UNIVERSITEIT GENT
FACULTEIT PSYCHOLOGIE EN PEDAGOGISCHE WETENSCHAPPEN

Approach-avoidance instructions and training as a method for changing implicit evaluations

Pieter Van Dessel

Promotor: Prof. Dr. Jan De Houwer

Co-promotor: Dr. Anne Gast

Proefschrift ingediend tot het behalen van de academische graad van Doctor in de Psychologie

2016

ACKNOWLEDGEMENTS

I started this PhD project with a lot of enthusiasm, really excited and raring to go. Four years later, my enthusiasm has only grown. This project has been exciting, fascinating, mind-altering, challenging, sometimes scary, but mostly really fulfilling. I would like to thank FWO-Vlaanderen for giving me the opportunity to work on this project. Also, I want to express my gratitude to everyone who contributed in making this project such a rich and enjoyable experience.

First and foremost I want to thank my supervisor, Jan De Houwer. Jan, thank you for all your guidance and support. Thank you for your ever constructive feedback and for all the time and effort you have put into discussing my research, into correcting errors in my manuscripts, into patiently teaching me to be coherent and clear in my writing, and so much more. I generally consider myself a fortunate man, but in this project my biggest fortune was to have been under your supervision.

Secondly I want to thank my co-promotor, Anne Gast. Anne, thank you for our interesting discussions, for helping me with the practical matters when starting my PhD, and for providing your support and feedback on my work and on my, sometimes whacky, ideas even after you returned to your home country.

I also want to thank all the members of my doctoral guidance committee, Prof Dr. Christian Unkelbach, Prof. Dr. Olivier Corneille and Adriaan Spruyt for their time and constructive comments. Also, Adriaan, thanks for your practical help with the design of my experiments and for giving me a pat on the back when I needed it. Additionally, I want to thank all previous and present members of the LIPLAB for their helpful remarks when discussing my work and for the many nice moments. Gaëtan, Jolien, Marijke, a special thank you goes out to you also for making our office a fun place to be and for embarking on many interesting journeys together.

I also owe a great deal of gratitude to Colin Tucker Smith, a collaborator on many of the different chapters of this PhD dissertation. Thank you for your ideas, for our interesting discussions, and for enabling me to conduct my research on an online platform. Also, Maarten, I want to thank you for the statistical help you provided for many of my studies and Arne, thank you for guiding me through the racial prejudice literature.

I would also like to thank Bertram Gawronski, Karl Christoph Klauer, Marco Perugini, Curtis Phills, Kerry Kawakami, Chris Jones, Ayumi Tanaka, and Mandy Hütter for their inspiration and for handing me practical guidelines to perform this research.

Finally, I want to thank my family and friends, for their continuous support. Most importantly, thank you Maya and Kasper, for being the most wonderful children a parent could ever hope for. I am ever so happy and proud of having two children so nice, joyful and kind. And finally, Charlotte I want to thank you for being who you are, ever supportive, inspiring, patient, and loving. The love and support that we have in our family is a beautiful thing and I cherish it with all my heart.

Pieter

February 2016

CONTENTS

CHAPTER 1

GENERAL INTRODUCTION

INTRODUCTION

It has been recognized for many decades that a person's behavior is a function of their likes and dislikes (Allport, 1935). The study of evaluation therefore constitutes an essential part of psychological research. The term "evaluation" can be used to refer to a behavioral phenomenon, that is, the impact of stimuli on evaluative responses (De Houwer, Gawronski, & Barnes-Holmes, 2013). As Zajonc (1980) argued in his seminal paper, evaluations can sometimes arise in a spontaneous manner. For instance, the sight of a stimulus (e.g., a kitten or a chocolate-chip cookie) may automatically evoke a smile. Evaluations that occur under certain conditions of automaticity (e.g., unintentional, uncontrolled, unconscious, or fast; see Moors & De Houwer, 2006) are typically referred to as implicit evaluations whereas evaluations that arise in a more deliberate and controlled manner (e.g., self-reported ratings of liking) are referred to as explicit evaluations.

Previous studies have indicated that indices of implicit evaluations are often better than explicit evaluations at predicting automatic or spontaneous behavior (e.g., Dovidio, Kawakami, & Gaertner, 2002; Friese, Hofmann, & Schmitt, 2009; Perugini, Richetin, & Zogmaister, 2010). Moreover, some research suggests that implicit evaluations are an important determinant of certain spontaneous behaviors (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). It is therefore of particular importance to understand how implicit evaluations are acquired and can be changed.

Interestingly, whereas many studies have investigated how explicit evaluations can be influenced by a variety of different manipulations (see Petty & Wegener, 2010 for an overview), only a limited number of studies have examined the acquisition and change of implicit evaluations, employing only a small number of paradigms (Gawronski & Bodenhausen, 2006). Moreover, most of the paradigms that have been used to change implicit evaluations involve some kind of repeated presentations or training. An important reason for this state of affairs is that most early models that accounted for the impact of evaluations on behavior considered implicit evaluations to result from representations in memory that are highly stable across time and context (see Wilson, Lindsey, & Schooler, 2000 for an example of such a model). The mental representations that underlie implicit evaluations are typically thought to consist of associations links that develop gradually over many experiences. In line with this idea, it is often assumed that implicit stimulus evaluations arise exclusively as the result of repeated pairings of events (Rydell & McConnell, 2006). For instance, evaluative conditioning (EC) research provides ample evidence that changes in the implicit evaluation of a stimulus (conditioned stimulus; CS) occur when it is repeatedly paired with a valenced stimulus (unconditioned stimulus; US; for a review see Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). Moreover, recent studies have shown that changes in implicit evaluations can also be obtained by pairing a stimulus with a valenced action (i.e., approach or avoidance). Typically, the repeated approaching of one stimulus and avoiding of another stimulus leads to more positive implicit evaluations for the former stimuli (e.g., Kawakami, Phills, Steele, & Dovidio, 2007; Woud, Maas, Becker, & Rinck, 2013). It is generally assumed that effects of approach-avoidance (AA) training on implicit evaluations arise as a result of automatic processes (i.e., the automatic formation of mental associations, see for instance Phills, Kawakami, Tabi, Nadolny, & Inzlicht, 2011). However, certain attempts to find AA training effects have failed (see Vandenbosch & De Houwer, 2011) and there is little evidence that informs us about the processes underlying the effects that have been observed.

In the present dissertation, we focus on the learning of implicit evaluation through AA training. We examine the conditions under which AA procedures lead to changes in implicit evaluations with the aim to learn more about the processes underlying these effects and the processes that determine implicit evaluation. In the current chapter, we first provide an overview of current mental-process accounts of implicit evaluation. Second, we describe research on the acquisition and change of implicit evaluation and discuss mental process accounts of AA training effects. Finally, we specify the aims and research questions of the present thesis.

Mental Process Accounts of Implicit Evaluation

As an effect, implicit evaluation can in principle be due to a variety of mental processes. Cognitive theories of evaluation have traditionally assumed that implicit evaluations reflect the automatic activation of associations in memory (for a review, see Hughes, Barnes-Holmes, & De Houwer, 2011). Associations can be defined as simple links between mental representations that form whenever two representations simultaneously become active (Shanks, 2007). One can conceptualize an *evaluative association* as an association between a representation of a target concept (e.g., a stimulus) and a representation of positive or negative valence. Implicit evaluation can then be explained in the following manner: Upon presentation of a certain stimulus, activation of its corresponding representation in memory will spread to a valenced representation through acquired evaluative associations. Once this valenced representation becomes active, it may influence the evaluative response to the stimulus.

Broadly speaking, one can distinguish two different classes of evaluation models that have adopted the assumption that associations underlie implicit evaluation. First, single-process associative models postulate 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 (i.e., associations that have built up sufficient strength to operate automatically). In contrast, explicit evaluations may also depend on the more effortful activation of other associations (when a person has the opportunity and motivation to consider additional information; Fazio &

Olson, 2014).

Second, in contrast to those purely associative models, dual-process models propose that two separate processes are responsible for evaluation. These models are based on dual-system models which assume that people process information through two qualitatively distinct cognitive systems (e.g., Sloman, 1996, Smith & DeCoster, 2000, Strack & Deutsch, 2004). One system is slow-learning and operates through automatic association formation processes whereas the other system is fast-learning and entails the rule-based formation of propositional representations. Contrary to associations, propositions are structured representations that include information not only about the strength of the relationship between concepts but also about the nature of the 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). Consistent with this idea, some studies have found that explicit evaluations are particularly sensitive to the acquisition of propositional information via instructions, whereas implicit evaluations are more sensitive to repeated evaluative pairings (Rydell, McConnell, Mackie, & Strain, 2006; Rydell &McConnell, 2006).

However, some contemporary dual-process models have argued against a strict separation of associative and propositional processes in evaluation. These models assume that associative processes interact with propositional processes to explain implicit and explicit evaluation. The currently most extensive theoretical account of how these two distinct processes interact and cause changes in implicit and explicit evaluations in different contexts is the Associative-Propositional Evaluation (APE) model (Gawronski & Bodenhausen, 2006, 2011, 2014). The APE model posits that implicit evaluations can form not only as the result of a pairing between events (i.e., associative learning, see De Houwer, Barnes-Holmes, & Moors, 2013) but also as the result of the acquisition of propositional information (i.e., propositional learning). More specifically, implicit evaluations may form when externally provided or internally generated propositions are considered valid, thereby creating new (strong) associations in memory (see Whitfield & Jordan, 2009). Several recent studies seem to support this idea. First, it has been shown that changes in implicit evaluations can occur when participants are provided with a single instance of new propositional information via instructions (Castelli, Zogmaister, Smith, & Arcuri, 2003; De Houwer, 2006; Gregg, Seibt, & Banaji, 2006). For instance, when participants are informed that the members of one fictitious social group have positive traits and the members of another fictitious social group have negative traits, an implicit preference for the former group is typically observed (Gregg et al.). Second, some studies recently provided evidence that instruction effects on implicit evaluation depend on qualities of the information other than the (one-time) pairing of stimulus and valenced event that are part of the instructions (e.g., the instruction “Bob is good” entails a paring of “Bob” and “good”), such as the perceived validity of the acquired information (Cone & Ferguson, 2015; Peters & Gawronski, 2011; Raes, De Houwer, De Schryver, Brass, & Kalisch, 2014). Because these findings indicate that propositional processes play an important role in the learning of implicit evaluations, they pose a challenge to associative and dual-process models of evaluation which assume that implicit evaluations result only from the gradual formation of associations in memory as the result of actual pairings (Rydell & McConnell, 2006; Smith & DeCoster, 2000).

