the impact of the SR manipulation on one's behavioral choice (serial mediation). Therefore, the genuineness of the SR effect on implicit attitude is corroborated by the fact that such a change on implicit attitude explains a subsequent change on explicit attitude and on a behavioral proxy measure also affected by the SR manipulation. Nevertheless, future SR studies should be aimed at testing its effectiveness in changing implicit attitudes assessed using measures other than the IAT.

**Attitude formation versus attitude change. The interplay between SR and persuasion**

In attitude research, it is important discriminating between the concept of forming or changing an attitude. Thus, a manipulation can contribute to the generation of attitudes that did not exist before or can alter an existing attitude. Related to this point, a relevant boundary condition of the SR effect that emerged from the meta-analysis is the stronger impact of the SR on fictitious, rather than existing, target objects. Therefore, it seems that the SR is highly effective in forming attitudes, while changing pre-existing attitudes is more complex. This may raise doubts about the potential application of the SR effect in the real world. Chapter 5 described a study in which we tested a) the impact of the SR in changing pre-existing negative (or moderately positive) attitudes towards green vegetables and b) the comparison between the SR manipulation and a persuasive communication. By measuring implicit and explicit attitudes and the readiness to change behavior in favor of the targeted stimuli (e.g., eating more green vegetables in future) we showed that the SR and the persuasive communication uniquely contributed in changing implicit and explicit attitudes (respectively). In essence, this study demonstrated the SR can indeed change, not merely form, attitudes, as shown by the fact that the effect on implicit attitudes was tested on a sample of participants who did not like green vegetables.

Moreover, this study showed that pairing green vegetables with the self in the SR task lead to stronger intention to consume green vegetables in future, but only for participants who were not

exposed to the persuasive communication. Taken together, these results pave the way for further investigations on the differential impact of either procedurally associative or propositional manipulations in determining implicit, explicit attitudes, and behavioral intentions when these manipulations are tested in interaction. Associative attitude change and persuasion have been traditionally considered as independent phenomena based on different underlying mental processes. It is generally assumed that whereas attitude change via persuasive messages requires effortful processing of words and sentences, associative attitude change (e.g., EC) is *"a primitive means of changing attitudes"* (Briñol, Petty, & McCaslin, 2009, p. 287) requires little cognitive effort and thought. To the best of our knowledge, the interplay between associative and persuasive manipulations in affecting simultaneously the same target object has not been explored in attitude literature. In trying to provide an interpretation of the findings just mentioned, we could speculate that an associative learning procedure based on the self is more effective when administered in isolation because, from the individuals' perspective, being also exposed to a persuasive message might be seen as a non-genuine and forced attempt to change people's behavior. If this hypothesis is correct, then the lower effectiveness of the double manipulation should be mediated by participants' reactance. In fact, people may change their attitudes or intention simply because they are motivated to restore their freedom, and disagreement is the most direct way to do so (Brehm & Brehm, 1981). That is, people might declare they are not willing to consume green vegetables because they feel coerced into doing so.

But these findings have direct implications also from a practical perspective. Let us consider for instance the case of everyone's favorite hazelnut-chocolate spread: Nutella. In 2013, Nutella came out with a marketing campaign based on the idea to give customers the opportunity to see their own names on its jars. Although operationalized in a different way, this can be considered as a self-referencing strategy: by seeing their own names printed with the same colors and font in which the brand name is usually depicted, people should end-up liking the brands as they think that it has something in common with them. A couple of years later, Ségolène Royal, France's ecology minister,

claimed that people should stop eating Nutella as it is made with palm oil, an environmentally destructive but commonly used ingredient. Nowadays, by surfing on Nutella's website one could find pictures of jars with customers' names printed on, as well as information about its ingredients, among which palm oil is taken in serious consideration. Based on what observed in the study described in chapter 5, this double way to build a marketing campaign might ironically reveal as counter-productive if the purpose of the brand is to persuade people who do not like Nutella to start consuming it. Relating the brand with the self (via the personalized jars) should develop a positive implicit attitude toward it. Conversely, reassuring customers about the way in which palm oil is used and the environmental implications related to it should increase attitude towards the brand at the explicit level. However, the combination of the two interventions might be less effective than the first one alone in directing one's intention to consume Nutella in future time. We acknowledge that the parallel between our findings and this real life example might sound as rather speculative. As a matter of fact, our aim is to stimulate the interest in investigating deeper this interplay between associative models of evaluative learning and persuasion, with a special focus on the role played by distinct moderating variables: for instance, whether the manipulations are meant to either change or form an attitude might matter; also the order in which persuasive information and associative learning procedures are presented might impact their interplay. In essence, we believe that the analysis of this interplay could represent an intriguing challenge for all researchers interested in distinct ways through which attitude and behavior can be changed.

**The SR effect beyond preferences**

The lines of research presented across this dissertation tested the idea of changing attitudes, identification, and behavioral intentions towards target objects by exploiting the positivity carried by the self-concept. However, the self is not only characterized by its positive valence. Rather, it is a central psychological structure (Greenwald & Pratkanis, 1984) that it is chronically active and therefore accessible (Markus & Kunda, 1986) and has a motivational relevance. The two studies presented in chapter 6 are based on the idea that information that involve the self tend to be retained

longer and are more likely to lead to permanent attitude change (Petty, Cacioppo, Strathman, & Priester, 1994). Plus, threatening messages tailored on the individual by means of visual stimuli has been investigated in the context of smoking behavior (e.g., Flett, Grogan, Clark-Carter, Gough, & Conner, 2015). Similarly, we created a modified SR task, where self-threatening visual stimuli were used to change attitudes towards health related goods (i.e., soft drinks). This approach is not only innovative from the perspective of the SR task; as far as we are aware, no studies on associative learning have used self-relevant information to induce attitudinal change. If study 1 offered convincing results about the effectiveness of this Fat Self Referencing task in changing implicit and explicit attitudes, study 2 went even further in showing that this effect can also have an impact on extra-evaluative dimensions. It showed that implicitly and (only in part) explicitly, the drink paired with the fat version of one's self was perceived as fatter and less healthy. Past research on attribute conditioning has shown that people ascribe certain attributes to objects because of their co-occurrence with stimuli possessing the attribute at issue (Kim, Allen, & Kardes, 1996; Förderer & Unkelbach, 2011; 2014; 2015; Glaser & Walther, 2013). The second study presented in chapter 6 suggests that this attribution of properties might be due to some sort of inferred causal relation between one's fat face and a health related good (e.g., *"My face gets fatter when I drink the soft drink X"*).

In more general terms, this study is also inspiring as it offers a nice and effective example of using the centrality of the self to transfer properties other than valence. In the study summarized above, we demonstrated that the SR task is effective on an extra-evaluative domain. These findings add to the previously demonstrated SR effect on identification (i.e., Richetin, Mattavelli, & Perugini, 2016). In addition to this, in a series of four studies not mentioned in the present dissertation, Richetin, Perugini and Mattavelli (2016) showed that objects characterized through the same action as the self acquire higher cognitive accessibility. Taken together, these findings offer the ideal basis for further exploiting the special status of the self to broaden the effect of the SR task beyond the evaluative domain.

**Conclusions**

Throughout the pages of the present dissertation, we developed the novel principle of intersecting regularities (Hughes et al., 2016) in associative evaluative learning. In a research context that has been largely dominated by a unique way of linking stimuli (i.e., through repeated pairings) the IR mechanism is a potential new pathway that has already shown effectiveness in modifying implicit and explicit attitudes by establishing commonalities between valenced stimuli and attitudinal target objects. We applied this learning mechanism through the SR task (Prestwich et al., 2010). We showed the effectiveness of this paradigm in changing both implicit and explicit attitudes, as well as identification, towards distinct target objects and we detailed the most important boundaries that characterize such an effect, among which, one pertains to the role played by IR memory in moderating the SR effect. In the empirical chapters, we first centered our attention on the self and examined which of its facets best explains the SR effect and then moved to the possibility to exploiting the IR mechanism to create (indirect) SR effect on multiple attitude stimuli belonging to either the same or different class. The findings from these lines of research are important from a theoretical perspective: in both cases, the SR effect on implicit and explicit attitude seemed to rely on the ability to relate stimuli in a propositional manner (De Houwer, 2009, 2014). This idea, already raised by the impact of IR memory observed in the meta-analysis, was further supported by the facts that the effect of the SR a) was driven by the intrinsic self-relevance, rather than the valence, of the source stimuli, b) it could spread on stimuli that are not presented in the SR task, and c) that this spreading was qualified by the similarity between stimuli as well as by individuals' ability to infer that stimuli directly or indirectly related were somehow equivalent. We also provided compelling support to the ecological validity of the SR task. We showed its effectiveness in both forming and changing attitudes in the health-domain. Besides, we proved the strength of the SR effect as well as its impact on behavioral intentions. Furthermore, we could not ignore the fact that the self, as a complex and central psychological structure (Greenwald & Pratkanis, 1984), is not only characterized by a positive valence. We showed that by simply tailoring the SR task in a way that made health-threatening messages as relevant for the self, negatively affected one's attitudes towards aversive target food

objects. Finally, in line with the idea that the self is not only positive, this research also provided initial evidence that the SR task can also be used to alter other psychological properties.

On the one hand, these distinct lines of research contributed in improving our knowledge about the many features of associative learning via the self through the IR mechanism. On the other hand, we believe that these findings should be considered as an initial, though robust, evidence of an evaluative learning effect (i.e., the SR effect) and of a learning mechanism (i.e., intersecting regularities) that are still relatively novel and unexplored. Therefore, the primary goal of this dissertation is to stimulate the research interest upon the present findings and to pave the way for potential new investigations.

**References**

Aertsens, J., Verbeke, W., Mondelaers, K., & Van Huylenbroeck, G. (2009). Personal determinants of organic food consumption: A review. *British Food Journal*, 111(10), 1140-1167.

Ajzen, I. (1991). The theory of planned behavior. *Organization Behavior and Human Decision Process, 50*(2), 179-211.

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology*, *49*(6), 1621-1630.

Allison, S. T., Messick, D. M., & Goethals, G. R. (1989). On being better but not smarter than others: The Muhammad Ali effect. *Social Cognition*, *7*(3), 275-295.

Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *Handbook of Social Psychology*. Worchester, MA: Clark University Press.

Alós-Ferrer, C., & Strack, F. (2014). From dual processes to multiple selves: Implications for economic behavior. *Journal of Economic Psychology*, *41*, 1–11.

Andreyeva, T., Chaloupka, F. J., & Brownell, K. D. (2011). Estimating the potential of taxes on sugar-sweetened beverages to reduce consumption and generate revenue. *Preventive Medicine*, *52*(6), 413-416.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*(4), 596-612.

Baeyens, F., Eelen, P., & Van Den Bergh, O. (1990). Contingency awareness in evaluative conditioning: A case for unaware affective-evaluative learning. *Cognition & Emotion*, *4*(1), 3-18.

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.

Bandura, A. (1998). Health promotion from the perspective of social cognitive theory. *Psychology & Health*, *13*(4), 623-649.

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods, 46*(3), 668-688.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3):255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version, 1(7).

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effect s and moderated mediation in multilevel models: new procedures and recommendations. *Psychological Methods, 11*, 142–163.

Beggan, J. K. (1992). On the social nature of nonsocial perception: The mere ownership effect. *Journal of Personality and Social Psychology*, *62*(2), 229-237.

Bhattacharya, C. B., & Sen, S. (2003). Consumer-company identification: A framework for understanding consumers' relationships with companies. *Journal of Marketing*, *67*(2), 76-88.

Bluemke, M., & Friese, M. (2006). Do features of stimuli influence IAT effects? *Journal of Experimental Social Psychology, 42*, 163-176.

Bluemke, M., & Friese, M. (2012). On the validity of idiographic and generic self-concept Implicit Association Tests: A core-concept model. *European Journal of Personality, 26*(5), 515-528.

Bodenhausen, G. V, & Gawronski, B. (2013). Attitude Change. In D. Reisberg (Ed.), *The Oxford Handbook of Cognitive Psychology* (pp. 957–969). New York, NY: Oxford University Press.

Bohner, G., & Dickel, N. (2011). Attitudes and attitude change. *Annual Review of Psychology*, 62, 391–417.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis version 2*. Englewood, NJ: Biostat.

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.

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of personality and Social Psychology*, *47*(6), 1191-1205.

Brehm, S. S., & Brehm, J. W. (1981). Psychological reactance: A theory of freedom and control. San Diego, CA: Academic.

Brewer, M. B., & Gardner, W. (1996). Who is this" We"? Levels of collective identity and self representations. *Journal of Personality and Social Psychology, 71*, 83-93.

Brown, J. D., Collins, R. L., & Schmidt, G. W. (1988). Self-esteem and direct versus indirect forms of self-enhancement. *Journal of Personality and Social Psychology*, *55*(3), 445-453.

Brug, J., Steenhuis, I., van Assema, P., & de Vries, H. (1996). The impact of a computer-tailored nutrition intervention. *Preventive Medicine*, *25*(3), 236-242.

Brunel, F. F., Tietje, B. C., & Greenwald, A. G. (2004). Is the implicit association test a valid and valuable measure of implicit consumer social cognition? *Journal of Consumer Psychology*, *14*(4), 385.

Burnkrant, R. E., & Unnava, H. R. (1995). Effects of self-referencing on persuasion. *Journal of Consumer Research*, *22*(1), 17-26.

Campbell, M. K., DeVellis, B. M., Strecher, V. J., Ammerman, A. S., DeVellis, R. F., & Sandler, R. S. (1994). Improving dietary behavior: the effectiveness of tailored messages in primary care settings. *American Journal of Public Health*, *84*(5), 783-787.

Caprara, G. V., & Perugini, M. (1991). L'approccio psicolessicale e l'emergenza dei Big Five nello studio della personalita'(The psycholexical approach and the emergency of the Big Five in the study of personality). *Giornale Italiano di Psicologia*, *18*, 721-747.

Carver, C. S., & Scheier, M. F. (1981). The self-attention-induced feedback loop and social facilitation. *Journal of Experimental Social Psychology, 17*, 545-568.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. *Dual-Process Theories in Social Psychology*, *15*, 73-96.

Cohen, D., & Babey, S. H. (2012). Contextual influences on eating behaviours: heuristic processing and dietary choices. *Obesity Reviews : An Official Journal of the International Association for the Study of Obesity*, *13*(9), 766–79.

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 & Social Psychology Bulletin*, *33*(12), 1727–1740.

Crawford, M. T., Sherman, S. J., & Hamilton, D. L. (2002). Perceived entitativity, stereotype formation, and the interchangeability of group members. *Journal of Personality and Social Psychology*, *83*(5), 1076-1094.

De Hart, T., Pelham, B. W., & Tennen, H. (2006). What lies beneath: Parenting style and implicit self-esteem. *Journal of Experimental Social Psychology*, *42*(1), 1-17.

De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental psychology*, *50*(2), 77-85.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, *10*(02), 230-241.

De Houwer, J. (2009a). Leerpsychologie [Learning Psychology]. Gent, Belgium: Academia Press.

De Houwer, J. (2009b). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior, 37*(1), 1-20.

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

De Houwer, J., & Hughes, S. (2016). Evaluative conditioning as a symbolic phenomenon: On the relation between evaluative conditioning, evaluative conditioning via instructions, and persuasion. *Social Cognition*.

De Houwer, J., Baeyens, F., & Eelen, P. (1994). Verbal evaluative conditioning with undetected US presentations. *Behaviour Research and Therapy*, *32*(6), 629-633.

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

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology*, *24*(1), 252-287.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*(6), 853-869.

Deacon, T. W. (1997). *The Symbolic Species: The Co-evolution of Language and the Brain*. New York: Norton.

Dholakia, U. M., & Morwitz, V. G. (2002). The Scope and Persistence of Mere-Measurement Effects: Evidence from a Field Study of Customer Satisfaction Measurement. *Journal of Consumer Research*, *29*(2), 159-167.

Di Noia, J. (2014). Defining Powerhouse Fruits and Vegetables: A Nutrient Density Approach. *Preventing Chronic Disease*. 11: E95.

Dickinson, A., & Brown, K. J. (2007). Flavor-evaluative conditioning is unaffected by contingency knowledge during training with color-flavor compounds. *Animal Learning & Behavior*, *35*(1), 36-42.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*(1), 62-68.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology*, *57*(6), 1082-1090.

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, *23*(3), 316-326.

Duval, S., & Wicklund, R. A. (1973). Effects of objective self-awareness on attribution of causality. *Journal of Experimental Social Psychology, 9*, 17-31.

Dwan, K., Gamble, C., Williamson, P. R., & Kirkham, J. J. (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias—an updated review. *PloS One, 8*(7).

Dymond, S., & Whelan, R. (2010). Derived relational responding: A comparison of Match-To-Sample and the Relational Completion Procedure. *Journal of the Experimental Analysis of Behavior, 94*(1), 37-55.

Eagly, A. H., & Chaiken, S. (1993). The psychology of attitudes. Fort Worth, TX: Harcourt Brace Jovanovich.

Eagly, A. H., & Chaiken, S. (2007). The advantages of an inclusive definition of attitude. *Social Cognition*, *25*(5), 582-602.

Ebert, I. D., Steffens, M. C., Von Stülpnagel, R., & Jelenec, P. (2009). How to like yourself better, or chocolate less: Changing implicit attitudes with one IAT task. *Journal of Experimental Social Psychology, 45*(5), 1098-1104.

Ellemers, N. (1993). The influence of socio-structural variables on identity management strategies. *European Review of Social Psychology, 4*, 27-57.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*(8), 709–724.

Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278.

Farnham, S. D. (1999). From implicit self-esteem to in-group favoritism. Unpublished doctoral dissertation, University of Washington.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 23, pp. 75-109). New York: Academic Press.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, *25*(5), 603-637.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*(1), 297-327.

Fazio, R. H., & Towles-Schwen, T. (1999). The MODE model of attitude-behavior process. In *Dual-Process Theories in Social Psychology*, 97–116.

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

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline?. *Journal of Personality and Social Psychology*, *69*(6), 1013-1027.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117-140.

Festinger, L. (1957). Cognitive dissonance theory. *1989) Primary Prevention of HIV/AIDS: Psychological Approaches. Newbury Park, California, Sage Publications*.

Fiedler, K., & Unkelbach, C. (2011). Evaluative conditioning depends on higher order encoding processes. *Cognition & Emotion, 25*(4), 639-656.

Flett, K., Grogan, S., Clark-Carter, D., Gough, B., & Conner, M. (2015). Male smokers' experiences of an appearance-focused facial-ageing intervention. *Journal of Health Psychology*.

Förderer, S., & Unkelbach, C. (2011). Beyond evaluative conditioning! Evidence for transfer of non-evaluative attributes. *Social Psychological and Personality Science*, *2,* 479-486.

Förderer, S., & Unkelbach, C. (2012). Hating the cute kitten or loving the aggressive pit-bull: EC effects depend on CS–US relations. *Cognition & Emotion, 26*(3), 534-540.

Förderer, S., & Unkelbach, C. (2014). The moderating role of attribute accessibility in conditioning multiple specific attributes. *European Journal of Social Psychology*, *44*(1), 69-81.

Förderer, S., & Unkelbach, C. (2014). The moderating role of attribute accessibility in conditioning multiple specific attributes. *European Journal of Social Psychology*, *44*(1), 69-81.

Förderer, S., & Unkelbach, C. (2015). Attribute conditioning: Changing attribute-assessments through mere pairings. *The Quarterly Journal of Experimental Psychology*, *68*(1), 144-164.

Friese, M., Hofmann, W., & Schmitt, M. (2009). When and why do implicit measures predict behaviour? Empirical evidence for the moderating role of opportunity, motivation, and process reliance. *European Review of Social Psychology*, *19*(1), 285-338.

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. *The British Journal of Social Psychology*, *47*(3), 397–419.

Friese, M., Hofmann, W., & Wänke, M. (2009). The impulsive consumer: Predicting consumer behavior with implicit reaction time measures. *Frontiers in social psychology: Social Psychology of Consumer Behavior. New York*, 335-364.

Fudenberg, D., & Levine, D. K. (2006). A dual-self model of impulse control. *The American Economic Review*, *96*(5), 1449–1476.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology, 78*, 708-724.

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.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692-731.

Gawronski, B., & Bodenhausen, G. V. (2011). 2 The Associative-Propositional Evaluation Model: Theory, Evidence, and Open Questions. *Advances in Experimental Social Psychology*, *44*, 59-127.

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.

Gawronski, B., & Strack, F. (2004). On the propositional nature of cognitive consistency: Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology*, *40*(4), 535-542.

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.

Gawronski, B., Walther, E., & Blank, H. (2005). Cognitive consistency and the formation of interpersonal attitudes: Cognitive balance affects the encoding of social information. *Journal of Experimental Social Psychology, 41*(6), 618-626.

Giner-Sorolla, R. (2012). Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspectives on Psychological Science, 7*, 562-571.

Glaser, T., & Walther, E. (2013). Two in one conditioning? The role of valence in concept learning. *Learning & Motivation*, *44*(3), 174-183.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4-27.

Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*(6), 1022-1038.

Greenwald, A. G., & Pratkanis, A. R. (1984). The self. In R. S. Wyer & TK Srull. *Handbook of Social Cognition*, *1*, 129-178.

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.

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.

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.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*(1), 17-41.

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, *90*(1), 1-20.

Grunert, S. C., & Juhl, H. J. (1995). Values, environmental attitudes, and buying of organic foods. *Journal of Economic Psychology*, *16*(1), 39-62.

Hammerl, M., & Grabitz, H. J. (1996). Human evaluative conditioning without experiencing a valued event. *Learning & Motivation*, *27*(3), 278-293.

Harris, P., Brearley, I., Sheeran, P., et al. (2014). Combining self-affirmation with implementation intentions to promote fruit and vegetable consumption. *Health Psychology*. *33*(7), 729-736.

Hayes S.C, Barnes D. (1997). Analyzing derived stimulus relations requires more than the concept of stimulus class. *Journal of the Experimental Analysis of Behavior, 68*, 235–244.

Hayes, A. F. (2012). PROCESS: *A versatile computational tool for observed variable mediation, moderation, and conditional process modeling* [white paper]. Retrieved from http://www.afhayes.com/ public/process2012.pdf

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. Springer Science & Business Media.

Haynes, A., Kemps, E., & Moffitt, R. (2015). Inhibitory self-control moderates the effect of changed implicit food evaluations on snack food consumption. *Appetite*, *90*, 114-122.

He, H., Li, Y., & Harris, L. (2012). Social identity perspective on brand loyalty. *Journal of Business Research*, *65*(5), 648–657.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics, 6*(2), 107-128.

Heider, F. (1958). The psychology of interpersonal relations. New York: Wiley.

Higdon, J.V., Delage, B., Williams, D.E., & Dashwood, R.H. (2007). Cruciferous vegetables and Human Cancer Risk: Epidemiologic Evidence and Mechanistic Basis. *Pharmacological Research, 55*(3), 224-236.

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.

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.

Hoorens, V., & Nuttin, J. M. (1993). Overvaluation of own attributes: Mere ownership or subjective frequency? *Social Cognition*, *11*(2), 177-200.

Houben, K., Havermans, R., & Wiers, R. (2010). Learning to dislike alcohol: Conditioning negative implicit attitudes toward alcohol and its effect on drinking behavior. *Psychopharmacology*, *211*(1), 79-86.

Hu, X., Gawronski, B., & Balas, R. (2016). Propositional Versus Dual-Process Accounts of Evaluative Conditioning: I. The Effects of Co-Occurrence and Relational Information on Implicit and Explicit Evaluations. *Personality and Social Psychology Bulletin*.

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

Hughes, S., De Houwer, J., & 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.

Ioannidis, J. P. (2008). Why most discovered true associations are inflated. *Epidemiology, 19*, 640-648.

Kahneman, D. (2011). *Thinking, fast and slow*, New York: Farrar, Straus and Giroux.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowmend effect and the coase theorem. *The Journal of Political Economy*, *98*(6), 1325-1348.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology*, *81*(5), 774–788.

Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91*(1), 16-32.

Kim, J., Allen, C. T., & Kardes, F. R. (1996). An investigation of the mediational mechanisms underlying attitudinal conditioning. *Journal of Marketing Research*, 318-328.

Koole, S. L., Dijksterhuis, A., & van Knippenberg, A. (2001). What's in a name: implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology*, *80*(4), 669-685.

Krieglmeyer, R., Deutsch, R., De Houwer, J., & De Raedt, R. (2010). Being moved valence activates approach-avoidance behavior independently of evaluation and approach-avoidance intentions. *Psychological Science*, *21*(4), 607-613.

Krosnick, J., & Petty, R. (1995) Attitude strength: An overview. In: Petty R, Krosnick J, eds. *Attitude strength: Antecedents and consequences*. Hillsdale, NJ, England: Lawrence Erlbaum Associates. 1-24.

Kuenzel, S., & Halliday, S. V. (2008). Investigating antecedents and consequences of brand identification. *Journal of Product & Brand Management*, *17*(5), 293-304.

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y. & Sloman, S. A. (2005) Beyond covariation: Cues to causal structure. In: Causal learning: Psychology, philosophy and computation, ed. A. Gopnik & L. Schultz, pp. 154–72. Oxford University Press.

Lakens, D., Hilgard, J., & Staaks, J. (2016). On the reproducibility of meta-analyses: Six practical recommendations. *BMC psychology, 4*, 1-10.

Lascelles, K. R., Field, A. P., & Davey, G. C. (2003). Using foods as CSs and body shapes as UCSs: A putative role for associative learning in the development of eating disorders. *Behavior Therapy*, *34*(2), 213-235.

Lebens, H., Roefs, A., Martijn, C., Houben, K., Nederkoorn, C., & Jansen, A. (2011). Making implicit measures of associations with snack foods more negative through evaluative conditioning. *Eating Behaviors*, *12*(4), 249–53.

Levey, A. B., & Martin, I. (1975). Classical conditioning of human 'evaluative' responses. *Behaviour Research and Therapy*, *13*(4), 221-226.

Lieberman, M. D. (2003). Reflexive and reflective judgment processes: A social cognitive neuroscience approach. In J. P. Forgas, K. D. Williams, & W. von Hippel (Eds.), *Social judgments: Implicit and explicit processes* (pp. 44–67). New York, NY. Cambridge University Press.

Lin, B. H., Wendt, & M., Guthrie, J. F. (2013). Impact on energy, sodium and dietary fibre intakes of vegetables prepared at home and away from home in the USA. *Public Health Nutrition*. *16*(11), 1937-1943.

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. Sage Publications, Inc.

MacCallumm, R. C., Kim, C., Malarkey, W., & Kiecolt-Glaser, J. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research, 32*, 215 -253.

Maison, D., Greenwald, A. G., & Bruin, R. (2001). The Implicit Association Test as a measure of implicit consumer attitudes. *Polish Psychological Bulletin*, *32*(1), 1–9.

Maison, D., Greenwald, A. G., & Bruin, R. H. (2004). Predictive validity of the Implicit Association Test in studies of brands, consumer attitudes, and behavior. *Journal of Consumer Psychology*, *14*(4), 405-415.

Malhotra, N. K. (1988). Self concept and product choice: An integrated perspective. *Journal of Economic Psychology*, *9*(1), 1–28.

Malik, V. S., Popkin, B. M., Bray, G. A., Després, J. P., Willett, W. C., & Hu, F. B. (2010). Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes a meta-analysis. *Diabetes Care*, *33*(11), 2477-2483.

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, *35*(2), 63-78.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224-253.

Markus, H., & Kunda, Z. (1986). Stability and malleability of the self-concept. *Journal of Personality and Social Psychology*, *51*(4), 858-866.

Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology, 38*, 299-337.

Markus, H., & Zajonc, R. B. (1985). The cognitive perspective in social psychology. *Handbook of Social Psychology*, *1*, 137-230.

Martin, I., & Levey, A. B. (1978). Evaluative conditioning. *Advances in Behaviour Research and Therapy*, *1*(2), 57-101.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, *37*(5), 435-442.

McConnell, A. R., Brown, C. M., Shoda, T. M., Stayton, L. E., & Martin, C. E. (2011). Friends with benefits: On the positive consequences of pet ownership. *Journal of Personality and Social Psychology, 101*, 1239-1252.