Recently, a third class of evaluation models has emerged that can also explain the effects of instructions on implicit evaluation. These models do not adopt the assumption that implicit evaluation reflects association activation. Rather, they argue that implicit and explicit evaluation is determined entirely by propositional processes (De Houwer, 2009, 2014; Mandelbaum, 2015; Mitchell, De Houwer, & Lovibond, 2009). Proponents of these single-process propositional models believe that evaluations depend on the formation and activation of propositions. Whereas explicit evaluations may reflect the subjective validity of activated propositions, implicit evaluation may reflect the automatic activation of propositions that are easily retrieved (De Houwer, 2014). For instance, the sight of a cookie (i.e., an evaluative stimulus) could allow for the retrieval of memories in which the proposition “cookies are tasty” was encoded. The activation of this proposition may automatically evoke a positive evaluative response such as a smile (i.e., implicit evaluation). A person’s deliberate response to the question whether they like the cookie (i.e., explicit evaluation) may more strongly reflect whether participants also consider this proposition to be valid and/or the importance of this belief in comparison to other beliefs about cookies (e.g., “cookies are unhealthy”).

Associative learning of Implicit Evaluations

Learning that involves a regularity in the presence of two events can be defined as associative learning (De Houwer et al., 2013). Traditionally, it was assumed that all instances of associative learning result from the automatic formation of mental associations (see Mitchell et al., 2009). A popular assumption of association formation theories of learning is that observed spatio-temporal contiguities (or co-occurrences) of events lead to a co-activation of their corresponding mental representations which automatically creates an associative link between the two representations (Shanks, 2007). Repeatedly observing the same co-occurrences strengthens this link, which facilitates the spread of activation from one concept to the other when one of the two events occurs. Given that most evaluation models assume that mental associations underlie implicit evaluations, it is not surprising that most research on the acquisition and change of implicit evaluations has employed paradigms that involve repeated co-occurrences.

The most extensively used procedure to induce changes in implicit evaluations is the EC procedure. EC refers to the change in liking of a stimulus (CS) that results from the pairing of that stimulus with a positive or negative stimulus (US; De Houwer, 2007). In line with the idea that automatic association formation mechanisms underlie these effects (Baeyens, Eelen, Crombez, & Van den Bergh, 1992), several studies have provided evidence that EC can occur under certain conditions of automaticity. For instance, some research suggests that EC is not impeded by attentional load (Fulcher & Hammerl, 2001), is not susceptible to intentional control (Gawronski, Balas, & Creighton, 2014; Gawronski, Mitchell, & Balas, 2015) and can occur in the absence of conscious awareness of the contingency between CS and US (Baeyens, Eelen, Crombez, & Van den Bergh, 1992; Hütter, Sweldens, Stahl, Unkelbach, & Klauer, 2012; Olson & Fazio, 2001).

However, recent studies have provided evidence that the processes underlying EC may be less automatic than is often assumed. For instance, some research has shown that EC depends on the availability of cognitive resources (Davies, El-Deredy, Zandstra, & Blanchette, 2012; Pleyers, Corneille, Yzerbyt, & Luminet, 2009), and momentary processing goals (Corneille, Yzerbyt, Pleyers, & Mussweiler, 2009; Gast & Rothermund, 2011). Moreover, recent evidence has challenged the idea that EC can occur in the absence of contingency awareness and suggested that EC occurs only after participants become aware of CS-US contingencies (e.g., Bar-Anan, De Houwer, & Nosek, 2010; Pleyers, Corneille, Luminet, & Yzerbyt, 2007; Stahl, Unkelbach, & Corneille, 2009). Though the question whether EC can occur without contingency awareness is still strongly debated (see Sweldens, Corneille, & Yzerbyt, 2014 for an overview), there is now general consensus that contingency awareness is at least an important moderator of EC (Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010).

Evidence that EC, at least partly, depends on controlled processes, has important theoretical implications. It implies that (stimulus) pairings do not (always) automatically produce changes in implicit evaluations and therefore contrasts with the idea that EC is (exclusively) the result of the automatic formation of associations. In accordance with propositional theories of evaluation, it has recently been suggested that EC results entirely from propositional processes (De Houwer, 2009; Hughes et al., 2011; Mitchell, De Houwer, & Lovibond, 2009). According to these propositional accounts, the acquisition of propositional knowledge about the relation between the CS and US causes changes in implicit and explicit evaluation. Supporting this idea, some recent studies have shown that changes in the implicit evaluation of a CS occur when verbal instructions link a CS with a valenced US in the absence of actual CS-US pairings (De Houwer, 2006; Gast & De Houwer, 2012). For instance, when participants are instructed that one nonword will always be followed by positive pictures and another nonword will be followed by negative pictures, they exhibit more positive implicit evaluations of the former nonword than of the latter nonword (De Houwer).

AA Training Effects on Implicit Evaluation

Recent research has shown that implicit evaluations can form also as the result of the repeated pairing of stimuli and approach or avoidance actions. The first demonstration of this effect was provided by Kawakami et al. (2007). They found that participants who repeatedly approached photographs of Black people and avoided photographs of White people exhibited more positive evaluations of Black relative to White people on the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). AA training effects have been observed for a variety of stimuli, such as pictures of unfamiliar faces (Woud, Maas, Becker, & Rinck, 2013), alcoholic drinks (Wiers, Eberl, Rinck, Becker, & Lindenmeyer, 2011), unhealthy foods (Zogmaister, Perugini, & Richetin, in press), insects and spiders (Jones, Vilensky, Vasey, & Fazio, 2013), or contamination-related objects (Amir, Kuckertz, & Najmi, 2013).

AA training effects resemble EC in that a change in liking occurs as the result of a contiguous relation between a neutral stimulus and a valenced event. In AA studies the valenced event corresponds to the execution of a valenced action (i.e., an approach or avoidance action) rather than the presentation of a valenced US. Approach and avoidance are typically considered primitive behavioral tendencies that must have come into place very early on in evolution. As a result, these actions are often assumed to be tightly linked with an impulsive, associative system (Strack & Deutsch, 2004). The processes underlying AA training effects might thus differ from those involved in EC in that they provide a more low-level route to changing stimulus evaluations. That is, AA training might change implicit evaluations in a way that circumvents propositional processes and relies entirely on automatic association formation processes. In line with this idea, Kawakami and colleagues (2007) observed changes in implicit evaluations when participants performed AA training under conditions that seriously limited conscious detection of the approached and avoided stimuli (i.e., presentation in between two masking stimuli and for a duration of 23 ms). The authors concluded that “these effects can occur in a largely automatic fashion, outside of participants’ awareness and beyond their conscious control” (p. 968).

Mental Process Accounts of AA Training Effects

Though the investigation of AA training effects is a recent phenomenon, a number of explanations for these effects have been described. First, motivational accounts suggest that motivational systems of approach and avoidance mediate the relation between AA behavior and stimulus evaluations (Cacioppo, Priester, & Berntson, 1993). Theorists have argued that behavior is driven by two distinct motivational circuits that direct the deployment of primitive approach and withdrawal behavior (Elliot, 2006; Lang, Bradley, & Cuthbert, 1990). Because AA actions are wired into these motivational systems, performing these actions may lead to the activation of these motivational systems which, in turn, may bias the evaluative processing of stimuli. Specifically, the processing of positive qualities of a stimulus may be facilitated when the approach system is triggered and the processing of negative qualities may be facilitated when the avoidance system is triggered (Neumann & Strack, 2000). Support for this account has primarily come from studies that used a motivation induction procedure such as the enduring performance of a specific approach behavior (e.g., arm flexion) or avoidance behavior (e.g., arm extension) during stimulus evaluation. For instance, Cacioppo and colleagues demonstrated that stimuli are evaluated more positively during arm flexion (established by pressing upwards on a table) than during arm extension (established by pressing downward on a table). Note that, in contrast with typical AA training studies, these studies involved the evaluation of stimuli during the long-during performance of specific AA actions (i.e., the performance of arm movements that were assumed to be intrinsically an approach action or an avoidance action, but see Eder & Rothermund, 2008). Though AA training effects result from repeated pairings of AA actions and stimuli, they can also be accounted for from a motivational perspective. For instance, each time a stimulus is approached or avoided, this may lead to the co-activation of the corresponding stimulus representation and approach or avoidance motivational system. Associations may form such that the subsequent presentation of an approached stimulus will now trigger the approach system and the presentation of an avoided stimulus will trigger the avoidance system, which may facilitate the automatic evaluation of the stimulus as positive or negative (Neumann, Förster, & Strack, 2003).

Second, according to the associative self-anchoring account of AA training effects, approach behaviors are fundamentally related to pulling objects closer to the self (Förster, 2001). As a result, the repeated performance of approach behavior in response to a stimulus may allow for the gradual formation of an association between positively valenced representations of the self and the to-be-approached stimulus (Phills, Kawakami, Tabi, Nadolny, & Inzlicht, 2011). Once these associations are established, positive valence may spread to the to-be-approached stimulus, thereby influencing implicit evaluations of this stimulus (Gawronski, Bodenhausen, & Becker, 2007).

Third, some have argued that representations of AA behaviors have become associated with representations of positive or negative valence because they usually lead to positive or negative outcomes (Woud, Becker, & Rinck, 2008; Woud, Maas, et al., 2013). The repeated performance of AA behavior in response to a stimulus will lead to co-activation of the positively or negatively valenced action representations and representations of the stimulus. As a result, associations between these representations are formed and these associations will influence implicit evaluation. This explanation accords with typical associative explanations of EC to the extent that the repeated presentation of a stimulus and a valenced event allows for the gradual formation of evaluative associations in memory which drive implicit evaluation (Baeyens et al., 1992).

Finally, Eder and Klauer (2009) described a common-coding account of AA training effects. This account builds on the assumption that sensory events and action events are represented in memory by means of structurally identical event codes (Hommel, 2004). During the performance of valenced actions, codes of positive or negative valence may be activated. These codes are bound to stimulus representations when they are activated at the same time. As a result, presentation of an approached or avoided stimulus may lead to the activation of valenced codes and evoke an evaluative response in line with the valence of the performed action. Similar to the account described by Woud and colleagues the common-coding account thus explains AA training effects on the basis of the automatic activation of valenced representations during the performance of valenced actions and the formation of an association between these representation and stimulus representations.

The other described accounts differ in a number of important ways. For instance, whereas the account proposed by Woud and colleagues and the common-coding account both assume that the repeated performance of any valenced action should produce similar effects, motivational accounts assume that effects are restricted to approach and avoidance actions. The self-anchoring account, on the other hand, assumes that only the approach actions determine AA training effects. Importantly, all described accounts ascribe to one key assumption, that automatic association formation processes underlie AA training effects.