McEachan, R.R.C., Conner, M., Taylor, N.J., & Lawton, R.J. (2011). Prospective prediction of health-related behaviours with the Theory of Planned Behaviour: a meta-analysis. *Health Psychological Review*, *5*(2), 97-144.

Meier, B. P., Robinson, M. D., & Caven, A. J. (2008). Why a big mac is a good mac: Associations between affect and size. *Basic and Applied Social Psychology*, *30*(1), 46-55.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*(02), 183-198.

Moors, A., & De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297-326.

Moran, T., & Bar-Anan, Y. (2013). The effect of object–valence relations on automatic evaluation. *Cognition & Emotion, 27*(4), 743-752.

Moreland, R. L., & Topolinski, S. (2010). The mere exposure phenomenon: A lingering melody by Robert Zajonc. *Emotion Review*, *2*(4), 329-339.

Morewedge, C. K., Shu, L. L., Gilbert, D. T., & Wilson, T. D. (2009). Bad riddance or good rubbish? Ownership and not loss aversion causes the endowment effect. *Journal of Experimental Social Psychology*, *45*(4), 947–951.

Mussweiler, T., & Strack, F. (2000). The" relative self": informational and judgmental consequences of comparative self-evaluation. *Journal of Personality and Social Psychology*, *79*(1), 23-38.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. WH Freeman/Times Books/Henry Holt & Co.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition*, *19*(6), 625-666.

Nosek, B.A. (2007). Implicit–explicit relations. *Current Direction in Psychological Science*, *16*(2), 65-69.

Nuttin, J. M. (1985). Narcissism beyond Gestalt and awareness: The name letter effect. *European Journal of Social Psychology*, *15*(3), 353-361.

Nuttin, J. M. (1987). Affective consequences of mere ownership: The name letter effect in twelve European languages. *European Journal of Social Psychology*, *17*(4), 381-402.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*, 413-417.

Olson, M. A., & Fazio, R. H. (2009). Implicit and explicit measures of attitudes: The perspective of the MODE model. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), Attitudes: Insights from the new implicit measures (pp. 19 – 63). New York, NY: Psychology Press.

Olson, M. A., Kendrick, R. V., & Fazio, R. H. (2009). Implicit learning of evaluative vs. non-evaluative covariations: The role of dimension accessibility. *Journal of Experimental Social Psychology*, *45*(2), 398-403.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*, aac4716.

Osgood, C. E., & Tannenbaum, P. H. (1955). The principle of congruity in the prediction of attitude change. *Psychological Review, 62*, 42–55.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1978). The measurement of meaning. 1957. *Urbana, IL: University of Illinois Press.*

Palfai, T. P., & Ostafin, B. D. (2003). Alcohol-related motivational tendencies in hazardous drinkers: assessing implicit response tendencies using the modified-IAT. *Behaviour Research and Therapy*, *41*(10), 1149-1162.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277-293.

Pelham, B. W., Mirenberg, M. C., & Jones, J. T. (2002). Why Susie sells seashells by the seashore: implicit egotism and major life decisions. *Journal of Personality and Social Psychology*, *82*(4), 469-487.

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.

Perloff, L. S., & Fetzer, B. K. (1986). Self–other judgments and perceived vulnerability to victimization. *Journal of Personality and Social Psychology*, *50*(3), 502-510.

Perugini, M. (2005). Predictive models of implicit and explicit attitudes. *British Journal of Social Psychology*, *44*(1), 29–45.

Perugini, M., Richetin, J., & Zogmaister, C. (2010). Prediction of behavior. *Handbook of implicit social cognition: Measurement, theory, and applications*, *10*, 255-278.

Perugini, M., Richetin, J., & Zogmaister, C. (2012). The formation of implicit and explicit attitudes for neutral and valenced stimuli using the self. *Learning & Motivation, 43*, 135-143.

Perugini, M., Richetin, J., & Zogmaister, C. (2014). Indirect measures as a signal for evaluative change. *Cognition & Emotion*, *28*(2), 208-229.

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.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In *Communication and Persuasion* (pp. 1-24). Springer New York.

Petty, R. E., & Wegener, D. T. (1999). The Elaboration Likelihood Model: Current status and controversies.

Petty, R. E., Cacioppo, J. T., Strathman, A. J., & Priester, J. R. (2005). To think or not to think. *Persuasion: Psychological Insights and Perspectives*, 81-116.

Phills, C. E., Kawakami, K., Tabi, E., Nadolny, D., & Inzlicht, M. (2011). Mind the gap: Increasing associations between the self and Blacks with approach behaviors. *Journal of Personality and Social Psychology, 100*, 197-210.

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis) liking: item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 130-144.

Poulton, R., & Menzies, R. G. (2002). Non-associative fear acquisition: A review of the evidence. from retrospective and longitudinal research. *Behaviour Research and Therapy, 40*, 127-149.

Powell, M. C., & Fazio, R. H. (1984). Attitude accessibility as a function of repeated attitudinal expression. *Personality and Social Psychology Bulletin, 10*(1), 139-148.

Pratkanis, A. R., & Greenwald, A. G. (1989). A sociocognitive model of attitude structure and function. *Advances in Experimental Social Psychology*, *22*, 245-285.

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.

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.

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.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (Vol. 1). Sage.

Reed, M. B., & Aspinwall, L. G. (1998). Self-affirmation reduces biased processing of health-risk information. *Motivation & Emotion*, 22(2), 99-132.

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.

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.

Rippetoe, P. A., & Rogers, R. W. (1987). Effects of components of protection-motivation theory on adaptive and maladaptive coping with a health threat. *Journal of Personality and Social Psychology*, *52*(3), 596-604.

Roefs, A., & Jansen, A. (2002). Implicit and explicit attitudes toward high-fat foods in obesity. *Journal of Abnormal Psychology*, *111*(3), 517–521.

Roefs, A., Huijding, J., Smulders, F. T., MacLeod, C. M., de Jong, P. J., Wiers, R. W., & Jansen, A. (2011). Implicit measures of association in psychopathology research. *Psychological Bulletin*, *137*(1), 149-193.

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.

Roheim, C., Asche, F., & Santos, J. (2011). The elusive price premium for ecolabelled products: evidence from seafood in the UK market. *Journal of Agronomic Economy*, *62*(3), 655–668.

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.

Rudman, L. A. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science*, *13*(2), 79-82.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: a systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*(6), 995-1008.

Samson, A., & Voyer, B. G. (2012). Two minds, three ways: Dual system and dual process models in consumer psychology. *AMS Review*, *2*(2-4), 48–71.

Schulze, M. B., Manson, J. E., Ludwig, D. S., Colditz, G. A., Stampfer, M. J., Willett, W. C., & Hu, F. B. (2004). Sugar-sweetened beverages, weight gain, and incidence of type 2 diabetes in young and middle-aged women. *Jama*, *292*(8), 927-934.

Scott, W. A., Kline, J. A., Faguy-Coté, E., & Peterson, C. (1980). Centrality of cognitive attributes. *Journal of Research in Personality*, *14*(1), 12-26.

Shanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *The Quarterly Journal of Experimental Psychology, 60*(3), 291-309.

Sheeran, P., Harris, P. R., & Epton, T. (2014). Does heightening risk appraisals change people's intentions and behavior? A meta-analysis of experimental studies. *Psychological Bulletin*, *140*(2), 511-543.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127-190.

Showers, C. J., & Zeigler-Hill, V. (2003). Organization of self-knowledge: Features, functions and flexibility. In M. R. Leary, & J. P. Tangney (Eds.), Handbook of self and identity (pp. 47 – 67). New York7 The Guilford Press.

Simonsohn, U., Nelson, L.D., Simmons, J.P. (2014a). p-Curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143*, 534–547.

Simonsohn, U., Nelson, L.D., Simmons, J.P. (2014b) p-Curve and Effect Size: Correcting for publication bias using only significant results. *Perspective on Psychological Science, 9*, 666–681.

Skinner, C. S., Strecher, V. J., & Hospers, H. (1994). Physicians' recommendations for mammography: do tailored messages make a difference?. *American Journal of Public Health*, *84*(1), 43-49.

Smith, E. R., Fazio, R. H., & Cejka, M. A. (1996). Accessible attitudes influence categorization of multiply categorizable objects. *Journal of Personality and Social Psychology*, *71*(5), 888-898.

Sparks, P., & Shepherd, R. (1992). Self-identity and the theory of planned behavior: Assesing the role of identification with "green consumerism". *Social Psychology Quarterly*, *55*(4), 388-399.

Sriram, N., & Greenwald, A. G. (2009). The Brief Implicit Association Test. *Experimental Psychology, 56*, 283-294.

Stahl, C., Unkelbach, C., & Corneille, O. (2009). On the respective contributions of awareness of unconditioned stimulus valence and unconditioned stimulus identity in attitude formation through evaluative conditioning. *Journal of Personality and Social Psychology*, *97*(3), 404-420.

Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *Journal of Clinical Epidemiology, 54*(10), 1046-1055.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), 220-247.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), 220–247.

Sweldens, S., Corneille, O., & Yzerbyt, V. (2014). The role of awareness in attitude formation through evaluative conditioning. *Personality and Social Psychology Review*, *18*(2), 187-209.

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: a meta-analysis. *Psychological Bulletin*, *121*(3), 371-394.

Tabachnick, B. G., & Fidell, L. S. (2007). Using multivariate statistics, 5th. Needham Height, MA: Allyn & Bacon. Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). Applied multiple regression/correlation analysis for the behavioral sciences. Routledge.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. The Social Psychology of Intergroup Relations, 33, 74.

Taylor, C., Webb, T. L., & Sheeran, P. (2014). 'I deserve a treat!': Justifications for indulgence undermine the translation of intentions into action. *British Journal of Social Psychology*, *53*(3), 501-520.

Tropp, L. R., & Wright, S. C. (2001). Ingroup identification as the inclusion of ingroup in the self. *Personality and Social Psychology Bulletin, 27*, 585-600.

Vantomme, D., Geuens, M., de Houwer, J., & de Pelsmacker, P. (2005). Implicit attitudes toward green consumer behaviour. *Psychologica Belgica*, *45*, 217–239.

Verhulst, F., Hermans, D., Baeyens, F., Spruyt, A., & Eelen, P. (2006). Determinants and predictive validity of direct and indirect measures of recently acquired food attitudes. *Appetite*, *46*(2), 137–143.

Verosky, S. C., & Todorov, A. (2010). Differential neural responses to faces physically similar to the self as a function of their valence. *NeuroImage, 49*(2), 1690-1698.

Verosky, S. C., & Todorov, A. (2013). When physical similarity matters: Mechanisms underlying affective learning generalization to the evaluation of novel faces. *Journal of Experimental Social Psychology, 49*(4), 661-669.

Walther, E. (2002). Guilty by mere association: evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology*, *82*(6), 919-934.

Walther, E., & Nagengast, B. (2006). Evaluative conditioning and the awareness issue: assessing contingency awareness with the four-picture recognition test. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(4), 454-459.

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.

Walther, E., Nagengast, B., & Trasselli, C. (2005). Evaluative conditioning in social psychology: Facts and speculations. *Cognition & Emotion*, *19*(2), 175-196.
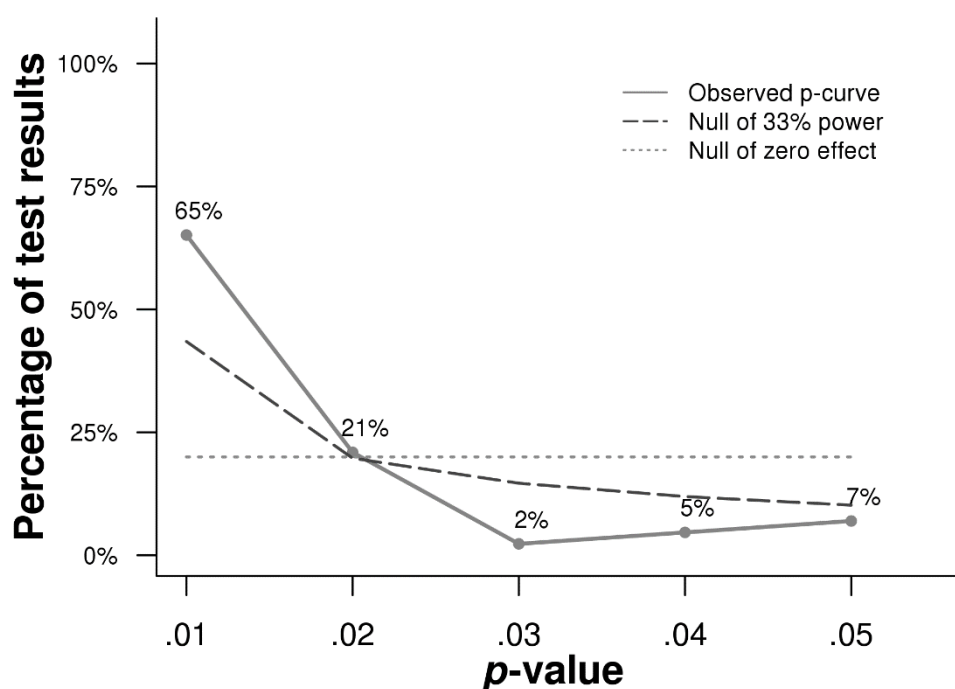
Welsh, J. A., Sharma, A. J., Grellinger, L., & Vos, M. B. (2011). Consumption of added sugars is decreasing in the United States. *The American Journal of Clinical Nutrition*, *94*(3), 726-734.

Wheeler, M., Sharp, A., & Nenycz-Thiel, M. (2013). The effect of "green" messages on brand purchase and brand rejection. *Australasian Marketing Journal (AMJ)*, *21*(2), 105–110.

Wiers, R. W., & Stacy, A. W. (2006). Implicit cognition and addiction. *Current Directions in Psychological Science*, *15*(6), 292-296.

Wilson, T., Lindsey, S., & Schooler, T. (2000). A model of dual attitudes. *Psychological Review*, *107*(1), 101-126.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, *72*(2), 262-274.

Yamaguchi, S., Greenwald, A. G. A., 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.

Young, W., Hwang, K., Mcdonald, S., & Oates, C. J. (2010). Sustainable consumption: Green consumer behaviour when purchasing products. *Sustainable Development*, *18*(1), 20-31.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2), 1-27.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American psychologist*, *35*(2), 151-175.

Zanna, M. P., and Rempel, J. K. "Attitudes: A New Look at an Old Concept," in The Social Psychology of Knowledge, D. Bar-Tal and A. W. Kruglanski (eds.), Cambridge University Press, New York, 1988, pp. 315-334.

Zanon, R., De Houwer, J., & Gast, A. (2012). Context effects in evaluative conditioning of implicit evaluations. *Learning & Motivation, 43*(3), 155-165.

Zanon, R., De Houwer, J., Gast, A., & Smith, C. T. (2014). When does relational information influence evaluative conditioning?. *The Quarterly Journal of Experimental Psychology, 67*(11), 2105-2122.

Zhang, F., Epperson, J. E., Huang, C. L., & Houston, J. E. (2009). Organic price premiums paid for fresh tomatoes and apples by U.S. households: Evidence from Nielsen Homescan Data. *Journal of Food Distribution Research*, *40*(3), 105–114

Zhang, H., & Chan, D. K. S. (2009). Self-esteem as a source of evaluative conditioning. *European Journal of Social Psychology*, *39*(6), 1065-1074.

**Appendix**

**Supplementary Material Chapter 1**

**Appendix A - P curve analyses on the three dependent variables**



| | Binomial Test | Continuous Test | |
| --- | --- | --- | --- |
| | Share of p<.025 | Aggregate pp-values (Stouffer Method) | |
| | | **Full p-curve (p's<.05)** | **Half p-curve (p's<.025)** |
| 1) Studies contain evidential value (Right-skew) | p<.0001 | Z=-12.98, p<.0001 | Z=-12.22, p<.0001 |
| 2) Evidential value, if any, is inadequate (Flatter than when power=33%) | p=.9982 | Z=7.88, p>.9999 | Z=12.29, p>.9999 |
| | **Statistical Power** | | |
| Power of tests included in p-curve (Correcting for publication bias) | Estimate: 91% Confidence interval: (84%,95%) | | |

The observed p-curve includes 43 statistically significant (p<.05) results, of which 38 are p<.025.
There were 10 additional results entered but excluded from p-curve because they were p>.05.

*Figure A3. Distribution of the significant p values among studies on implicit attitude (K = 53).*

| | Binomial Test<br>*Share of p<.025* | Continuous Test<br>*Aggregate pp-values (Stouffer Method)* | |
|---|---|---|---|
| | | **Full p-curve (p's<.05)** | **Half p-curve (p's<.025)** |
| 1) Studies contain evidential value<br>    *(Right-skew)* | *p<.0001* | *Z=-5.45, p<.0001* | *Z=-3.91, p<.0001* |
| 2) Evidential value, if any, is inadequate<br>    *(Flatter than when power=33%)* | *p=.9909* | *Z=1.86, p=.9684* | *Z=5.62, p>.9999* |
| | **Statistical Power** | | |
| Power of tests included in p-curve<br>    *(Correcting for publication bias)* | Estimate: 58%<br>Confidence interval: (36%,76%) | | |

The observed p-curve includes 26 statistically significant (p<.05) results, of which 23 are p<.025.
There were 26 additional results entered but excluded from p-curve because they were p>.05.

*Figure A2. Distribution of the significant p values among studies on explicit attitude (K = 52).*

| | Binomial Test<br>*Share of p<.025* | Continuous Test<br>*Aggregate pp-values (Stouffer Method)* | |
|---|---|---|---|
| | | *Full p-curve (p's<.05)* | *Half p-curve (p's<.025)* |
| 1) Studies contain evidential value<br>*(Right-skew)* | *p=.0312* | *Z=-5.31, p<.0001* | *Z=-4.64, p<.0001* |
| 2) Evidential value, if any, is inadequate<br>*(Flatter than when power=33%)* | *p>.9999* | *Z=3.4, p=.9997* | *Z=4.27, p>.9999* |
| | **Statistical Power** | | |
| Power of tests included in p-curve<br>*(Correcting for publication bias)* | Estimate: 94%<br>Confidence interval: (76%,99%) | | |

The observed p-curve includes 5 statistically significant (p<.05) results, of which 5 are p<.025.
There were 9 additional results entered but excluded from p-curve because they were p>.05.

*Figure A3. Distribution of the significant p values among studies on identification (K = 14).*

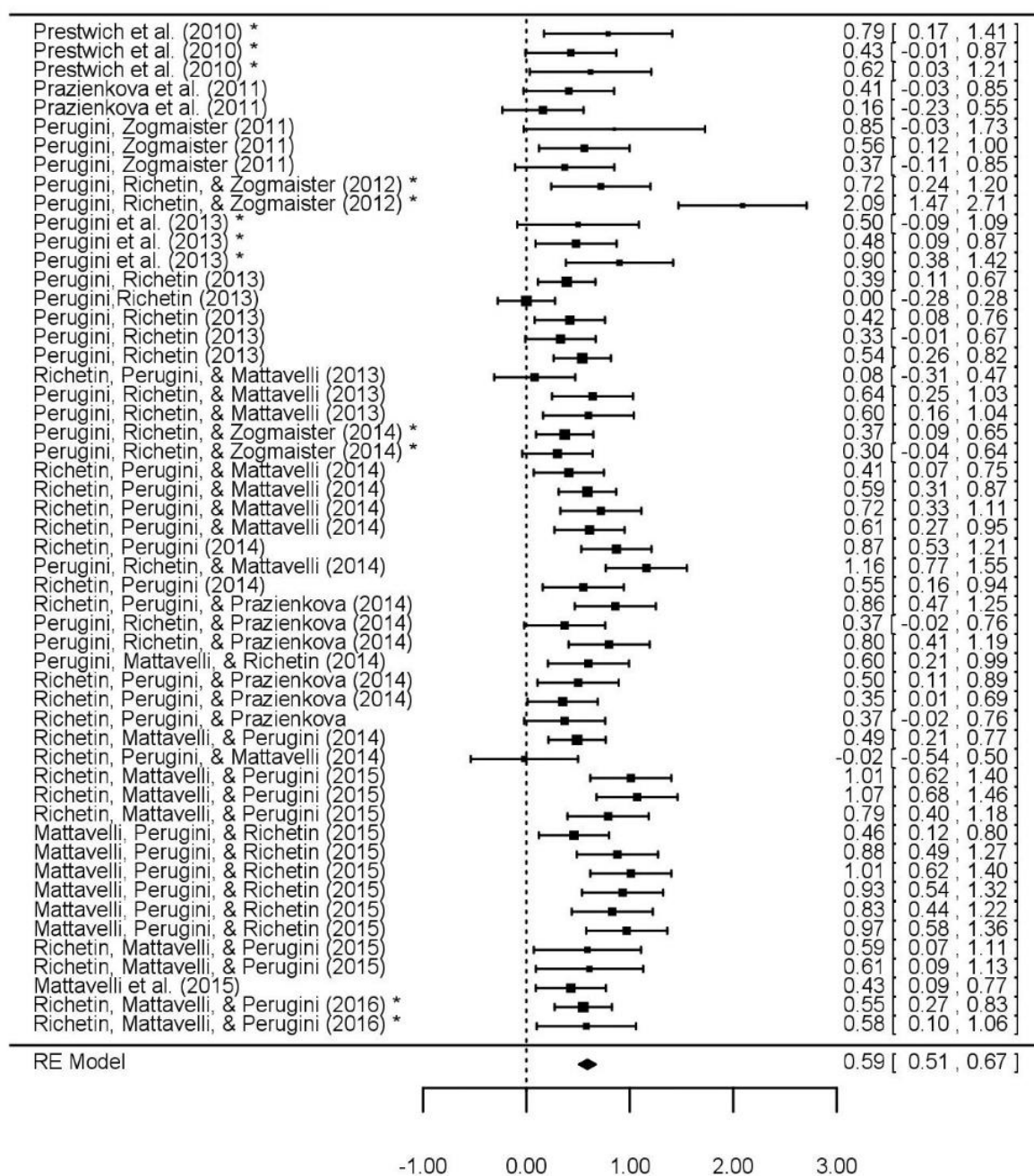**Appendix B - Forest and Funnel Plots for the three dependent variables**



*Figure B1. Forest plot for implicit attitude (K = 53). The left column reports authors and years as key identifiers for each study (* is for published studies). The middle column shows a graphical representation of the single and the overall effect sizes. Cohen's d and confidence intervals are reported on the right columns.*
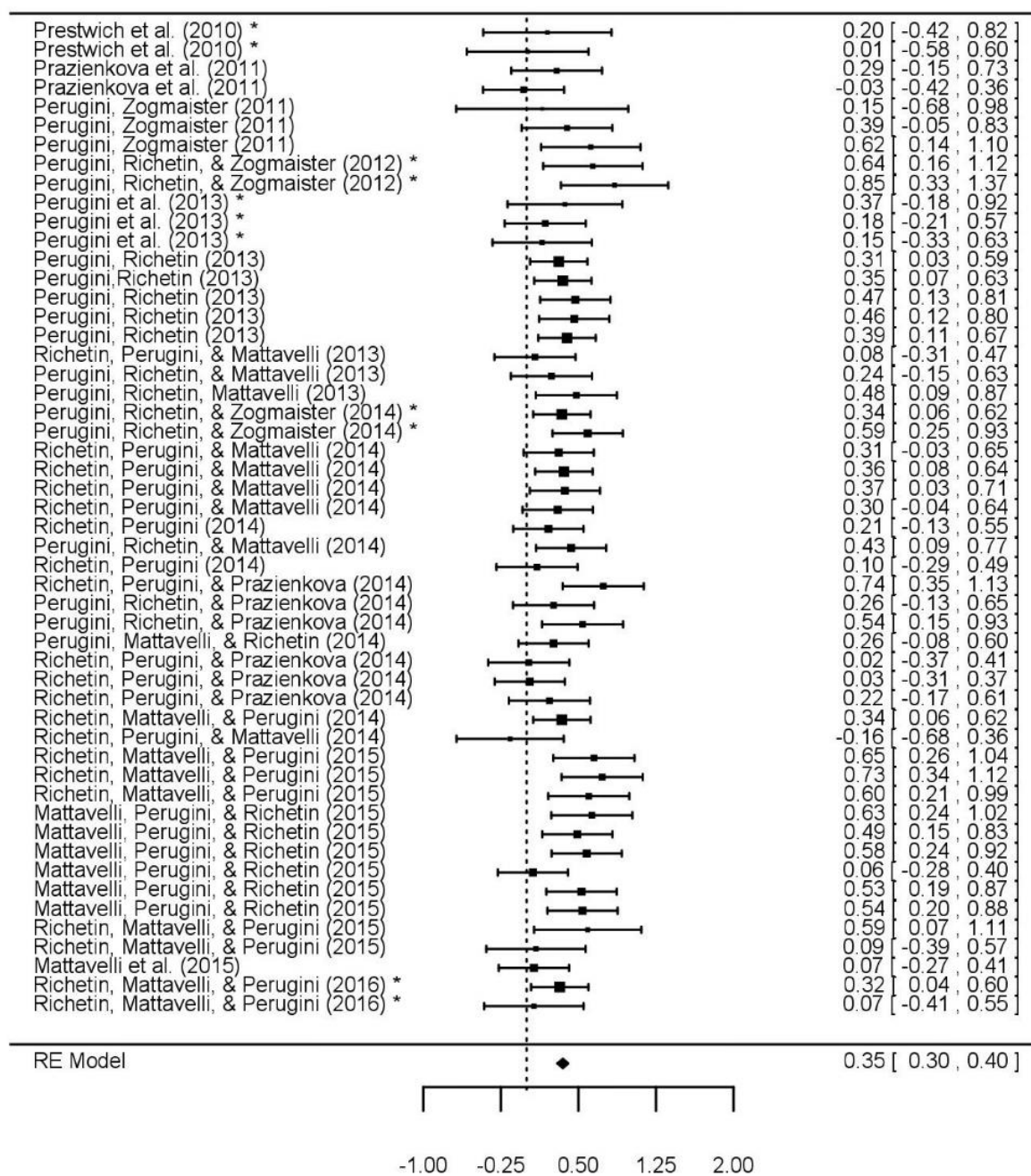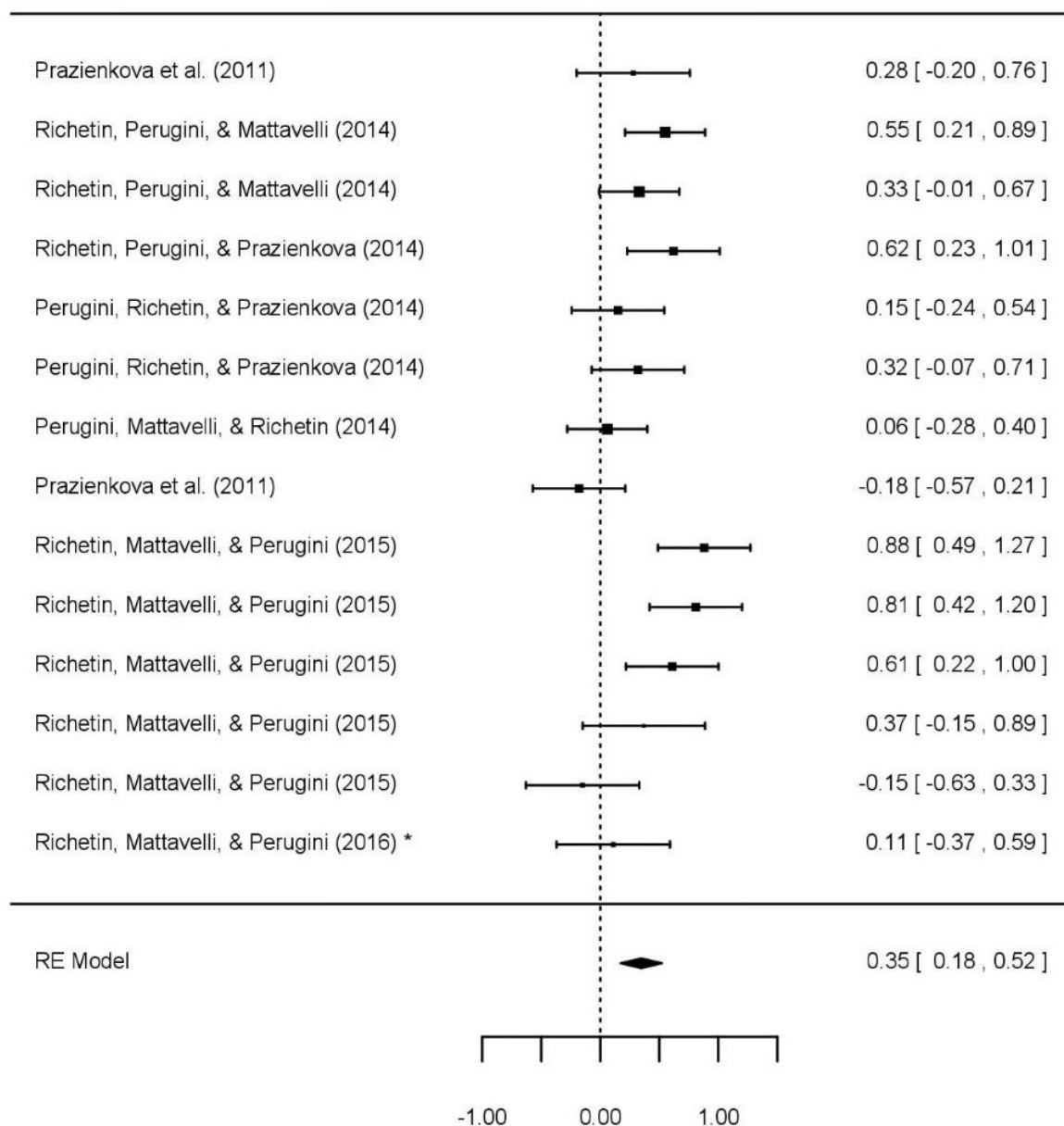
*Figure B2. Forest plot for explicit attitude (K = 52). The left column reports authors and years as key identifiers for each study (\* is for published studies). The middle column shows a graphical representation of the single and the overall effect sizes. Cohen's d and confidence intervals are reported on the right columns.*
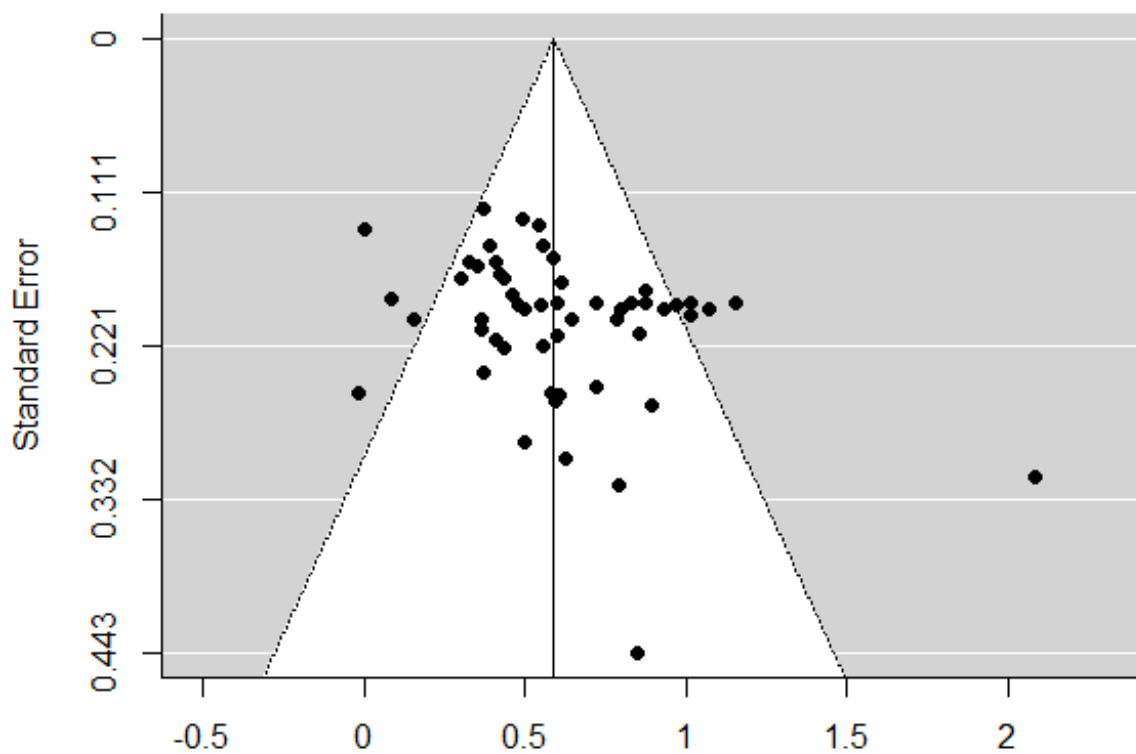
*Figure B3. Forest plot for identification (K = 14). The left column reports authors and years as key identifiers for each study (\* is for published studies). The middle column shows a graphical representation of the single and the overall effect sizes. Cohen's d and confidence intervals are reported on the right columns.*