The Associative Assumption In Accounts of AA Training Effects

The assumption that AA training effects arise exclusively as a result of the automatic formation of associations in memory is widely present in current theorizing but has never been explicitly tested. Moreover, some findings seem to provide evidence that contrasts with this assumption. First, several studies have failed to show effects of AA training on evaluations of well-known stimuli (Becker, Jostmann, Wiers, & Holland, 2015) or even novel stimuli (Vandenbosch & De Houwer, 2011). Other studies have found that AA training effects were restricted to certain stimuli (e.g., for face stimuli with neutral but not angry or smiling expressions, Woud, Becker, Lange, & Rinck, 2013) or found effects on explicit but not implicit evaluation (Huijding, Muris, Lester, Field, & Joosse, 2011). These findings contrast with the idea that AA training effects depend on the automatic installation of associations. Rather, AA training effects seem to depend on subtle boundary conditions that yet have to be specified. For instance, whereas Woud et al. (2008) reported that participants who repeatedly performed AA movements in response to pictures of faces with neutral emotional expressions exhibited an implicit preference for approached faces, Vandenbosch and De Houwer (2011) failed to find any evidence for AA training effects on evaluations of novel faces in five experiments with similar methods.

Second, Laham, Kashima, Dix, Wheeler, and Levis (2014) recently provided evidence that AA training effects depend on the motivational framing of AA actions. When participants repeatedly performed approach or avoidance movements in response to novel stimuli, this led to the development of implicit evaluations when the actions were contextually framed as collect and discard actions within a foraging context. Importantly, no AA training effects were observed when contextual framing of AA actions was not elaborated. This suggests that it is not the pairing of AA action and stimulus per se that produces the effects. Rather, AA training effects may critically depend on certain moderators that indicate the involvement of high-level controlled processes. Importantly, however, this question has received little or no attention in AA training research.

THE CURRENT PHD THESIS

In the present project, we aim to learn more about the conditions under which AA training causes changes in implicit evaluations. This question is relevant for a number of reasons. First, as we previously described, it is often assumed that AA training might provide a special, low-level route to changing implicit evaluations that circumvents propositional processes and relies entirely on automatic associative processes. The idea that repeatedly approaching or avoiding stimuli changes implicit evaluations, fits well with the commonly held assumption that associations can be altered only gradually as the result of many different experiences. However, it has never been tested whether propositional processes play a role in these effects. By studying the effects of AA training on implicit evaluations, we could obtain unique information about the scope of propositional processes or the nature of associative processes. Second, at present, only a handful of studies inform us about whether propositional processes can directly influence implicit evaluations, that is, without the involvement of associative processes (Hughes et al., 2011). By examining the role of propositional and associative processes in the formation of implicit evaluations via AA training, we may learn more about the processes underlying implicit evaluation in general. Third, because implicit evaluations have such a profound impact on behavior, this new information could have important practical implications. For example, studies have already provided evidence that AA training can be effective in reducing unwanted behaviors such as racial prejudice (Kawakami et al., 2007) or can be used in the treatment of pathological conditions such as social anxiety (Taylor & Amir, 2012), contamination anxiety (Amir et al., 2013), alcohol-dependence (Wiers, Gladwin, & Rinck, 2013), or smoking addiction (Wittekind, Feist, Schneider, Moritz, & Fritzsche, 2015).

In the present project, we focus on two specific questions about when and how AA training changes implicit evaluations. First, we examine if and under what circumstances AA instructions can influence implicit (and explicit) evaluation. Second, we examine the role of contingency awareness in the effects of AA training.

Effects of AA instructions

The first question we address is whether the acquisition of propositional knowledge about stimulus-response contingencies via instructions (i.e., about which stimulus is approached and which stimulus is avoided) can lead to typical AA training effects. To the extent that propositional processes underlie AA training effects, the acquisition of propositional contingency knowledge may be crucial. For instance, participants who acquire stimulus-action contingency information may elaborate on this information and infer that the approached stimulus is positive and the avoided stimulus is negative. Such inference could arise because of the fact that often during their lifetime, good things were approached and bad things were avoided. Once participants make this inference and form the proposition that the approached stimulus is positive (or that the avoided stimulus is negative), this may influence implicit stimulus evaluation, either directly (see De Houwer, 2014) or via its influence on explicit evaluations (see Gawronski & Bodenhausen, 2006).

It is typically assumed that propositions about events can be formed not only on the basis of the repeated experience of those events (e.g., the pairing of AA action and stimulus) but also as the result of a single instruction or inference concerning those events (see De Houwer, 2009; Lagnado et al., 2007; Mitchell et al., 2009). Therefore, effects that result from the acquisition of propositional contingency information should occur also via mere instructions (see Gast & De Houwer, 2012 for a similar reasoning in the context of EC instruction effects). By examining to what extent instructions that specify stimulus-action contingencies (i.e., AA instructions) may also cause effects on stimulus evaluations we may gain important information about the role of propositional processes in AA effects.

The role of contingency awareness in AA training effects

The second question we address is whether conscious propositional knowledge about stimulus-response contingencies moderates AA training effects. Given the important role of contingency awareness for EC and the strong resemblance between EC and AA training procedures, contingency awareness may be an important factor also for AA training effects. Laham and colleagues (2014) recently alluded to this possibility when they suggested that AA training effects may depend on the motivational framing of AA actions because elaborated framing instructions increase the likelihood that participants become aware of the stimulus-action contingencies.

The question of (contingency) awareness in AA training effects, however, has already been addressed in studies that examined whether subliminal AA training causes changes in evaluations (Jones et al., 2013; Kawakami et al., 2007). The results of these studies provided evidence that AA training effects can occur outside of participants' awareness of the contingencies in the training task. This is an important conclusion because it implies that AA training effects depend on processes that do not necessitate awareness (e.g., automatic association formation processes). Moreover, such unambiguous evidence in favor of unconscious associative learning is scarce (see Mitchell et al., 2009). However, given that only few studies have investigated this matter, more research is warranted to establish if and under what circumstances AA training causes changes in evaluations in the absence of conscious knowledge of stimulus-action contingencies. Also, even if contingency awareness might not prove to be necessary for AA training effects, it may still be an important moderator. By examining the role of contingency awareness in AA training effects we aim to learn more about the propositional or associative processes that underlie these effects.

Overview of the empirical chapters

The current thesis contains four empirical chapters, each of which aims at shedding light on the conditions under which AA training and AA instructions can produce changes in implicit evaluations. In the first two chapters we focused on the question whether implicit evaluations can change as the result of AA instructions. The next two chapters investigated to what extent AA training effects are moderated by awareness of stimulus-action contingencies.

In **Chapter 2**, we examined whether merely instructing participants to approach a certain stimulus and to avoid a second stimulus could influence implicit (and explicit) evaluations of these stimuli. First, we examined whether AA instructions cause changes in implicit evaluations of *novel* stimuli. In Experiments 1 and 2, we informed participants that they would perform a task in which they would encounter specific stimuli (novel non-words: Experiment 1, or names of members of two fictitious social groups: Experiment 2). We instructed participants to approach one stimulus (e.g., the nonword UDIBNON) by pulling a lever towards them and to avoid another stimulus (e.g., the nonword BAYRAM) by pushing a joystick away from them. After receiving these instructions, participants performed an IAT that was designed to measure implicit evaluations of the stimuli they were instructed to approach and the stimuli they were instructed to avoid. Experiment 2 also examined effects of AA instructions on explicit stimulus evaluations (i.e., self-reported ratings of warmth and liking). In a second set of studies we examined whether AA instructions cause changes in implicit (and explicit) evaluations of well-known stimuli. More specifically, Experiment 3 investigated AA instruction effects on the relative evaluation of a social group that participants belonged to (i.e., Flemish people) and a social group that Flemish people are prejudiced towards (i.e., Turkish people; see Dhont & Van Hiel, 2009). Experiment 4 examined AA instruction effects on White people's evaluations of White people and Black people, in accordance with the training-based AA studies by Kawakami et al. (2007). This experiment also compared AA instruction effects on evaluations of these well-known social groups with effects on evaluations of fictitious social groups. Finally, Experiment 5 was a replication of Experiment 4 with the exception that a different task was used to measure implicit evaluations (i.e., the evaluative priming task, Fazio, Sanbonmatsu, Powell, & Kardes, 1986).

Chapter 3 examined whether effects of AA instructions and AA training on implicit evaluations are mediated by changes in explicit evaluations. This is an important question because it may provide essential information about the processes underlying these effects and may constrain mental process models of evaluation. For instance, according to dual-process models such as the APE model, changes in implicit evaluation as the result of instructions (but not training) should necessarily be mediated by changes in explicit evaluation (Gawronski & Bodenhausen, 2006). We used two different approaches to examine whether instructions to approach or avoid members of fictitious social groups and actual AA training can cause a direct effect on implicit evaluation (i.e., unmediated by changes in explicit evaluation). First, we employed statistical mediation analyses to test the extent to which the impact of AA instruction and AA training on implicit evaluation is mediated by changes in explicit evaluation. Second, we took an experimental approach to test the question of mediation by manipulating the proposed mediating variable (i.e., changes in explicit evaluation). To this end, we provided participants with ‘trait instructions’ that should prevent an impact of AA instructions on explicit evaluation. Specifically, we asked participants to imagine that the members of one fictitious social group had very positive traits and the members of another fictitious social group had very negative traits. We expected that participants who received trait instructions would not take the AA instructions into account when explicitly evaluating the stimuli. We examined whether, under these circumstances, AA instructions and AA training would still cause changes in implicit evaluation.

Chapter 4 investigated the relationship between, on the one hand, effects of AA training and, on the other hand, contingency awareness. This study employed generally the same procedures as the studies by Woud et al. (2008) and Vandenbosch and De Houwer (2011) that provided mixed evidence for effects of AA training on evaluations of novel face stimuli. In these studies, participants did not receive any information about the stimulus-action contingencies. Rather, faces had a subtle brown or red filter placed over them and participants were instructed to approach or avoid on the basis of the color of the presented face. Unbeknownst to the participants, some stimuli were always presented in the to-be-approached color whereas other stimuli were always presented in the to-be-avoided color. We examined whether the effects of AA training on implicit and explicit evaluations of novel faces that are produced in this paradigm depend on participants’ awareness of the contingency between face stimulus and AA action. We took both a correlational and an experimental approach to address the role of contingency awareness in AA training effects. First, we tried to capture participants’ awareness of the experienced face-action contingencies by measuring participants’ memory of the relation between faces and actions. We compared AA training effects for faces that were correctly linked to the action they were paired with, faces that were linked to the incorrect action, and faces for which participants did not remember the correct action. Second, we manipulated contingency awareness by providing one group of participants with instructions that specified the face-action contingencies whereas a second group did not receive these instructions. We examined whether this manipulation caused changes in contingency awareness and whether these changes affected AA training effects on implicit and explicit evaluation.