*Figure B4. Funnel plot representing the distribution of the effect sizes of the SR task on implicit attitude change (K = 53). Each dot represents a study; the y-axis represents study precision (i.e., standard error) and the x-axis shows the study's effect size.*
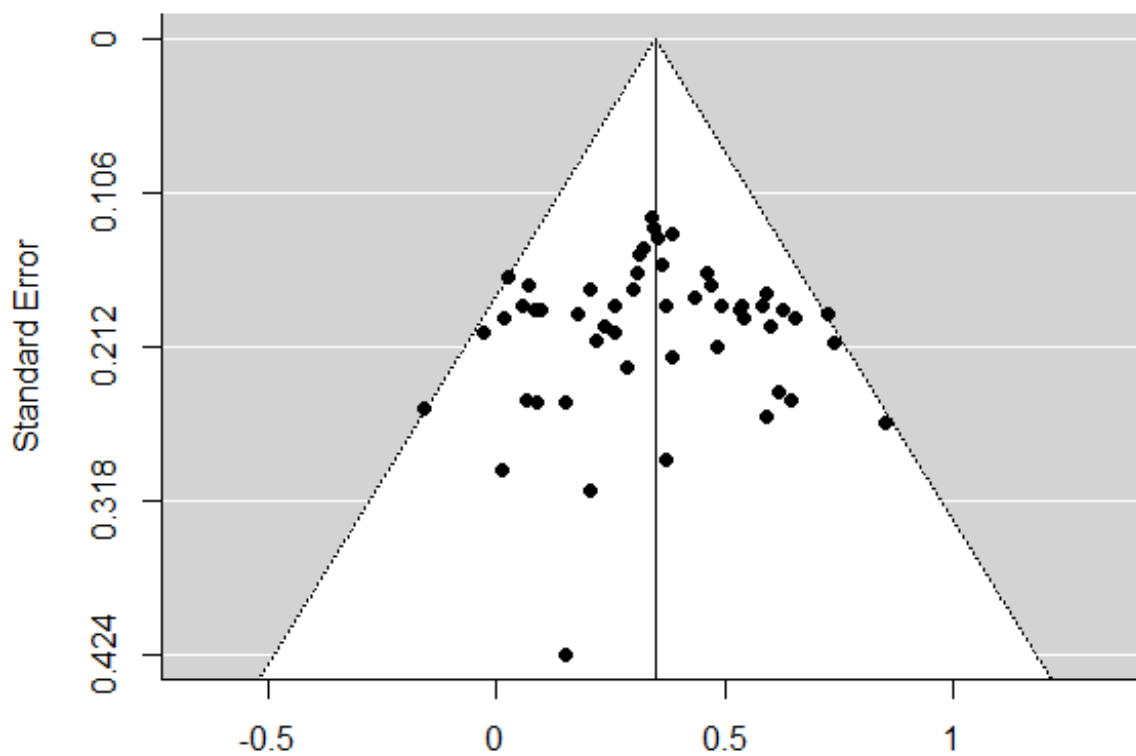
*Figure B5. Funnel plot representing the distribution of the effect sizes of the SR task on explicit attitude change (K = 52). Each dot represents a study; the y-axis represents study precision (i.e., standard error) and the x-axis shows the study's effect size.*
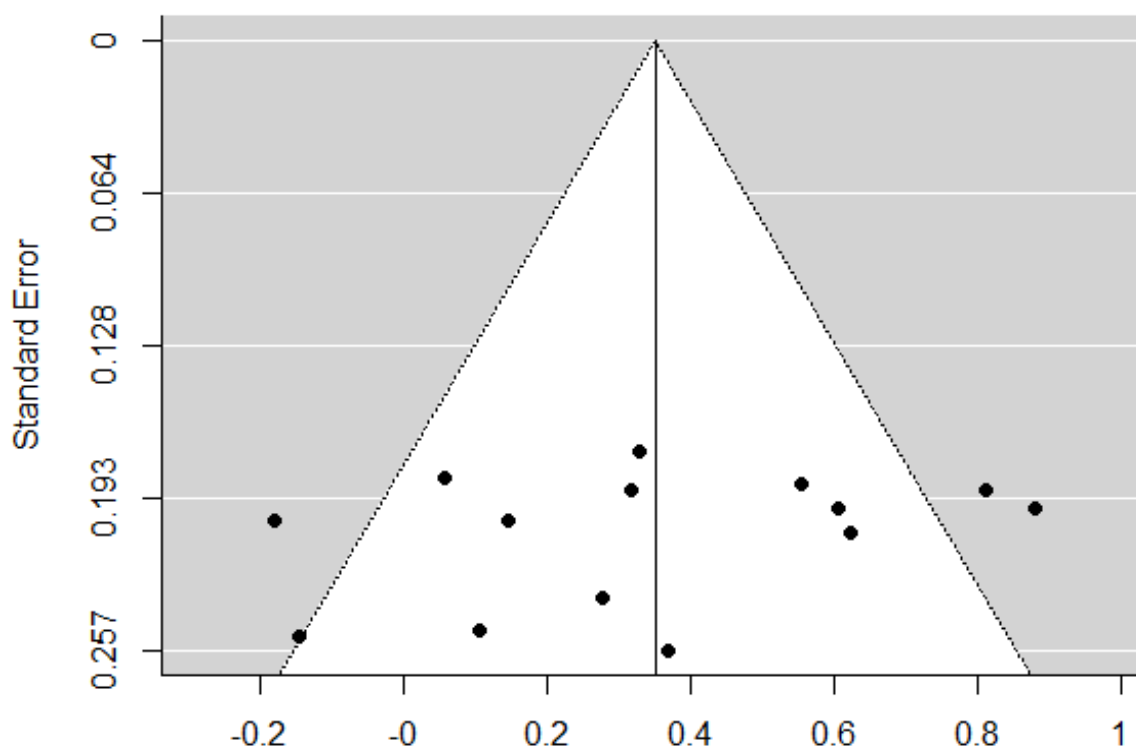
*Figure B6. Funnel plot representing the distribution of the effect sizes of the SR task on identification*

*(K = 14). Each dot represents a study; the y-axis represents study precision (i.e., standard error) and*

*the x-axis shows the study's effect size.*

**Supplementary Material Chapter 2**

**Appendix C - Analyses on all participants**

**Study 1**

We first looked at the effect of the SR manipulation on implicit attitude change. We ran a one-way-ANOVA with SR manipulation as predictor and IAT score as dependent variable. We observed a significant SR effect, $F(1,116) = 33.10$, $p < .001$, $\eta_P^2 = .22$. Participants who pressed the same key for Lestea and self-related moderately positive adjectives showed greater implicit liking for Lestea, compared to those who paired the same brand with non-self-referent adjectives. We repeated the same analysis on explicit attitude change. As dependent variable, we used the differential mean of the four items used in the Semantic differential for each brand, such that higher scores indicated a preference for Lestea. The effect of the SR manipulation was significant and in the predicted direction, $F(1,116) = 15.94$, $p < .001$, $\eta_P^2 = .12$. Finally, the significant effect of the manipulation also occurred at the identification level, $F(1,101) = 19.74$, $p < .001$, $\eta_P^2 = .15$. Participant identified more with the brand categorized with self-referent adjectives. The three dependent variables were positively and significantly correlated (all $r$s > .317).

We then looked at the evaluation of personality adjectives selected. We compared the score for the adjectives selected as relevant for the self with those that were not. Moreover, we compared the two post manipulation evaluations of the adjectives with the normative scores observed in the pretest. A paired-sample t-test revealed that the adjectives selected as representative for one's self where better liked than those that were not self-relevant, $t(117) = 15.53$, $p < .001$. Moreover, selecting adjectives as representative of one's self also increased their liking compared to a normative evaluation of the same adjectives, $t(117) = 11.19$, $p < .001$: Individuals rated more positively the adjectives used as stimuli compared to individuals not in the study. Conversely, selecting adjectives as not representative of one's self decreased their liking compared to a normative evaluation, , $t(117) = -16.22$, $p < .001$.

We ran a series of moderation analysis to see whether the evaluation of the adjectives selected as representative for one's self qualified the impact of the experimental manipulation on any of the outcome measure. Both the predictor (SR manipulation) and the proposed moderator (differential evaluation of self-related versus unrelated adjectives) were standardized. On implicit attitude, the whole model explained 23.61% of the total variance. We found a main effect of the experimental manipulation ($\beta = .48$, $p < .001$), no significant main effect of adjectives rating ($\beta = .10$, $p = .213$) and the interaction term was not significant either ($\beta = .06$, $p = .475$). On explicit attitude, the model explained 16.16% of the total variance. There was a main effect of the experimental manipulation ($\beta = .35$, $p < .001$), no significant main effect of the self-adjectives rating ($\beta = .05$, $p = .586$) but a significant interaction between the two factors ($\beta = .20$, $p = .024$). A simple slope analysis revealed significant SR effect when the self-adjectives rating score was either medium or high ($B = .35$, $p < .001$ and $B = .55$, $p < .001$) but not significant when it was low ($B = .15$, $p = .222$). Johnson-Neyman technique revealed that the SR effect was significant for self-related adjectives values above -.70. We ran the same moderation analysis with identification as dependent variable in the model. The model explained 17.25% of the total variance. We found a main effect of the experimental manipulation ($\beta = .40$, $p < .001$), but not of the self-adjectives rating scores ($\beta = .15$, $p = .082$). The interaction term was not significant ($\beta = .07$, $p = .437$).

**Study 2**

We ran a one-way-ANOVA with SR manipulation as predictor and IAT score as dependent variable. We observed a significant SR effect, $F(1,109) = 29.49$, $p < .001$, $\eta_P^2 = .21$. Participants who categorized Lestea products and self-related moderately negative adjectives showed greater implicit liking for Lestea. The same analysis was conducted on explicit attitude change. The effect of the SR manipulation was also significant on explicit attitude, $F(1,109) = 12.08$, $p = .001$, $\eta_P^2 = .10$. The manipulation also occurred at the identification level, $F(1,109) = 21.90$, $p < .001$, $\eta_P^2 = .17$. Participant

identified more with the brand categorized with self-referent adjectives. The three dependent variables were positively and significantly correlated (all $r$s > .291).

We then looked at the evaluation of personality adjectives selected. Although we did not have an evaluation for those adjectives selected as not representative of one's self, we compared the post manipulation evaluation of the self-related adjectives with the normative scores observed in the pretest. Selecting adjectives as representative of one's self also increased their liking compared to a normative evaluation of the same adjectives, $t(110) = 8.60, p < .001$: Individuals rated more positively the adjectives used as stimuli compared to individuals not in the study. This rating score was used as potential moderator of the SR effect on the outcome measures. On implicit attitude, the whole model explained 22.15% of the total variance. We found a main effect of the experimental manipulation ($\beta = .46, p < .001$), no significant main effect of self-adjectives rating ($\beta = .04, p = .651$) and the interaction term was not significant either ($\beta = -.08, p = .327$). On explicit attitude, the model explained 12.54% of the total variance. There was a main effect of the experimental manipulation ($\beta = .32, p = .004$), no significant main effect of the self-adjectives rating ($\beta = .08, p = .391$) and a non-significant interaction between the two factors ($\beta = .14, p = .124$). Although the interaction did not reach the significance level, a simple slope analysis revealed a pattern of results comparable to what we observed in study 1: the SR effect was significant when the self-adjectives rating score was either medium or high ($B = .32, p < .001$ and $B = .46, p < .001$) but not significant when it was low ($B = .18, p = .172$). We ran the same moderation analysis with identification as dependent variable in the model. The model explained 17.57% of the total variance. We found a main effect of the experimental manipulation ($\beta = .41, p < .001$), and no significant effect of the self-adjectives rating scores ($\beta = .05, p = .562$). The interaction term was not significant ($\beta = .08, p = .385$).

**Study 3**

A one-way-ANOVA revealed the SR manipulation to be significant on implicit attitude towards the two brands, $F(1,103) = 16.37, p < .001, \eta_p^2 = .14$. Participants who categorized Lestea

products and self-related moderately negative adjectives showed greater implicit liking for Lestea. The effect of the SR manipulation was also significant on explicit attitude, $F(1,103) = 9.54$, $p < .003$, $\eta_P^2 = .09$. The manipulation also occurred at the identification level, $F(1,103) = 9.71$, $p = .002$, $\eta_P^2 = .09$. Participant identified more with the brand categorized with self-referent adjectives. The three dependent variables were positively and significantly correlated (all $r$s $> .284$).

We then looked at the evaluation of personality adjectives selected. Contrary to previous studies, we did not have a normative evaluation of the adjectives. Rather, all the adjectives included in the selection list were pre-rated by participants and they all received the same moderately negative rating score (i.e., 3 on a scale from 0 to 6). Therefore, we compared a) the score for the adjectives selected as relevant for the self with those that were not, b) the post manipulation evaluations of the adjectives with the baseline score (i.e., 3). A paired-sample t-test revealed that the adjectives selected as representative for one's self where better liked than those that were not self-relevant, $t(104) = 9.64$, $p < .001$. We also ran one-sample t-tests, setting the critical value as equal to 3, to see whether the post-manipulation rating scores for both self-relevant and irrelevant selected adjectives differed from the baseline rating. Self-relevant adjectives received significantly greater liking after the SR manipulation, $t(104) = 4.16$, $p < .001$. On the other hand, self-irrelevant selected adjectives were more negative, $t(104) = -10.03$, $p < .001$.

Post manipulation rating score towards self-relevant adjectives was used as moderator of the effect of the manipulation on the three outcomes. The model explained 21.31% of the total variance on implicit attitude change. We found a main effect of the experimental manipulation ($\beta = .40$, $p < .001$) and no significant main effect of self-adjectives rating ($\beta = -.08$, $p = .410$). Crucially, the interaction term was significant ($\beta = .26$, $p = .005$). We decomposed the interaction with a simple slope analysis. The SR effect was significant when the self-adjectives rating score was either medium or high ($B = .40$, $p < .001$ and $B = .66$, $p < .001$) but not significant for low scores ($B = .14$, $p = .295$). Johnson-Neyman technique revealed that the SR effect was significant for self-related adjectives values above -.68. On explicit attitude, the model accounted for 9.93% of variance. There was a main

effect of the experimental manipulation ($\beta = .32$, $p = .001$) and no effect of self-adjectives rating ($\beta = -.12$, $p = .216$). The interaction term was not significant either ($\beta = -.05$, $p = .648$). The same analysis was conducted on identification. The whole model explained 14.30% of the total variance. We found a main effect of the experimental manipulation ($\beta = .35$, $p < .001$), a significant main effect of self-adjectives rating ($\beta = -.20$, $p = .038$) and the interaction term was not significant either ($\beta = .11$, $p = .241$).

**Study 4**

A one-way-ANOVA revealed the SR manipulation to be significant on implicit attitude towards the two brands, $F(1,112) = 21.50$, $p < .001$, $\eta_P^2 = .16$. Participants who categorized one brand with self-related moderately negative adjectives showed greater implicit liking for that brand, compared to a contrast brand categorized with self-irrelevant moderately positive adjectives. The effect of the SR manipulation was also significant on explicit attitude, $F(1,112) = 11.47$, $p = .001$, $\eta_P^2 = .09$. The manipulation also occurred at the identification level, $F(1,112) = 21.59$, $p < .001$, $\eta_P^2 = .16$. Participant identified more with the brand categorized with self-referent adjectives, regardless of their negative valence. The three dependent variables were positively and significantly correlated (all $r$s > .421).

We looked at the evaluation of personality adjectives selected. The adjectives included in the selection list were pre-rated by participants. All the adjectives attributed to the self received the same moderately negative rating score (i.e., 3 on a scale from 1 to 6). All the adjectives not attributed to the self received the same moderately positive rating score (i.e., 4 on a scale from 1 to 6). Therefore, we compared a) the post SR score for the adjectives selected as relevant for the self with those that were not, b) the post SR scores of the self-relevant adjectives with the baseline score (i.e., 3) and c) b) the post SR scores of the self-irrelevant adjectives with the baseline score (i.e., 4). A paired-sample t-test revealed that, after being attributed to the self, moderately negative adjectives no longer differed in liking from the moderately positive ones, $t(113) = -.17$, $p = .093$. We also ran a one-sample t-test,

setting the critical value as equal to 3, to see whether the post-manipulation rating scores for self-relevant adjectives differed from the baseline rating. Self-relevant adjectives received significantly greater liking after the SR manipulation, $t(113) = 9.61$, $p < .001$. Another one-sample t-test was conducted, setting the critical value as equal to 4, to see whether the post-manipulation rating scores for self-irrelevant adjectives differed from their baseline rating. Self-irrelevant adjectives received significantly lower liking after the SR manipulation, $t(113) = -13.43$, $p < .001$.

We conducted some moderation analyses to test whether the SR effect on implicit attitude, explicit attitude and identification was qualified by participants' evaluation of the self-referent adjectives. For implicit attitude the model explained 16.81% of the total variance. We found a main effect of the experimental manipulation ($\beta = .40$, $p < .001$), no significant main effect of self-adjectives rating ($\beta = -.07$, $p = .474$) and the interaction term was not significant either ($\beta = .03$, $p = .710$). On explicit attitude the model accounted for 11.45% of variance. There was a main effect of SR manipulation ($\beta = .31$, $p = .002$), but neither a main effect of a main effect of self-adjectives rating nor an interaction between the two factors ($\beta = .12$, $p = .207$ and $\beta = -.05$, $p = .579$, respectively). The same pattern was detected on identification. The model accounted for 17.56% of variance. The effect of the SR manipulation was significant ($\beta = .40$, $p < .001$), the effect of self-adjectives ratings and the interaction term were not ($\beta = .11$, $p = .244$ and $\beta = .06$, $p = .505$, respectively).

**Appendix D – Example of rating and selection tasks for personality traits**

Valuta l'aggettivo seguente cliccando il numero corrispondente alla tua risposta.

**testarda**

| del tutto negativo | molto negativo | un po' negativo | un po' positivo | molto positivo | del tutto positivo |
|:---:|:---:|:---:|:---:|:---:|:---:|
| 1 | 2 | 3 | 4 | 5 | 6 |

Sotto sono presentati degli aggettivi. Ti chiediamo di sceglierne 10: i 5 aggettivi che ti descrivono di più (quelli che possono applicarsi di più a te) e i 5 aggettivi che ti descrivono di meno (quelli che possono applicarsi di meno a te)

Scegli i 5 aggettivi che ti descrivono di più cliccando su l'opzione 'io' di ogni dei 5 aggettivi scelti.
Scegli i 5 aggettivi che ti descrivono di meno cliccando su l'opzione 'non io' di ogni dei 5 aggettivi scelti.

| accomodante | enigmatica | rigida | sfrontata | indiscreta |
|---|---|---|---|---|
| ☐ io  ☐ non io | ☐ io  ☐ non io | ☐ io  ☐ non io | ☐ io  ☐ non io | ☐ io  ☐ non io |

| docile | incoerente | solitaria | introversa | consuetudinaria |
|---|---|---|---|---|
| ☐ io  ☐ non io | ☐ io  ☐ non io | ☐ io  ☐ non io | ☐ io  ☐ non io | ☐ io  ☐ non io |

| sincera | remissiva | circospetta | timida | arrogante |
|---|---|---|---|---|
| ☐ io  ☐ non io | ☐ io  ☐ non io | ☐ io  ☐ non io | ☐ io  ☐ non io | ☐ io  ☐ non io |

| egocentrica | sarcastica | maliziosa | ipersensibile | scettica |
|---|---|---|---|---|
| ☐ io  ☐ non io | ☐ io  ☐ non io | ☐ io  ☐ non io | ☐ io  ☐ non io | ☐ io  ☐ non io |

| irascibile | contestatrice | impaziente | dominante | mutevole |
|---|---|---|---|---|
| ☐ io  ☐ non io | ☐ io  ☐ non io | ☐ io  ☐ non io | ☐ io  ☐ non io | ☐ io  ☐ non io |

**Appendix E – Stimuli used in the tasks of the experimental procedure**

| **Familiarization Task (pictures only) and SR task** | |
| --- | --- |
| Lestea | Sabea |



| **IAT** | |
| --- | --- |
| Lestea | Sabea |



| *Positive (Positivo)* | *Negative (Negativo)* |
| --- | --- |
| positive, joy, happy, paradise, nice | negative, hell, ugly, sad, pain |
| *(positivo, gioia, contento, paradiso, bello)* | *(negativo, inferno, brutto, triste, dolore)* |

**Explicit attitude measure**

Bipolar dimensions for single evaluation

bad-good, unpleasant-pleasant, negative-positive useless-worthwhile

(*brutto-bello, spiacevole-piacevole, negativo-positivo, inutile-utile)*

**Supplementary Material Chapter 3**

**Appendix F - Analyses on all participants**

**Study 1**

The effect of the SR manipulation significantly predicted implicit attitude towards the two groups, $F(1,116) = 7.50$, $p = .007$, $\eta_p^2 = .06$. For the explicit attitude score a one-way ANOVA revealed a significant effect of SR manipulation, $F(1,116) = 11.72$, $p = .001$, $\eta_p^2 = .09$. Thus, implicitly and explicitly the group paired with the self is better liked.

For the indirect effect of the SR manipulation we found a marginally significant effect on IAT score, $F(1,116) = 2.85$, $p = .094$, $\eta_p^2 = .02$. Also, we found a non significant effect on explicit attitude, $F(1,116) = 2.54$, $p = .114$, $\eta_p^2 = .02$.

**Study 2**

The effect of the SR manipulation significantly predicted implicit attitude towards the brand1 opposed to brand2, $F(1,118) = 22.82$, $p < .001$, $\eta_p^2 = .16$. A one-way ANOVA revealed a significant effect of SR manipulation, $F(1,118) = 7.22$, $p = .008$, $\eta_p^2 = .06$.

The indirect effect of the SR task was significant on implicit attitude, $F(1,118) = 7.05$, $p = .009$, $\eta_p^2 = .06$. Also, we found a non significant effect on explicit attitude, $F(1,118) = 2.67$, $p = .105$, $\eta_p^2 = .02$.

**Study 3**

The effect of the SR manipulation significantly predicted implicit attitude towards the two brands, $F(1,118) = 28.61$, $p < .001$, $\eta_p^2 = .20$. Also on explicit attitude score we found a significant effect of SR manipulation, $F(1,118) = 8.68$, $p = .004$, $\eta_p^2 = .07$.

A one-way-ANOVA showed a significant effect on implicit attitude, $F(1, 116) = 4.58$, $p = .034$, $\eta_p^2 = .04$. Conversely, and contrary to what observed in study 3, the effect was not significant on explicit attitude change, $F(1, 116) = .46$, $p = .457$. This indicates that people like implicitly, but not explicitly, the group related to the brand previously related with the self.

**Study 4**

We ran a one-way ANOVA, with SR condition (group1-Self vs. group2-Self) as factor. The IAT score was calculated such as higher score indicated a preference for Group1. The effect of the SR manipulation significantly predicted implicit attitude towards the two groups, $F(1,117) = 19.90$, $p < .001$, $\eta_p^2 = .15$. Also on explicit attitude score we found a significant effect of SR manipulation, $F(1,117) = 8.55$, $p = .004$, $\eta_p^2 = .07$.

A one-way-ANOVA showed a significant effect of the same factor on explicit attitude, $F(1,116) = 7.68$, $p = .007$, $\eta_p^2 = .06$. However, the effect was no longer significant at the implicit level, $F(1,116) = 1.87$, $p = .174$, $\eta_p^2 = .02$.

**Study 5**

We ran a one-way ANOVA, with SR condition as factor. The IAT score was calculated such as higher score indicated a preference for Target1. The effect of the SR manipulation significantly predicted implicit attitude towards the two groups, $F(1,120) = 13.64$, $p < .001$, $\eta_p^2 = .10$. Also on explicit attitude score we found a significant effect of SR manipulation, $F(1,120) = 8.73$, $p = .004$, $\eta_p^2 = .07$.