Finally, **in Chapter 5**, we examined whether reliable AA training effects can be observed with subliminal stimulus presentations. Experiment 1 was a direct replication of the experiment by Kawakami et al. (2007, Exp. 2) that found effects of subliminal AA training on implicit evaluations of well-known social groups. In Experiment 2, participants received subliminal training to approach or avoid novel faces. Before the training, however, participants received either correct or incorrect face-action contingency information (i.e., information about which face stimulus would be presented with which action during the subliminal AA training task). This experimental set-up allowed us to compare effects that are the result of subliminal AA training with effects that result from AA instructions. In Experiment 3, participants performed an AA training task with supraliminal stimulus presentations of two novel non-existing words prior to the performance of an AA training task with subliminal presentations of two other non-existing words. We compared effects of subliminal and supraliminal AA training and examined whether participants' awareness of the contingencies fuelled the effects.

REFERENCES

Allport, G. (1935). Attitudes. In Murchison, C. (Ed.). *A Handbook of Social Psychology* (pp. 789-843). Worcester, MA: Clark University Press.

Amir, N., Kuckertz, J. M., & Najmi, S. (2013). The effect of modifying automatic action tendencies on overt avoidance behaviors. *Emotion*, *13*, 478–484. doi: 10.1037/a003044

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*, 133–142. doi: 10.1016/0005-7967

Bar-Anan, Y., De Houwer, J., & Nosek, B. (2010). Evaluative conditioning and conscious knowledge of contingencies: a correlational investigation with large samples. *Quarterly Journal of Experimental Psychology*, *63*, 2313–2335. doi: 10.1080/17470211003802442

Becker, D., Jostmann, N. B., Wiers, R. W., & Holland, R. W. (2015). Approach avoidance training in the eating domain: Testing the effectiveness across three single session studies. *Appetite*, *85*, 58–65. doi: 10.1016/j.appet.2014.11.017.

Cacioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary determinants of attitudes II: Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology*, *65*, 5–17. doi: 10.1037/0022-3514.65.1.5

Castelli, L., Zogmaister, C., Smith, E. R., & Arcuri, L. (2004). On the automatic evaluation of social exemplars. *Journal of Personality and Social Psychology*, *86*, 373–387. doi: 10.1037/0022-3514.86.3.373

Cone, J., & Ferguson, M. J. (2015). He Did What ? The Role of Diagnosticity in Revising Implicit Evaluations. *Journal of Personality and Social Psychology*, *108*, 37–57. doi: 10.1037/pspa0000014

Corneille, O., Yzerbyt, V. Y., Pleyers, G., & Mussweiler, T. (2009). Beyond awareness and resources: Evaluative conditioning may be sensitive to processing goals. *Journal of Experimental Social Psychology*, *45*, 279–282. doi: 10.1016/j.jesp.2008.08.020

Davies, S. R., El-Deredy, W., Zandstra, E. H., & Blanchette, I. (2012). Evidence for the role of cognitive resources in flavour–flavour evaluative conditioning. *The Quarterly Journal of Experimental Psychology, 65*, 2297-2308. doi: 10.1080/17470218.2012.701311

De Houwer, J. (2006). Using the Implicit Association Test does not rule out an impact of conscious propositional knowledge on evaluative conditioning. *Learning and Motivation*, *37*, 176–187. doi: 10.1016/j.lmot.2005.12.002

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, *10*, 230–241. doi: S1138741600006491

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

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

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*, 631-642. doi: 10.3758/s13423-013-0386-3

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology*, *24*, 252–287. doi: 10.1080/10463283.2014.892320

Dhont, K., & Van Hiel, A. (2009). We must not be enemies: Interracial contact and the reduction of prejudice among authoritarians. *Personality and Individual Differences*, *46*, 172–177. doi: 10.1016/j.paid.2008.09.022

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*, 62–68. doi: 10.1037/0022-3514.82.1.62

Eder, A. B., & Klauer, K. C. (2009). A common-coding account of the bidirectional evaluation–behavior link. *Journal of Experimental Psychology: General*, *138*, 218–235. doi: 10.1037/a0015220

Eder, A. B., & Rothermund, K. (2008). When do motor behaviors (mis)match affective stimuli? An evaluative coding view of approach and avoidance reactions. *Journal of Experimental Psychology: General*, *137*, 262–281. doi: 10.1037/0096-3445.137.2.262

Elliot, A. J. (2006). The hierarchical model of approach-avoidance motivation. *Motivation and Emotion, 30*, 111–116. doi: 10.1007/s11031-006-902807

Fazio, R. H. (1990). Multiple Processes by which Attitudes Guide Behavior: The MODE Model as an Integrative Framework. *Advances in Experimental Social Psychology*, *23*, 75–109. doi: 10.1016/S0065-2601(08)60318-4

Fazio, R. H. (2007). Attitudes as Object–Evaluation Associations of Varying Strength. *Social Cognition, 25,* 603−637. doi: 10.1521/soco.2007.25.5.603

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition. research: their meaning and use. *Annual Review of Psychology*, *54*, 297–327. doi: 10.1146/ annurev.psych.54.101601.145225

Fazio, R. H., & Olson, M. A. (2014). The MODE model: Attitude-behavior processes as a function of motivation and opportunity. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.). Dual process theories of the social mind. New York: Guilford Press.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*, 229–238. doi: 10.1037/0022-3514.50.2.229.

Förster, J. (2001). Success/Failure Feedback, Expectancies, and Approach/Avoidance Motivation: How Regulatory Focus Moderates Classic Relations. *Journal of Experimental Social Psychology*, *37*, 253–260. doi: 10.1006/jesp.2000.1455

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*, 285-338. doi: 10.1080/10463280802556958

Fulcher, E. P., & Hammerl, M. (2001). When all is revealed: a dissociation between evaluative learning and contingency awareness. *Consciousness and Cognition*, *10*, 524–549. doi: 10.1006/ccog.2001.0525

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*, 1657–1674. doi: 10.1080/17470218.2012.665061

Gast, A., & Rothermund, K. (2011). What you see is what will change: evaluative conditioning effects depend on a focus on valence. *Cognition & Emotion*, *25*, 89–110. doi: 10.1080/02699931003696380

Gawronski, B., Balas, R., & Creighton, L. (2014). Can the formation of conditioned attitudes be intentionally controlled? *Personality & Social Psychology Bulletin*, *40*, 419–432. doi: 10.1177/0146167213513907

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692–731. doi: 10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model. Theory, evidence, and open questions. Advances in Experimental Social Psychology, 44, 59-127. doi: 10.1016/B978-0-12-385522-0.00002-0

Gawronski, B., & Bodenhausen, G. V. (2014). Implicit and explicit evaluation: A brief review of the associative-propositional evaluation model. *Social and Personality Psychology Compass*, *8*, 448–462. doi: 10.1111/spc3.v8.8

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*, 221–232. doi: 10.1016/j.jesp.2006.04.001

Gawronski, B., Mitchell, D. G. V, & Balas, R. (2015). Is Evaluative Conditioning Really Uncontrollable? A Comparative Test of Three Emotion-Focused Strategies to Prevent the Acquisition of Conditioned Preferences. *Emotion*, *15*, 556-568. doi: 10.1037/emo0000078

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*, 1464–80. doi: 10.1037/0022-3514.74.6.1464

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*, 17–41. doi: 10.1037/a0015575

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*, 1–20. doi: 10.1037/0022-3514.90.1.1

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

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences*, *8*, 494–500. doi: 10.1016/j.tics.2004.08.007

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). the Dominance of Associative Theorizing in Implicit Attitude Research: Propositional and Behavioral Alternatives. *Psychological Record*, *61*, 465–496.

Huijding, J., Muris, P., Lester, K. J., Field, A. P., & Joosse, G. (2011). Training children to approach or avoid novel animals: Effects on self-reported attitudes and fear beliefs and information-seeking behaviors. *Behaviour Research and Therapy*, *49*, 606–613. doi: 10.1016/j.brat.2011.06.005

Hütter, M., Sweldens, S., Stahl, C., Unkelbach, C., & Klauer, K. C. (2012). Dissociating contingency awareness and conditioned attitudes: Evidence of contingency-unaware evaluative conditioning. *Journal of Experimental Psychology: General*, *141*, 539–557. doi: 10.1037/a0026477

Jones, C. R., Vilensky, M. R., Vasey, M. W., & Fazio, R. H. (2013). Approach behavior can mitigate predominately univalent negative attitudes: evidence regarding insects and spiders. *Emotion*, *13*5, 989–996. doi: 10.1016/S0896-6273(02)00858-9

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder:Improving implicit racial attitudes and interracial interactions through approach behaviors. *Journal of Personality and Social Psychology*, *92*, 957–971. doi: 10.1037/0022-3514.92.6.957

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (2007). Beyond covariation: Cues to causal structure. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 154–172). Oxford University Press.

Laham, S. M., Kashima, Y., Dix, J., Wheeler, M., & Levis, B. (2014). Elaborated contextual framing is necessary for action-based attitude acquisition. *Cognition & Emotion, 28*, 1119-1126. doi: 10.1080/02699931.2013.867833

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, Attention, and the Startle Reflex. *Psychological Review*, *97*, 377–395. doi: 10.1037/0033-295X.97.3.377

Mandelbaum, E. (2015). Attitude, Inference, Association: On the Propositional Structure of Implicit Bias. *Noûs*. doi: 10.1111/nous.12089

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *The Behavioral and Brain Sciences*, *32*, 183–198. doi: 10.1017/s0140525x09000855

Moors, A., & De Houwer, J. (2006). Automaticity: A Theoretical and Conceptual Analysis. *Psychological Bulletin*, *132*, 297–326. doi: 10.1037/0033-2909.132.2.297

Neumann, R., Förster, J., & Strack, F. (2003). Motor compatibility: The bidirectional link between behavior and evaluation. In J. Musch & K.C. Klauer (Eds.), *The psychology of evaluation* (pp. 371–391). London: Erlbaum.

Neumann, R., & Strack, F. (2000). Approach and avoidance: the influence of proprioceptive and exteroceptive cues on encoding of affective information. *Journal of Personality and Social Psychology*, *79*, 39–48. doi: 10.1037/0022-3514.79.1.39

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*, 413–417. doi: 10.1111/1467-9280.00376

Perugini, M., Richetin, J., & Zogmaister, C. (2010). Prediction of behavior. In Gawronski, B., & Payne, B. K. (Eds.), *Handbook of implicit social cognition* (pp. 255–278). New York, NY: Guilford Press.