For the indirect effect of the SR, a one-way-ANOVA showed no significant effect on implicit attitude, $F(1,117) = 1.20$, $p = .277$, $\eta_p^2 = .01$. The effect was not significant at the explicit level either, $F(1,117) = .16$, $p = .686$, $\eta_p^2 = .001$.

We tested a series of moderation analysis to see whether individuals' ability to identify equivalence between stimuli did impact the indirect effect of the self-referencing manipulation. As a first moderator, we considered the number of correct responses on the sixteen MTS trials that measured indirect equivalences - the ability to relate stimuli never encountered together in the same task throughout the experimental session (i.e., self/other words and brand stimuli). All variables were centered (standardized). A first moderation analysis was conducted on implicit attitude. The model accounted for 14.21% of variance. We did not observe a main effect of either the SR manipulation (β

= .13, $p$ = .179) or the MTS score ($\beta$ = .14, $p$ = .148). However, the interaction term was significant ($\beta$ = .35, $p$ < .001). We decomposed the effect with a simple slope analysis and observed that for participants whose score on the MTS was low or medium the manipulation was not significant ($B$ = -.22, $p$ = .121 and $B$ = .13, $p$ = .179, respectively) while it was significant for those scoring high on the MTS ($B$ = .47, $p$ < .001). The same analysis was conducted on explicit attitude change. The model accounted for 11.96% of variance. We did not observe a main effect of either the SR manipulation ($\beta$ = -.04, $p$ = .699) or the MTS score ($\beta$ = .07, $p$ = .470). However, the interaction term was significant ($\beta$ = .36, $p$ < .001). We decomposed the effect with a simple slope analysis and observed that for participants whose score on the MTS was medium the manipulation was not significant ($B$ = -.04, $p$ = .699) while it was significant and in the opposite direction, for those scoring low and high on the MTS ($B$ = -.39, $p$ = .007 and $B$ = .31, $p$ = .013, respectively).

On the same dependent variables we also ran moderation analysis by considering MTS score on the thirty-two trials that measured directly trained equivalences - the ability to identify equivalence between object categorized together in the same task (i.e., self/other stimuli and groups as well as groups and brands). The model accounted for 9.16% of variance on implicit attitude change. We did not observe a main effect of either the SR manipulation ($\beta$ = .10, $p$ = .285) or this MTS score ($\beta$ = .04, $p$ = .673). However, the interaction term was significant ($\beta$ = .29, $p$ = .002). We decomposed the effect with a simple slope analysis and observed that for participants whose score on the MTS was low or medium the manipulation was not significant ($B$ = -.19, $p$ = .136 and $B$ = .10, $p$ = .285, respectively) while it was significant for those scoring high on the MTS ($B$ = .39, $p$ = .003). On explicit attitude, the model explained 8.18% of variance on explicit attitude. We did not observe a main effect of either the SR manipulation ($\beta$ = -.04, $p$ = .663) or this MTS score ($\beta$ = .04, $p$ = .649). However, the interaction term was significant ($\beta$ = .29, $p$ = .002). We decomposed the effect with a simple slope analysis and observed that for participants with medium MTS score the manipulation was not significant ($\beta$ = -.04, $p$ = .663) while it was significant and negative for those scoring low ($B$ = -.33, $p$ = .013) and borderline significant, but positive, for high scores ($B$ = .25, $p$ = .054).
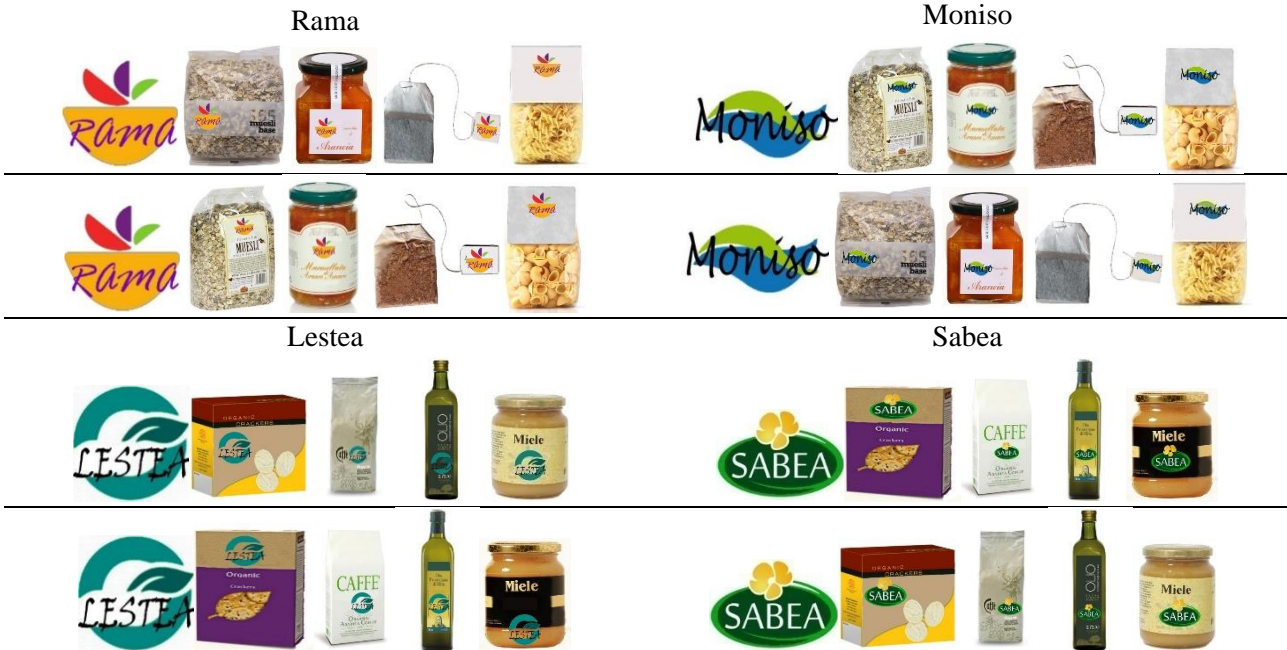
# Appendix G – Stimuli used in the tasks of the experimental procedure

## Study 1

| TC task | |
|---|---|
| Dattiani | Lerriani |
| pagadat, beradat, moradat, calidat, punidat | panaler, docoler, lisaler, mugiler, rediler |
| Craviani | Zimmiani |
| cumacrav, paricrav, delecrav, cidacrav, dalocrav | nofizim, vorozim, polezim, solazim, binazim |

| SR task | |
|---|---|
| Dattiani | Lerriani |
| dattiani, tugadat, zabudat, nisedat, tabedat | lerriani, gecoler, limiler, zopaler, vatuler |
| Me (*Io*) | Others (*Altri*) |
| me, mine, I, myself, my | others, they, their, she, he |
| (*io, mio, me, mia, miei*) | (*altri, altre, loro, esse, essi*) |

| IAT Dattiani/Lerriani | |
|---|---|
| Dattiani | Lerriani |
| dattiani, betodat, lododat, giladat, sobidat | lerriani, piviler, lineler, banaler, neruler |
| Positive (*Positivo*) | Negative (*Negativo*) |
| positive, joy, happy, paradise, nice | negative, hell, ugly, sad, pain |
| (*positivo, gioia, contento, paradiso, bello*) | (*negativo, inferno, brutto, triste, dolore*) |

| IAT Craviani/Zimmiani | |
|---|---|
| Craviani | Zimmiani |
| craviani, semacrav, tadicrav, silicrav, binocrav | zimmiani, manazim, lonazim, tilazim, daruzim |
| Positive (*Positivo*) | Negative (*Negativo*) |
| positive, joy, happy, paradise, nice | negative, hell, ugly, sad, pain |
| (*positivo, gioia, contento, paradiso, bello*) | (*negativo, inferno, brutto, triste, dolore*) |

| Explicit attitude measure |
|---|
| Bipolar dimensions for single evaluation |
| bad-good, unpleasant-pleasant, negative-positive useless-worthwhile |
| (*brutto-bello, spiacevole-piacevole, negativo-positivo, inutile-utile*) |

**Study 2**

---

**TC task**

| Rama | Moniso |
|:---:|:---:|
|  |  |
|  |  |

| Lestea | Sabea |
|:---:|:---:|
|  |  |
|  |  |

---

**SR task**

| Rama | Moniso |
|:---:|:---:|
|  |  |
|  |  |

| Me (*Io*) | Others (*Altri*) |
|:---:|:---:|
| me, mine, I, myself, my | others, they, their, she, he |
| (*io, mio, me, mia, miei*) | (*altri, altre, loro, esse, essi*) |

---

**IAT Rama/Moniso**

| Rama | Moniso |
|:---:|:---:|
|  |  |
|  |  |

| Positive (*Positivo*) | Negative (*Negativo*) |
|:---:|:---:|
| positive, joy, happy, paradise, nice | negative, hell, ugly, sad, pain |
| (*positivo, gioia, contento, paradiso, bello*) | (*negativo, inferno, brutto, triste, dolore*) |

---

**IAT Lestea/Sabea**

| Lestea | Sabea |
|:---:|:---:|

Positive (*Positivo*)
positive, joy, happy, paradise, nice
(*positivo, gioia, contento, paradiso, bello*)

Negative (*Negativo*)
negative, hell, ugly, sad, pain
(*negativo, inferno, brutto, triste, dolore*)

**Explicit attitude measure**

Bipolar dimensions for single evaluation
bad-good, unpleasant-pleasant, negative-positive useless-worthwhile
(*brutto-bello, spiacevole-piacevole, negativo-positivo, inutile-utile)*

**Study 3**

---

| TC task | |
|---|---|
| Lerriani | Datttiani |
| panaler, docoler, lisaler, mugiler, rediler | dattiani, tugadat, zabudat, nisedat, tabedat |
| Gester | Vabel |



---

| SR task | |
|---|---|
| Gester | Vabel |



| Me (*Io*) | Others (*Altri*) |
|---|---|
| me, mine, I, myself, my | others, they, their, she, he |
| (*io, mio, me, mia, miei*) | (*altri, altre, loro, esse, essi*) |

---

**IAT Gester/Vabel**

| Gester | Vabel |
|---|---|



| Positive (*Positivo*) | Negative (*Negativo*) |
|---|---|
| positive, joy, happy, paradise, nice | negative, hell, ugly, sad, pain |
| (*positivo, gioia, contento, paradiso, bello*) | (*negativo, inferno, brutto, triste, dolore*) |

---

**IAT Lerriani/Dattiani**

| Lerriani | Dattiani |
|---|---|
| lerriani, piviler, lineler, banaler, neruler | dattiani, betodat, lododat, giladat, sobidat |
| Positive (*Positivo*) | Negative (*Negativo*) |
| positive, joy, happy, paradise, nice | negative, hell, ugly, sad, pain |
| (*positivo, gioia, contento, paradiso, bello*) | (*negativo, inferno, brutto, triste, dolore*) |

---

**Explicit attitude measure**

Bipolar dimensions for single evaluation

bad-good, unpleasant-pleasant, negative-positive useless-worthwhile

(*brutto-bello, spiacevole-piacevole, negativo-positivo, inutile-utile*)

**Studies 4 and 5**

---

**TC task**

| Lerriani | Datttiani |
|---|---|
| panaler, docoler, lisaler, mugiler, rediler | dattiani, tugadat, zabudat, nisedat, tabedat |

| Gester | Vabel |
|---|---|

**SR task**

| Lerriani | Dattiani |
|---|---|
| lerriani, gecoler, limiler, zopaler, vatuler | dattiani, tugadat, zabudat, nisedat, tabedat |
| Me (*Io*) | Others (*Altri*) |
| me, mine, I, myself, my | others, they, their, she, he |
| (*io, mio, me, mia, miei*) | (*altri, altre, loro, esse, essi*) |

**IAT Gester/Vabel**

| Gester | Vabel |
|---|---|

| Positive (*Positivo*) | Negative (*Negativo*) |
|---|---|
| positive, joy, happy, paradise, nice | negative, hell, ugly, sad, pain |
| (*positivo, gioia, contento, paradiso, bello*) | (*negativo, inferno, brutto, triste, dolore*) |

**IAT Lerriani/Dattiani**

| Lerriani | Dattiani |
|---|---|
| lerriani, piviler, lineler, banaler, neruler | dattiani, betodat, lododat, giladat, sobidat |
| Positive (*Positivo*) | Negative (*Negativo*) |
| positive, joy, happy, paradise, nice | negative, hell, ugly, sad, pain |
| (*positivo, gioia, contento, paradiso, bello*) | (*negativo, inferno, brutto, triste, dolore*) |

**Explicit attitude measure**

Bipolar dimensions for single evaluation

bad-good, unpleasant-pleasant, negative-positive useless-worthwhile

(*brutto-bello, spiacevole-piacevole, negativo-positivo, inutile-utile)*

**MTS task (Study 5 only)**

| Lerriani | Dattiani |
|---|---|
| Lerriani, Lebaler, Vimoler | Dattiani, Delodat, Gemadat |
| Gester | Vabel |

Me (*Io*)

mine, myself

*(il mio, me stesso)*



Others (*Altri*)

theirs, themselves

*(il loro, loro stessi)*

**Supplementary Material Chapter 4**

**Appendix H - Analyses on all participants**

**Study 1**

The analyses involved a 2 (SR Condition: "Ambio" + "Self" vs. "Ecove" + "Self") x 2 (Order of measures: IAT first vs. Behavioral choice first) analysis of variance (ANOVA). There was a significant main effect of the SR manipulation on the IAT, $F(1, 186) = 14.52$, $p < .001$, $\eta_p^2 = .07$.

For the explicit attitude score, a Principal Components Analysis revealed one factor accounting for 55.80 % (factor loadings ranging from .52 to .86). Higher factor scores indicated a more positive evaluation for "Ambio" over "Ecove". The main effect of SR manipulation on the relative explicit evaluation was marginally significant, $F(1, 186) = 2.95$, $p = .086$, $\eta_p^2 = .02$. Moreover, this SR manipulation also resulted in a non-significant effect on the shopping score, $F(1, 186) = 1.83$, $p = .178$. There was no main effect of Order of measures on the IAT score and on the shopping score, $F(1, 186) = .00$, $p = .985$ and $F(1, 186) = .39$, $p = .532$, respectively. There was no interaction effect either on the IAT score, $F(1, 186) = .42$, $p = .518$ or on the shopping score, $F(1, 186) = .00$, $p = .977$.