Peters, K. R., & Gawronski, B. (2011). Are we puppets on a string? Comparing the impact of contingency and validity on implicit and explicit evaluations. *Personality and Social Psychology Bulletin*, *37*, 557–569. doi: 10.1177/0146167211400423

Petty, R. E., & Wegener, D. T. (2010). Attitude change: Multiple roles for persuasion variables. In R. F. Baumeister & E. J. Finkel (Eds.), *Advanced social psychology: The state of the science* (pp. 217–259). Oxford: Oxford University.

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. doi: 10.1037/a0022159

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*, 130–144. doi: 10.1037/0278-7393.33.1.130

Pleyers, G., Corneille, O., Yzerbyt, V., & Luminet, O. (2009). Evaluative conditioning may incur attentional costs. *Journal of Experimental Psychology. Animal Behavior Processes*, *35*, 279–285. doi: 10.1037/a0013429

Raes, A. K., De Houwer, J., De Schryver, M., Brass, M., & Kalisch, R. (2014). Do CS-US pairings actually matter? A within-subject comparison of instructed fear conditioning with and without actual CS-US pairings. *PLoS ONE, 9*(1): e84888. doi:10.1371/journal.pone.0084888

Roefs, A., Huijding, J., Smulders, F. T. Y., MacLeod, C. M., de Jong, P. J., Wiers, R. W., & Jansen, A. T. M. (2011). Implicit measures of association in psychopathology research. *Psychological Bulletin*, *137*, 149–193. doi: 10.1037/a0021729

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*, 995–1008. doi: 10.1037/0022-3514.91.6.995

Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of Two Minds: Forming and Changing Valence-Inconsistent Implicit and Explicit Attitudes. *Psychological Science*, *17*, 954–958. doi: 10.1111/j.1467-9280.2006.01811.x

Shanks, D. R. (2007). Associationism and cognition: human contingency learning at 25. *Quarterly Journal of Experimental Psychology*, *60*, 291–309. doi: 10.1080/17470210601000581

Sloman, S. a. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22. doi: 10.1.1.130.7987

Smith, E. R., & DeCoster, J. (2000). Dual-Process Models in Social and Cognitive Psychology: Conceptual Integration and Links to Underlying Memory Systems. *Personality and Social Psychology Review*, *4*, 108–131. doi: 10.1207/S15327957PSPR0402_01

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*, 404–420. doi: 10.1037/a0016196

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*, 220-247. doi: 10.1207/s15327957pspr0803_1

Sweldens, S., Corneille, O. & Yzerbyt, V. (2014). The Role of Awareness in Attitude formation Through Evaluative Conditioning. *Personality and Social Psychology Review, 18*, 187-209. doi: 10.1177/1088868314527832

Taylor, C. T., & Amir, N. (2012). Modifying automatic approach action tendencies in individuals with elevated social anxiety symptoms. *Behaviour Research and Therapy*, *50*, 529–536. doi: 10.1016/j.brat.2012.05.004

Vandenbosch, K., & De Houwer, J. (2011). Failures to induce implicit evaluations by means of approach–avoid training. *Cognition & Emotion*, *25*, 1311–1330. doi: 10.1080/02699931.2011.596819

Whitfield, M., & Jordan, C. H. (2009). Mutual influence of implicit and explicit attitudes. *Journal of Experimental Social Psychology*, *45*, 748–759. doi: 10.1016/j.jesp.2009.04.006

Wiers, R. W., Eberl, C., Rinck, M., Becker, E. S., & Lindenmeyer, J. (2011). Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science*, *22*, 490–497. doi: 10.1177/0956797611400615

Wiers, R. W., Gladwin, T. E., & Rinck, M. (2013). Should we train alcohol-dependent patients to avoid alcohol? *Frontiers in Psychiatry*, *4*, 33. doi: 10.3389/fpsyt.2013.00033

Wiers, R. W., & Stacy, A. W. (2006). Implicit cognition and addiction. *Current Directions in Psychological Science*, *15*, 292–296. doi: 10.1111/ j.1467-8721.2006.00455.x

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*, 101–126. doi: 10.1037/0033-295X.107.1.101

Wittekind, C. E., Feist, A., Schneider, B. C., Moritz, S., & Fritzsche, A. (2015). The approach-avoidance task as an online intervention in cigarette smoking: A pilot study. *Journal of Behavior Therapy and Experimental Psychiatry*, *46*, 115–120. doi: 10.1016/j.jbtep.2014.08.006

Woud, M. L., Becker, E. S., Lange, W.-G., & Rinck, M. (2013). Effects of Approach-Avoidance Training on Implicit and Explicit Evaluations of Neutral, Angry, and Smiling Face Stimuli, *Psychological Reports*, *113*, 1211–28. doi: 10.2466/21.07

Woud, M. L., Becker, E. S., & Rinck, M. (2008). Implicit evaluation bias induced by approach and avoidance. *Cognition & Emotion*, *22*, 1187–1197. doi: 10.1080/02699930801906843

Woud, M. L., Maas, J., Becker, E. S., & Rinck, M. (2013). Make the manikin move: Symbolic approach–avoidance responses affect implicit and explicit face evaluations. *Journal of Cognitive Psychology*, *25*, 738–744. doi: 10.1080/20445911.2013.817413

Zajonc, R. B. (1980). Feeling and thinking. *American Psychologist*, *35*, 151–175. doi: 10.1037/0003-066X.35.2.151

Zogmaister, C, Perugini, M., Richetin, J (in press). Motivation modulates the effect of approach on implicit preferences. *Cognition & Emotion*. doi: 10.1080/02699931.2015.1032892

CHAPTER 2

INSTRUCTION-BASED APPROACH-AVOIDANCE EFFECTS[1]

Prior research suggests that repeatedly approaching or avoiding a certain stimulus changes the liking of this stimulus. We investigated whether these effects of approach and avoidance training occur also when participants do not perform these actions but are merely instructed about the stimulus–action contingencies. Stimulus evaluations were registered using both implicit (Implicit Association Test and evaluative priming) and explicit measures (valence ratings). In Experiment 1, typical approach and avoidance effects were observed when participants were merely instructed to approach or avoid nonwords. In Experiments 2 through 5, instruction-based approach-avoidance effects were observed also for fictitious social groups, but not for well-known social groups. We conclude that instructions to approach or avoid stimuli provide sufficient bases for establishing both implicit and explicit evaluations of novel stimuli. We discuss several possible reasons for why similar instruction-based approach-avoidance effects were not found for well-known stimuli.

[1] Based on Van Dessel, P., De Houwer, J., Gast, A., & Smith, C.T. (2015). Instruction-Based Approach-Avoidance Effects: Changing Stimulus Evaluation via the Mere Instruction to Approach or Avoid Stimuli. *Experimental Psychology, 62*, 161–169. doi: 10.1027/1618-3169/a000282.

INTRODUCTION

In recent years, it has been argued that there is a bi-directional link between attitudes and approach-avoidance motor actions (Neumann, Förster, & Strack, 2003). On the one hand, attitudes are thought to determine the speed with which people perform approach and avoidance motor actions (Solarz, 1960, Chen & Bargh, 1999). On the other hand, the execution of approach and avoidance actions during stimulus processing is said to influence attitude formation and change (Cacioppo, Priester, & Berntson, 1993). As a result, repeatedly approaching a stimulus leads to more positive evaluations of this stimulus whereas repeated avoidance leads to more negative evaluations. Approach and avoidance (AA) training has been found to influence not only explicit (i.e., non-automatic) evaluations of stimuli but also implicit (i.e., automatic) evaluations (e.g., Kawakami, Phills, Steele, & Dovidio, 2007). In this paper, we extend earlier work by exploring the possibility that instructions about AA training can impact implicit evaluations without necessitating the actual execution of these AA actions.

A number of studies have provided evidence that AA training influences implicit evaluations of *novel* stimuli such as unknown persons or fictitious social groups (e.g., Woud, Becker, & Rinck, 2008; Huijding, Muris, Lester, Field, & Joosse, 2011; Slepian, Young, Rule, Weisbuch, & Ambady, 2012; Woud, Maas, Becker, & Rinck, 2013; Laham et al., in press). Additionally, Kawakami and colleagues (2007) observed effects of AA training on implicit evaluations of *well-known* social groups. In a series of studies, they found significant reductions in White people's implicit preference for faces of White people over Black people after they had responded with approach actions to photos of Black faces and with avoidance actions to photos of White faces. In line with these results, typical AA training effects have been reported in studies with other well-known stimuli, such as pictures of familiar alcoholic drinks (Wiers, Eberl, Rinck, Becker, & Lindenmeyer, 2011), insects and spiders (Jones, Vilensky, Vasey, & Fazio, 2013), or contamination-related objects (Amir, Kuckertz, & Najmi, 2013). Not all attempts to find effects of AA training, however, have been successful (e.g., Vandenbosch & De Houwer, 2011), suggesting that there are as yet undiscovered boundary conditions (Laham et al., in press; Vandenbosch & De Houwer, 2011).

At a mental process level, AA training effects are typically interpreted within the framework of embodied cognition. From this perspective, mental representations are assumed to be grounded in modality specific systems of perception and motor action (Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005). AA processes are given a special status as they are considered essential for successful adaptation to the environment. AA behavior supplies humans with an adaptive response when encountering stimuli that are potentially beneficial or harmful (Elliot, 2006). Embodiment theories assume that, as a result of this evolutionary benefit, evaluative processing is closely tied to representations of AA behavior. More specifically, they postulate that motivational systems of AA mediate the relation between AA behavior and stimulus evaluations (Cacioppo et al., 1993). Motivational systems of AA are activated automatically during the processing of positive or negative stimuli, thereby triggering AA actions (Chen & Bargh, 1999). In turn, because AA actions are wired into these motivational systems, performing AA actions also leads to the activation of these motivational systems, which can bias the automatic evaluative processing of stimuli (Neumann & Strack, 2000). Most important for the purposes of our paper, approaching or avoiding a stimulus is assumed to have long term effects on the evaluation of that stimulus via the formation of associations in memory (Strack & Deutsch, 2004). Each time that the stimulus is approached or avoided, the corresponding stimulus representation and motivational representation are both activated, thereby gradually strengthening the association between those representations. Consequently, AA training effects are assumed to necessitate a large number of trials in which the AA behavior is performed in response to the stimulus (Woud et al., 2008, Phills, Kawakami, Tabi, Nadolny, & Inzlicht, 2011).

There are, however, reasons to believe that the standard embodiment theory of AA training is incomplete at best. First, it has been argued that AA behavior is not simply hard-wired into motivational systems. Instead, the motivational implication of AA responses seems to depend on how these responses are coded cognitively (e.g., Eder and Klauer, 2009). For instance, the same AA action (e.g., pushing a lever toward a stimulus) can be activated both by positive and by negative stimuli depending on how the action is framed (e.g., as moving toward the stimulus or as pushing the stimulus away; e.g., Eder & Rothermund, 2008). Even the mere planning or anticipation of the AA response might result in the activation of motivational representations (e.g., Eder & Klauer, 2009; Hommel, 2004). Second, contrary to the standard view that associations are formed in a slow, gradual manner, some have argued that associations in memory can emerge very quickly, even as the result of mere instructions or propositional reasoning (e.g., Fazio, 2007, p. 609; Field, 2006, pp. 867-868). Likewise, recent non-associative accounts of learning allow for learning via the rapid formation of propositions via instructions or inferences (De Houwer, 2009; Mitchell, De Houwer, & Lovibond, 2009). Once acquired, these propositions might even be activated automatically and hence underlie not only explicit but also implicit evaluations (see Hughes, Barnes-Holmes & De Houwer, 2011).

Based on these theoretical considerations, we put forward the hypothesis that a stimulus does not actually have to be physically approached or avoided in order for AA training effects to arise. Instead, the mere instruction to approach or avoid a stimulus might suffice to produce changes in the evaluation of that stimulus, even its implicit evaluations.

Although we are the first to examine AA training via instructions, it has already been demonstrated that mere instructions about future events can influence both implicit and explicit evaluations. For instance, in studies on evaluative conditioning (EC) via instructions, De Houwer (2006) told participants that they would see trials on which a first neutral stimulus is paired with positive pictures and trials on which a second neutral stimulus is paired with negative pictures. Despite the fact that the participants never actually saw the stimulus pairings, the instructions did result in a preference for the first neutral stimulus over the second one, even on measures of implicit evaluation. Following up on this finding, a study by Gast and De Houwer (2013) provided evidence that instruction-based EC is moderated by variables that also moderate practice-based EC. Of course, these findings do not imply that instructions about stimulus-action contingencies also induce changes in liking, especially because of the special motivational significance of actually performing AA responses. Nevertheless, if mere instructions about stimulus-stimulus relation can produce changes in liking, than it is at least plausible that mere instructions about stimulus-action relations also produce changes in liking.

In our studies, we therefore adapted the instruction-based EC procedure of De Houwer (2006) in such a way that participants received instructions about a later phase in which they would be asked to approach or avoid stimuli. If we could demonstrate instructed AA effects on (implicit) evaluations, it would raise a number of interesting issues. Most importantly, it would highlight the possibility that at least some instances of AA training effects with actual practice are due (in part) to the instructions that participants receive about the stimulus-action mappings rather than the AA training itself. Hence, demonstrating AA effects on the basis of mere instructions would set the stage for a range of new questions about whether, when, and how instructions and actual AA practice influence evaluation. Examining these questions is bound to have important implications for theories of AA training effects, most importantly the relative role of the top-down and bottom-up processes that mediate these effects. Note that similar issues would need to be addressed also in studies in which instructions do not directly link attitude objects with AA responses, such as studies in which participants approach or avoid stimuli based on an arbitrary stimulus features (e.g., the color or orientation of pictures; see Wiers et al., 2011). Even in these studies, participants might gain conscious knowledge about the stimulus-action relations and hence effectively instruct themselves about these contingencies. Studies in which AA training effects occur in the absence of conscious knowledge of the stimulus-action relations would not be affected (e.g., Kawakami et al., 2007, Exp. 2) but currently the evidence for such unconscious AA training effects is limited at best.

Demonstrating instruction-based AA effects could also have practical implications. If mere AA instructions prove to be effective in changing the implicit evaluation of stimuli under certain conditions, they could replace or complement actual AA training as an efficient means of inducing implicit evaluations (e.g., of novel products) or changing existing implicit evaluations (e.g., toward Black people; Kawakami et al., 2007) or addictive behaviors (Wiers et al., 2011) or social anxiety (Taylor & Amir, 2012). Of course, much of the applied value of AA instructions will depend on how strong and general these effects are, but this can be established only by actually examining the power and limitations of instruction-based AA effects.

Although the main aim of our work was to examine whether AA instruction can influence implicit and explicit evaluations, we already looked at a first potential boundary condition of these effects, being the type of attitude object. More specifically, in Experiment 1, we investigated effects on unfamiliar nonwords whereas in Experiments 2 through 5 we investigated effects on fictitious and well-known social groups. Previous studies suggest that instructions might be more effective in changing the implicit evaluations of novel, affectively neutral attitude objects than in altering the existing evaluations of known, affectively laden attitude objects. For instance, Gregg, Seibt, and Banaji (2006) observed that implicit evaluations of novel social groups could be induced quite easily on the basis of instructions about the behavior and traits of those groups, but could not be undone by giving additional instructions about those groups. Although Gregg and colleagues did not manipulate directly whether the attitude objects were novel or affect-laden, their results are in line with the common sense idea that instructions might not be powerful enough to change existing (implicit) evaluations of well-known attitude objects.

EXPERIMENT 1

Method

Participants

Forty native Dutch-speaking undergraduates (31 women) participated in exchange for 4 euros. All participants had normal or corrected-to-normal vision and were naive to the purpose of the experiment.

Apparatus and Materials

Two nonwords were used as evaluation stimuli, namely 'BAYRAM' and 'UDIBNON'. These were two of the four stimuli used in the study by De Houwer (2006, Exp. 1). We decided to use only two stimuli, in order to reduce the possibility that participants would be unable to correctly memorize the instructions for all stimuli. During the Implicit Association Test (IAT, Greenwald, McGhee, & Schwarz, 1998), four positive words (the Dutch words for HAPPY, HONEST, NICE, and SINCERE) and four negative words (the Dutch words for MEAN, BRUTAL, AGGRESSIVE, and FAKE) were presented as attribute stimuli, in addition to the two nonwords, which were used as target stimuli. All words were presented in uppercase letters in Arial Black font with font size 36. The experiment was programmed and presented using the INQUISIT Millisecond Software package (Inquisit 3.0, 2011) on a Tori PC with a 19-inch monitor (120 Hz refresh rate), which had a keyboard and a joystick (Wingman attack 2) attached to it.

Procedure

After participants had given informed consent, they were seated in front of a computer screen. Half of the participants read the following instructions (translated from Dutch):

In this experiment, you will see two non-existing words. It is your task to make a certain action with the joystick each time you see one of these words.

If you see the word BAYRAM, you will have to approach it by pulling the joystick *towards you each time this word is presented.*

If you see the word UDIBNON, you will have to avoid it by pushing the joystick away from you each time this word is presented.

It is very important that you remember which action you will have to perform with each word. You need this information to complete the task successfully.

Note that this information will not be presented again later on, so remember well which word goes together with which action.
Before we will present the words, you need to complete a reaction time task. This will last approximately 10 minutes. Make sure that you do not forget which action you will have to perform later on with which word.

The other participants received identical instructions except that they were told that they would have to approach UDIBNON and avoid BAYRAM. Participants then pressed the space bar to proceed to the next screen where the instructions for the IAT were presented.

In the IAT, participants categorized positive words, negative words, and both nonwords into one of four categories: positive, negative, BAYRAM, or UDIBNON by pressing a left (Q) or right (M) key depending on the category of the presented word. The name of the categories assigned to the left key in any particular phase would always be presented in the left top corner of the screen and the name of the categories assigned to the right key would always appear in the right top corner. It was counterbalanced across participants which categories were assigned to which response in which phase. Half of the participants always pressed the left key for positive and the right key for negative words whereas the other participants pressed the left key for negative words and the right key for positive words. Orthogonal to this manipulation, half of the participants started by pressing the left key for the nonword UDIBNON and the right key for the nonwords BAYRAM, whereas the other half started with the opposite assignment. All participants started with a practice block of 16 trials during which both nonwords were each presented eight times. Next came a practice block of 32 trials during which four positive and four negative words were each presented four times. The practice blocks were followed by two test blocks of 32 trials. During each test block, each of the nonwords was presented on eight trials, and each of the positive and negative words was presented on two trials. The practice block with nonwords was then repeated but the response assignments were reversed. Finally, the two test blocks were repeated but this time with the reversed response assignment for the nonwords. The order of the trials was determined randomly for each block and each participant separately. On each trial, a word or nonword was presented in the center of the screen until the participant pressed one of the two valid keys (i.e., Q or M). If the response was correct, the word disappeared and the next word was presented 400 ms later. If the response was incorrect, the word was replaced by a red “X” for 400 ms. The next word appeared 400 ms after the red “X” was removed from the screen.

After the IAT, we assessed whether participants had correctly remembered which actions they had to perform with which nonword. They were asked two questions: (1) “According to the instructions, what will you have to do when the word BAYRAM is presented?”, and (2) “According to the instructions, what will you have to do when the word UDIBNON is presented?” Participants chose between the words ‘approach’ or ‘avoid’ for each question. The order in which the questions were presented was determined randomly for each participant.

Finally, even though performance on this task was irrelevant for our hypotheses, participants performed twenty trials of an AA training task in which they were instructed to act as stated in the instructions they had received at the start of the experiment. Each of the two nonwords was presented ten times in the center of the screen until participants responded by pushing the joystick forward or by pulling it backward. This task was included in order not to deceive participants in the earlier instructions.

Results

The data of two participants were discarded because they did not correctly answer the memory questions.[2] The IAT score was calculated using the D4 procedure outlined by Greenwald, Nosek, and Banaji (2003) in a way that positive values indicate an implicit preference for BAYRAM over UDIBNON and negative values indicate an implicit preference for UDIBNON over BAYRAM. Split-half reliability of the IAT, computed by correlating IAT D scores from the first half of the blocks with scores from the second half, was $r(38) = .57$. Across groups, participants did not display an implicit preference for any of the nonwords (M = 0.07, SD =0.43), $t(37) = 1.02$, $p = .31$. Crucially, a between-groups t-test indicated a successful effect of the instructions, $t(36) = 2.10$, $p = .043$, $d = 0.36$ (Figure 1). Participants who had been instructed to approach BAYRAM and avoid UDIBNON showed a stronger implicit preference for BAYRAM (M = 0.21, SD = 0.50) than participants who had been instructed to avoid BAYRAM and approach UDIBNON (M = -0.07, SD = 0.30).