The three dependent variables were correlated (all $r$'s > .37). We investigated whether the SR effect on explicit attitude was mediated by the implicit attitude. The analysis revealed a significant indirect effect, $M = .11$, 95% CI [.05, .19]. Moreover, the non significant effect of the SR manipulation on the explicit attitude when controlling for the effect of the implicit attitude (direct effect $\beta = .01$, $p = .835$, 95% CI [-.12, .15]) indicated a full mediation. Second, we tested a serial mediation hypothesis in which implicit and explicit attitude mediated serially the SR effect on the choice in the shopping task. The analysis revealed a non significant direct effect of the SR manipulation on the shopping task score $\beta = -.01$, $p = .934$, 95% CI [-.14, .13], indicating a full double mediation. There was a significant indirect effect of the SR manipulation on the shopping task score through the implicit attitude, $M = .06$, 95% CI [.03, .12]. No significant indirect effect of the SR manipulation on the shopping task score through the explicit attitude emerged, $M = -.004$, 95% CI [-.04, .05]. Moreover, there was a

significant indirect effect of the SR manipulation on the shopping task score through the implicit attitude and the explicit attitude, $M = .04$, 95% CI [.02, .08], indicating a serial mediation. The effect of SR manipulation on the shopping task score was mediated by the implicit and by the explicit attitude in a serial manner.

**Study 2**

For the implicit attitude score, we ran a 2 (SR condition: Ambio+Self vs. Ecove+Self) x 2 (Pairing removal: simple brand categorization vs. control) analysis of variance (ANOVA). A main effect of SR condition was observed, $F(1,122) = 15.37$, $p < .001$, $\eta_p^2 = .11$. Neither the SR x Pairing Removal interaction nor the main effect of the latter factor were significant, $F(1,122) = .16$, $p = .692$ and $F(1,122) = .61$, $p = .437$, respectively.

We repeated the same analysis for explicit attitude score. A Principal Components Analysis revealed one factor accounting for 52.62 % (factor loadings ranging from .55 to .85). A main effect of experimental manipulation was found, $F(1,122) = 6.43$, $p = .013$, $\eta_p^2 = .05$. There was neither a significant interaction between SR manipulation and the pairing removal manipulation, $F(1,122) = 1.80$, $p = .182$, nor a main effect of the latter factor, $F(1,122) = .42$, $p = .517$.

We repeated the same analysis for participants' identification with the two brands. There was a borderline significant effect of experimental manipulation, $F(1,122) = 3.14$, $p = .079$, $\eta_p^2 = .03$; also, there was neither a significant interaction between SR manipulation and pairing removal manipulation, $F(1,122) = 1.20$, $p = .275$, nor a main effect of the latter factor, $F(1,122) = .36$, $p = .552$.
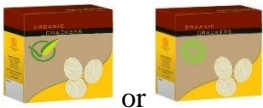
The three dependent variables were all significantly correlated (all $r$'s > .43). We tested two mediation hypotheses. First, we investigated whether the SR effect on explicit attitude was mediated by the implicit attitude. The analysis revealed a full mediation with a non significant effect of the SR manipulation on the explicit attitude score when controlling for the effect of the implicit attitude score, $\beta = .07$, $p = .412$, 95% CI [-.10, .23] and a significant indirect effect, $M = .15$, 95% CI [.07,

.27]. Second, we tested a serial mediation hypothesis in which both implicit and explicit attitude mediated the SR effect on the brand identification score. The analysis revealed a non significant direct effect of the SR manipulation of the identification score, $\beta$ = -.04, $p$ = .545, 95% CI [-.16, .08], indicating a full mediation. There was a non significant indirect effect of the SR manipulation on the identification score through the implicit attitude, $M$ = .03, 95% CI [-.005, .08]. Also, a non significant indirect effect of the SR manipulation on identification through the explicit attitude emerged, $M$ = .05, 95% CI [-.06, .17]. Finally, there was a significant indirect effect of the SR manipulation on the identification through the implicit attitude and the explicit attitude, $M$ = .11, 95% CI [.06, .21], indicating a serial mediation.

**Overall SR effect Study 1 and Study 2**

For the implicit attitude, Cochran's $Q$ statistic yielded a non significant effect, $Q(1) = 0.34$, $p$ = .562, indicating homogeneity and therefore suggesting to apply a fixed-effect model. The overall effect was significant ($z$ = 5.29, $p$ < .001) with an average effect size $d$ = 0.61, 95% CI: [0.38, .84]. For the explicit attitude, Cochran's $Q$ statistic yielded also a non significant effect, $Q(1) = 0.73$, $p$ = .394, indicating homogeneity and therefore suggesting to apply a fixed-effect model. The overall effect was significant ($z$ = 2.90, $p$ = .004) with an average effect size $d$ = 0.33, 95% CI: [0.11, 0.55].

## Appendix I – Stimuli used in the tasks of the experimental procedure

| **Familiarization Task (pictures only) and SR task** | |
|---|---|
| Ambio | Ecove |
|  |  |
|  |  |
| Me *(Io)* | Others *(Altri)* |
| Me, myself, mine, my, I | others, they, their, she, he |
| (io, mio, me, mia, miei) | (altri, altre, loro, esse, essi) |

| **IAT** | |
|---|---|
| Ambio | Ecove |
|  |  |
|  |  |
| Positive *(Positivo)* | Negative *(Negativo)* |
| positive, joy, happy, paradise, nice | negative, hell, ugly, sad, pain |
| *(positivo, gioia, contento, paradiso, bello)* | *(negativo, inferno, brutto, triste, dolore)* |

**Explicit attitude measure**

Bipolar dimensions for single evaluation

bad-good, unpleasant-pleasant, negative-positive useless-worthwhile

*(brutto-bello, spiacevole-piacevole, negativo-positivo, inutile-utile)*

**Shopping task**

| | | | |
|---|---|---|---|
|  or | Caffè al gusto cremoso e grande energia grazie alla selezione attenta delle migliori coltivazioni |  or | Chicchi di caffè lavorati con attenzione donano un sapore denso e risvegliano ogni senso |
|  or | Un piacevole break con sfoglia di grano saporitissima |  or | Spezzare la fame con un morso di spighe dal gusto delicato |
|  or | Genuinità garantita da olive di qualità, indicato per tutte le generazioni |  or | Olio autentico per ogni età, certificato per la sua naturalità |

| | | | | |
|---|---|---|---|---|
| or | Un cucchiaino al giorno per addolcire tutto intorno | or | Atteso appuntamento per deliziare i sensi |
| or | Chicco che non scuoce mai, ideale per tutte le ricette | or | Riso prelibato sempre al dente per soddisfare tutti i cuochi |
| or | Un'oasi di tè, il gusto da sorseggiare | or | Bere tè, il piacere che rinfresca |

**Supplementary Materials Chapter 5**

**Appendix L – Stimuli used in the task of the experimental procedure**

| SR task |
|---|



| Me | Other |
|---|---|
| *Me, My, Mine, I, Myself* | *They, Them, Their, Other, Theirs* |

| ST IAT |
|---|



| Pleasant | Unpleasant |
|---|---|
| *good, pleasant, nice, love* | *bad, unpleasant, nasty, hate* |

**Supplementary Materials Chapter 6**

**Appendix M - Analyses on all participants**

**Study 2**

We first look at the effect of the manipulation on the implicit association between fat-related words and either the drinks. A one-way ANOVA showed a significant main effect of the FSR condition on the BIAT score, $F(1,73) = 7.75$, $p = .007$, $\eta_P^2 = .10$. Participants were faster in categorizing fat words with Shasta when this was paired with their fat face in the SR task.

Then we tested the effect of the SR on the amount of calories attributed to a 355ml Shasta's can and the amount of time required to walk the drink off. The two variables were correlated each other ($r = .40$). Even though the means followed the predicted pattern, with more calories attributed to Shasta when this was paired the fat self, the effect of the SR was not significant, $F(1,73) = .96$, $p = .330$, $\eta_P^2 = .01$. We observed no SR effect when we entered walking time as dependent variable, $F(1,73) = .21$, $p = .651$, $\eta_P^2 < .01$.

We ran a one-way ANOVA on the difference between the initial and the final weight of the 355ml Shasta's can and show no FSR effect, $F(1,73) = 1.30$, $p = .258$, $\eta_P^2 = .02$.

We looked at whether pairing Shasta with one's fat face would influence the ratings of the drink after the tasting phase. We ran a Principle Component Analyses on all thirteen items. A two factors solution explained 46.77% of the overall variance. Given that a preliminary Oblimin rotation showed very little correlation between the two factors ($r = .18$) we performed a Varimax rotation. Three items of the scale (i.e., bitterness, sourness and aroma/smell) showed no clear loadings and were thus excluded. The new model explained 56.00% of variance. The first component explained 38.73% of variance and was interpreted as 'Taste pleasantness' and included item assessing overall liking, overall acceptability, tastiness, sweetness, taste/flavor and texture (factor loadings from .55 to .86). The second component explained 17.27% of variance pertained to the further properties of the drink related to its healthiness and looking (extra-tasting properties). They included naturalness, product looking, healthiness and appearance (factor loadings from .43 to .74). A first one-way

ANOVA showed that despite we observed that the means were in the expected direction, with Shasta judged as less pleasant after its consumption, the effect of the FSR condition was not significant, $F(1,67) = .07$, $p = .791$, $\eta_P^2 < .01$. Similarly, the effect of the FSR on the extra-tasting properties of the drink did not reach the conventional level of significance, $F(1,67) = 2.40$, $p = .126$, $\eta_P^2 = .03$.

**Appendix N – Stimuli used in the tasks of the experimental procedure**

**Study 1**

| SR task | |
| --- | --- |
| Shasta | Faygo |



| Approach/Avoidance IAT | |
| --- | --- |
| Shasta | Faygo |



| Approach | Avoid |
| --- | --- |
| forward, advance, closer, toward, approach | escape, away, withdraw, leave, avoid |

**Explicit measure of liking**

Prefer FAYGO □ □ □ □ □ □ □ □ □ □ Prefer SHASTA

**Thermometer measure**



0  100
Don't like it  Love it

**Study 2**

| SR task | |
| --- | --- |
| Shasta | Faygo |



**Brief IAT**

| Shasta | Faygo |
| --- | --- |



| Slimming | Fattening |
| --- | --- |
| slimming, smaller, lighter, slender | fattening, bigger, heavier, chubby |

## Appendix N - Questionnaires

Product's features evaluation

Please rate the product by <u>circling the appropriate number</u>:

Tastiness

1---------------2---------------3---------------4---------------5---------------6---------------7

Not a lot                                                                         Very much

Naturalness

1---------------2---------------3---------------4---------------5---------------6---------------7

Not a lot                                                                         Very much

Product look (attractiveness)

1---------------2---------------3---------------4---------------5---------------6---------------7

Not a lot                                                                         Very much

Sweetness

1---------------2---------------3---------------4---------------5---------------6---------------7

Not a lot                                                                         Very much

Healthiness

1---------------2---------------3---------------4---------------5---------------6---------------7

Not a lot                                                                         Very much

Sourness

1---------------2---------------3---------------4---------------5---------------6---------------7

Not a lot                                                                         Very much

Bitterness

1---------------2---------------3---------------4---------------5---------------6---------------7

Not a lot                                                                         Very much

Overall liking

1---------------2---------------3---------------4---------------5---------------6---------------7

Not a lot                                                                         Very much

## Sensory Evaluation

Circle one rating in the yellow boxes for each of the following: Appearance, Taste/Flavor, Texture/Consistency, Aroma/Smell, and Overall Acceptability

| Appearance | Extremely Attractive | Moderately Attractive | Attractive | Unappetizing | Unattractive |
|---|---|---|---|---|---|
| | | | | | |
| Taste/Flavor | Tasted great | Flavorful | Acceptable | Off flavor | Flavor did not appeal to me |
| | | | | | |
| Texture Rating | Wonderful texture | Good texture | Acceptable texture | Off texture | Inappropriate texture/flat/runny |
| | | | | | |
| Aroma/ Smell Rating | Wonderful aroma | Appealing aroma | Acceptable aroma | Aroma is not appealing | Unappetizing aroma |
| | | | | | |
| Overall Acceptability | Extremely Acceptable | Moderately Acceptable | Acceptable | Moderately Unacceptable | Unacceptable |
| | | | | | |

# Acknowledgments

These words are dedicated to all the people who will never read any of the previous pages, which is perfectly understandable and is probably why I love you all.

I want to start by thanking my parents for being constantly by my side and for helping me in fixing my troubles, which sometimes is a really tough job. This 'thank you' clearly goes also to my brother, my sister and her lovely children, Carlotta, Francesco and Emanuele.

This PhD was a lot of working and a lot of fun. For the fun part I am about to mention you guys in a chronological order (those I met first come first).

I wish to thank my 'old' friend Roberta, for the chats, the time spent together on a train (or waiting for it) and for many other reasons. Thanks to all my office mates, especially Giulio and Francesco. Outside my office (I mean, officially 'outside'), I want to say 'thank you' to Susanna, for being nice when things went ok and supportive when things went wrong, and to Marco, for the fun I had in listening to his adventures. Thank you guys!

Thanks to all the people I had the chance to meet during my periods abroad.

From Chapel Hill, I want to thank my dear friend Ludovica (best shopping partner ever), my housemate Mike, his lovely daughter Marian and my office mates Aya and Katelyn. A special thanks also goes to the Open Eye Cafè, for being my second house every time I needed some bad American coffee and some good American music (for the record, both Chapter 1 and Chapter 5 were born in there).

From Ghent, thanks to the not-from-Ghent guys. The Italians, Marta1, Marta2 and Valerio, for all the days and the nights spent together and for that trip somewhere in the Netherlands. Thanks to the lovely Irish, above any others, Colin and Ciara. And thanks to the locals, Gaëtan, Maarten, Tom and my two crazy housemates, Helena and Elisabeth.

As this dissertation was mostly about how to change the way people look at things (and other people) by relying on one's self, my last 'thank you' goes to who decided to rely on myself to change her mind about me. Thank you!