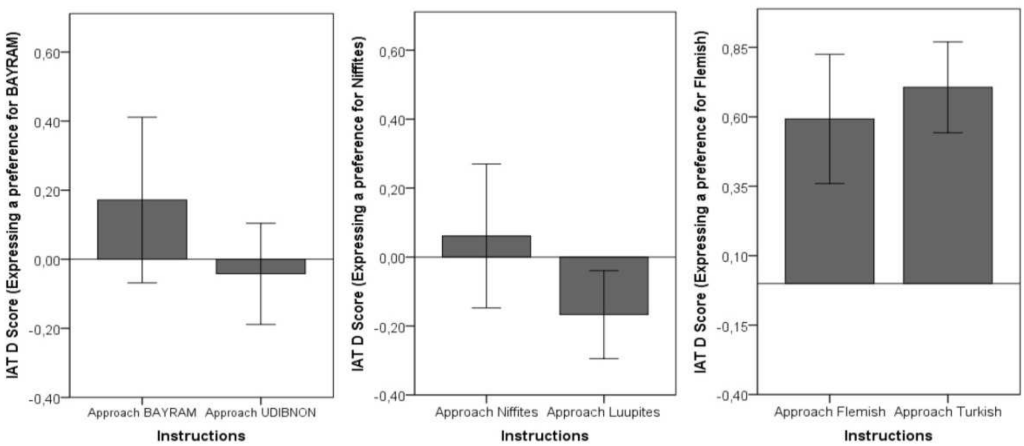

Figure 1. Mean IAT D scores indicating an implicit preference for nonword BAYRAM over UDIBNON, Niffites names over Luupites names or Flemish names over Turkish names, respectively, as a function of instructions, in Experiments 1, 2, and 3. Error bars represent 95% confidence intervals.

[2] Across all five experiments we observed that, when participants did not correctly remember the instructions, they did not show any effects of AA instructions. This suggests that knowledge of the stimulus–action contingencies is a necessary condition for instruction-based AA effects. This is in line with evidence showing that EC effects are stronger or only existent if participants know which US was paired with which CS (see Gast, De Houwer & De Schryver, 2012). Note that the lack of memory could result from processes involved in the encoding, storage, or recall of the contingencies. Importantly, including the data from these participants in the analyses did not result in any shift in significance for the effects of AA instructions on novel or well-known stimuli.

Discussion

Experiment 1 showed that merely instructing participants to approach or avoid certain nonwords leads to more positive evaluations of the nonword they were instructed to approach than of the nonword that they were instructed to avoid. This effect is in line with typical AA training effects, but it occurred even though participants performed the AA training task only *after* their evaluations were probed. Hence, our results demonstrate that the actual execution of AA responses is not necessary for AA effects.

There are important limitations to this study. First, evaluations were only probed using an implicit measure. We have no information on whether the instructions were also sufficient to change evaluations when probed with explicit measures. Second, in previous AA training studies, stimuli were used for which AA behavior is more relevant, such as pictures of spiders (Jones et al., 2013) or human faces (Taylor & Amir, 2012). Moreover, AA training effects have also been observed for evaluations of well-known stimuli. For example, in the study by Kawakami and colleagues (2007), participants approached photos of faces that could be categorized as belonging to White or Black people and evaluations towards these groups were measured.

EXPERIMENT 2 AND 3

To deal with the limitations of Experiment 1, we performed two experiments investigating effects of AA instructions on implicit and explicit evaluations of both novel and well-known stimuli. In Experiment 2, we replicated the first experiment with names of fictitious social groups (i.e., names of Niffites and Luupites, see Gregg et al., 2006). In Experiment 3, we used a similar approach to change evaluations of well-known social groups. In accordance with Kawakami and colleagues (2007), we measured changes in the relative evaluation of two social groups, namely a social group that participants belonged to (i.e., Flemish people) and a social group that Flemish people are prejudiced towards (i.e., Turkish people; see Dhont & Van Hiel, 2009).

Method

Participants

Participants were eighty native Dutch-speaking undergraduates. Forty participants took part in Experiment 2 (33 women) and forty participants took part in Experiment 3 (32 women). None of the participants performed both experiments or had previously participated in Experiment 1.

Procedure

Experiment 2 was identical to Experiment 1 except for the following points. First, participants were instructed that they would be presented with the names of two groups, called Luupites and Niffites. They were told that all the names of Luupites have two consecutive vowels in them and end with “lup”. They were then shown two examples of Luupites’ names (i.e., Loomalup, Ageelup). Subsequently, participants were told that all the names of Niffites would contain two consecutive consonants and end with “nif.” Again, this statement was followed by two Niffites names (i.e., Borrinif, Kennunif). Next, half of the participants were told that they would have to approach each name of a Luupite and avoid each name of a Niffite by respectively pulling a joystick towards them or pushing a joystick away from them. The other participants were given the opposite instruction. Second, in the IAT, the labels LUUPITES, NIFFITES, POSITIVE, and NEGATIVE were used. During the practice blocks with names, four Luupites names (i.e., Meesolup, Naanolup, Omeelup, Wenaalup) and four Niffites names (i.e., Cellanif, Eskannif, Lebbunif, Zallunif) were categorized, each on four occasions (i.e., 32 trials in total). During the test blocks, each of the eight names and each of the positive and negative words were presented twice (i.e., 32 trials in each block). Third, the memory questions were directed at assessing memory for whether the instructions linked Niffites or Luupites to either an approach or avoidance action. Finally, before the AA training task was performed, participants completed two self-report measures tapping into their evaluations of Niffites and Luupites (Iyengar, Messing, Hahn, Banaji, & Dial, 2011).

Participants completed thermometer ratings of self-reported warmth or cold feelings towards Niffites and Luupites on a 9-point Likert scale (1= not warm/liked at all; 9 = completely warm/liked). Additionally, participants completed two trait ratings. They indicated whether they thought Niffites and Luupites were aggressive and friendly on two seven-point Likert scales (1: not friendly/aggressive, 7: very friendly/aggressive).

Experiment 3 followed the same procedure as Experiment 2 with the following exceptions. First, participants were instructed that they would be presented with names of Turkish and Flemish people. Half of the participants were instructed to approach the names of Turkish people and avoid the names of Flemish people. The other participants were given the opposite instructions. Second, in the IAT, four traditionally Flemish (i.e., Pieter, Bruno, Mark, Steven) and four traditionally Turkish names (i.e., Ali, Mohammed, Moustafa, Achmed) were used as stimuli in addition to the positive and negative words. Third, as explicit measures we used the warmth thermometer measure as well as the Overt Racism Scale (ORS). This measure of explicit racial attitudes consists of three trait ratings and has been shown to successfully predict prejudiced behavior (Iyengar et al., 2011). Participants completed this measure by indicating whether they thought Flemish and Turkish people were aggressive, friendly and lazy on seven-point Likert scales.

Results

Experiment 2

The data of four participants who did not correctly answer the memory questions were discarded. The IAT difference score (i.e., the D4-score) was calculated in a way that positive values indicated an implicit preference for Niffites over Luupites and negative values indicated an implicit preference for Luupites over Niffites. Split-half reliability was $r(36) = .17$. The IAT scores indicated that overall participants did not display a significant implicit preference for either Niffites or Luupites (M = -0.07, SD =0.35), $t(35) = -1.14$, $p = .26$. Crucially, a between-groups t-test revealed a significant effect of the instructions, *t*(34) = 2.06, *p* = .047, *d* = 0.33 (Figure 1). Participants who had been instructed to approach Niffites and avoid Luupites had a stronger implicit preference for Niffites (*M* = 0.06, *SD* = 0.39) than participants who had been instructed to avoid Niffites and approach Luupites (*M* = -0.17, *SD* = 0.27).

The explicit evaluations were collapsed into two scores. The warmth score signified the difference in warmth feelings for Niffites and Luupites and was calculated by subtracting each participants' warmth score rating for Luupites from the score for Niffites. The trait rating score was calculated by subtracting the ratings for friendliness and the inverted ratings for aggressiveness of Luupites from the corresponding ratings for Niffites and averaging the resulting difference scores. The internal consistency of this measure was moderate (Cronbach's Alpha = .58). Positive values for the warmth and trait rating scores indicate a preference for Niffites over Luupites and negative scores indicate a preference for Luupites. Both measures revealed a significant preference for Luupites (warmth score: *M* = -1.58, *SD* = 2.49, *t*[35] = -3.82, *p* = .001; trait rating: *M* = -0.60, *SD* =1.40, *t*[35] = -2.56, *p* = .015). Importantly, we found a significant instruction effect for both the warmth score (approach Niffites: *M* = -0.19, *SD* = 2.14; approach Luupites: *M* = -2.70, *SD* = 2.20), *t*(34) = 3.45, *p* = .002, *d* = 1.16, and the trait rating score (approach Niffites: *M* = 0.03, *SD* = 1.53; approach Luupites: *M* = -1.10, *SD* = 1.08), *t*(34) = 2.59, *p* = .014, *d* = 0.85. That is, participants who had been instructed to approach Luupites and avoid Niffites preferred Luupites more than participants who were instructed to approach Niffites and avoid Luupites.

Experiment 3

All participants correctly remembered which action they had to perform with names from each group. The IAT score was calculated in a way that positive values indicated a preference for Flemish over Turkish names. In line with earlier prejudice literature (e.g., Kawakami et al., 2007) positive values indicate an implicit prejudice towards the prejudiced group (i.e., Turkish people). Split-half reliability of the IAT was *r*(40) = .43. An analysis of the IAT performance demonstrated that participants displayed a strong implicit preference for Flemish names over Turkish names (*M* = 0.65, *SD* = 0.43), *t*(39) = 9.60, *p* < .001. Crucially, an independent-samples t-test did not indicate a significant effect of the instructions, *t*(38) = -0.84, *p* = .41, *d* = -0.26 (Figure 1). Participants who had been instructed to approach Flemish names (*M* = 0.59, *SD* = 0.50) did not prefer Flemish people more than participants who were instructed to approach Turkish names (*M* = 0.71, *SD* = 0.35).

The explicit evaluations were collapsed into two scores, a warmth score and a trait rating (ORS) score. The internal consistency of the ORS score was moderate (Cronbach's Alpha = .55). For the ORS and warmth score, positive numbers indicate an explicit preference for Flemish people. As expected, both scores revealed a preference for Flemish people (warmth score: *M* = 1.53, *SD* = 1.81, *t*[39] = 5.32, *p* < .001; ORS score: *M* = 0.73, *SD* = 0.93, *t*[39] = 5.00, *p* < .001). However, we found no significant difference due to the instructions that participants had received, neither on the warmth score (approach Flemish: *M* = 1.90, *SD* = 1.92; approach Turkish: *M* = 1.15, *SD* = 1.66, *t*[38] = 1.32, *p* = .19, *d* =0.42), nor on the ORS score (approach Flemish: *M* = 0.95, *SD* = 1.01; approach Turkish: *M* = 0.52, *SD* = 0.81, *t*[38] = 1.50, *p* = .14, *d* = 0.47).

Discussion

The data of Experiment 2 show that evaluations towards fictitious social groups can be readily induced by instructing participants to approach or avoid members of the group. Whereas Experiment 1 examined an AA instruction effect only on implicit but not explicit evaluations of novel stimuli, in Experiment 2, the results showed that both implicit and explicit evaluations of fictitious social groups were influenced by the instructions to approach or avoid names of group members. This is an important extension from Experiment 1, because it indicates that instructions to approach or avoid novel stimuli operate similarly when participants consider novel social groups as when they consider novel words. In contrast, in Experiment 3 we did not observe an instruction effect for implicit or explicit evaluations of well-known social groups. That is, no evidence was found that participants who were instructed to approach or avoid names of Turkish or

Flemish people had changed their prejudice towards Turkish people as a result of these instructions. In fact the direction of the non-effect was opposite for implicit as compared to explicit measures.

There are a number of possible explanations for not observing an instruction effect on evaluations of well-known social groups in Experiment 3. First, evaluations of these groups might be more resistant to change due to the prior knowledge people have about these stimuli (e.g., Cacioppo, Marshall-Goodell, Tassinary, & Petty, 1992; Priester, Cacioppo, & Petty, 1996). In line with this, EC effects are generally larger for neutral CSs than for valenced CSs (Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). Second, it is possible that an effect of evaluations was not observed due to a lack of power in this experiment. Indeed, performing a power analysis revealed that the achieved power for revealing a medium or small effect was low (i.e., for a medium effect: power = 0.46; for a small effect: power = 0.15). This is important because in practice-based AA training studies, the observed effects are typically of small or medium size (e.g., Kawakami et al., 2007; Phills et al., 2011). Moreover, these studies have provided evidence that implicit prejudice is changed only when participants are trained to approach members of the prejudiced group and avoid members of the group that participants belong to, and not when participants are trained to avoid members of the prejudiced group and approach members of the group that participants belong to (Kawakami et al.). Additionally, Wennekers (2013) observed that only highly prejudiced people are influenced by approach training. It is possible that the number of strongly prejudiced participants in the approach Turkish condition might have been too small to observe any effect in Experiment 3.

EXPERIMENT 4

In order to obtain more power and directly compare the effect of AA instructions on the evaluation of novel and well-known socials groups, we ran a new study that involved a large number of participants who were assigned to either a condition with novel socials groups or well-known social groups. A priori power analyses indicated that, to detect a small effect (i.e., effect size *d* = 0.20; see Cohen, 1992) with sufficient power (power > .75) approximately 270 participants needed to be included in each between-subjects condition. We were able to recruit this high number of participants by implementing our study on the internet. Performing this study in a larger sample also allowed us to subdivide the sample based on different kind of criteria to gain additional information about the scope of instruction-based AA effects. To this end, we asked participants to indicate whether they had inferred that the purpose of the experiment was to change their attitudes and to what extent they believed that performing this experiment might have changed their attitudes. This allowed us to investigate whether, in line with AA training studies (e.g., Kawakami et al., 2007), effects can be observed even in subgroups of participants who do not infer, or even believe, that approach or avoidance would influence their attitudes.

Method

Participants

Participants were 949 visitors of the Project Implicit research website (618 women, mean age = 34.44, *SD* = 13.84). Participation was restricted to United States citizens. 91 participants (9.6%) were either African American or of mixed Black-White race. In line with the study by Kawakami and colleagues (2007), their data were not included in the analyses.[3]

Procedure

All participants were randomly assigned to either the condition with fictitious social groups (i.e., Niffites and Luupites) or the condition with well-known social groups (i.e., Whites and Blacks). Similar to Experiment 2, in the

[3] In both Experiments 4 and 5, including the data of participants who were Black or of mixed White-Black race did not significantly change the data pattern. Including a 'race' variable in the ANOVA's did reveal an additional effect of race on implicit and explicit evaluations for participants in the Whites/Blacks condition, Fs > 11, ps < .002. That is, participants of Black or mixed White-Black race had less implicit and explicit preference for White names.

Niffites/Luupites condition one group of participants was instructed to approach names of Niffites and avoid names of Luupites and a second group was instructed to approach names of Luupites and avoid names of Niffites. Similar to Experiment 3, in the Whites/Blacks condition one group was instructed to approach typical names of members of a social group they belong to (i.e., White people) and avoid typical names of members of a group that people of the group participants belong to are known to be prejudiced towards (i.e., Black people). The second group received reversed instructions (i.e., to approach names of Black people and avoid names of White people).

In comparison to the previous experiments, a number of adaptations were made to the procedure. First, the IAT procedure was altered in two ways. If participants made an error, a red “X” appeared on the screen and the participant had to correct their mistake in order to continue. Latencies were recorded until a correct response was made. Also, in the Whites/Blacks condition the names that were used in the IAT were five prototypical names of Black people (i.e., Darnell, Leroy, Terrence, Tyrone, Jerome) and five prototypical names of White people (i.e., Alfred, Hank, Edmund, Wilbur, Marty). These names were matched on word familiarity in a US-sample and have been used in previous studies on implicit prejudice (Ottaway, Hayden, & Oakes, 2001). Second, the explicit measure that participants completed consisted of four ratings. Participants in the Niffites/Luupites condition indicated the extent to which they liked and had warm feelings for Niffites and Luupites, and participants in the Whites/Blacks condition completed these ratings for Black and White people. Third, after completing the memory test, participants answered three additional questions. Specifically, participants indicated whether they thought that the purpose of the experiment was to change their attitude towards the social groups. Subsequently, participants indicated when they first started thinking that this was a purpose of the experiment (i.e., when reading the AA instructions, when performing the IAT, or when completing the explicit measure). Finally, participants indicated to what extent they believed that performing this experiment could have changed their attitude towards the social groups on a 5-

point Likert scale. Fourth, the approach and avoidance training task was adapted to ensure that also participants who did not possess a joystick would be able to participate. During this task participants pushed away names by pressing the up arrow on the keyboard (i.e., avoided) and pulled names towards them by pressing the down arrow on the keyboard (i.e., approached). A zoom effect enhanced the visual experience of approaching or avoiding.

Results

Data preparation

Data from 83 participants (10.7%) who did not fully complete all questions and tasks were not included in the analysis. Additionally, in line with our previous experiments, the data of 171 participants (Niffites/Luupites: 19.0 %, Blacks/Whites: 22.2 %) who did not correctly answer the memory questions were discarded.

The IAT-scores were calculated using the D2-algorithm (Greenwald et al., 2003). This procedure is very similar to the D4-algorithm with the distinction that the D2-algorithm takes into account that participants had to self-correct their errors. In line with standard analyses of Project Implicit IAT scores (e.g., Smith, De Houwer, & Nosek, 2013), data from 25 participants (3.9%) were dropped because of error rates above 30% across the entire task, or above 40% for any one of the four critical blocks. After removing participants based on the previous two criteria, there were no additional participants who needed to be removed because of completing more than 10% of trials faster than 400ms. The analyses were performed on the data of 625 participants (i.e., 271 in the Niffites/Luupites condition and 352 in the Whites/Blacks condition). Split-half reliability of the IAT score was $r(271) = .63$ for participants in the Niffites/Luupites condition and $r(352) = .53$ for participants in the Whites/Blacks condition.

The responses on the explicit measures were collapsed into two scores. The rating scores (i.e., warmth score and liking score) were calculated by subtracting the score rating for Luupites/Blacks from the corresponding score rating for Niffites/Whites. Positive scores indicate a preference for

Niffites/Whites.

Performance on implicit and explicit measure in Niffite/Luupite condition

Analysis of the IAT scores indicated that participants preferred Luupites over Niffites (*M* = -0.14, *SD* =0.52), *t*(270) = -4.29, *p* < .001. Crucially, a between-groups t-test revealed a significant effect of the instructions, *t*(269) = 7.98, *p* < .001, *d* = 0.97 (Figure 2). When participants had been instructed to approach Niffites and avoid Luupites, the former was preferred (*M* = 0.08, *SD* = 0.51), *t*(139) = 1.98, *p* = .050, and when participants had been instructed to avoid Niffites and approach Luupites, the latter was preferred (*M* = -0.37, *SD* = 0.43) , *t*(130) = -9.90, *p* < .001.

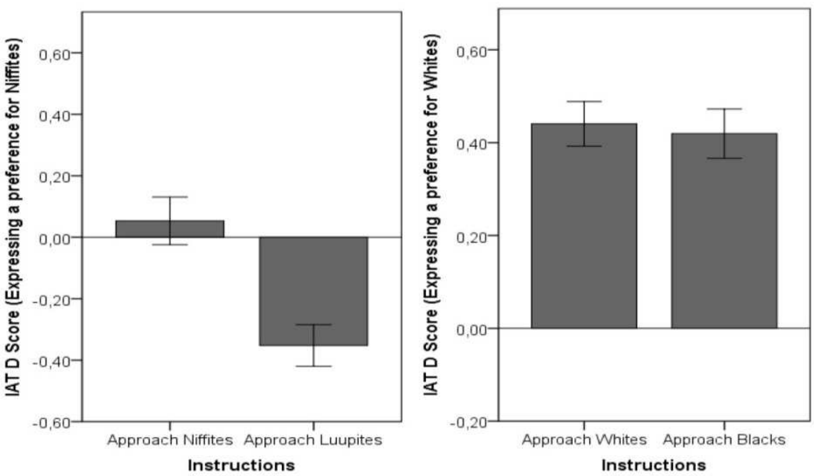

Figure 2. Mean IAT D scores indicating an implicit preference for Niffites names over Luupites names or names of Whites over names of Blacks, respectively, as a function of instructions, for participants who correctly remembered instructions in Experiment 4. Error bars represent 95% confidence intervals.

The explicit liking score revealed a significant preference for Luupites (*M* = -0.31, *SD* = 2.34, *t*[270] = -2.18, *p* = .03), whereas no significant difference was observed on the warmth score (*M* = -0.22, *SD* = 2.47, *t*[270] = -1.47, *p* = .14). Between-groups t-tests revealed a significant instruction effect both on the warmth score (approach Niffites: *M* = 0.66, *SD* = 2.21; approach Luupites: *M* = -1.16, *SD* = 2.40), *t*(269) = 6.49, *p* < .001, *d* = 0.79, and the liking score (approach Niffites: *M* = 0.51, *SD* = 2.06; approach Luupites: *M* = -1.19, *SD* = 2.31), *t*(269) = 6.42, *p* < .001, *d* = 0.78).

Performance on implicit and explicit measure in Whites/Blacks condition

Analysis of the IAT performance replicated previous research on implicit prejudice (e.g., Dasgupta, McGhee, Greenwald, & Banaji, 2000), demonstrating that participants displayed a strong implicit preference for Whites (M = 0.42, SD = 0.40) , $t(353)$ = 20.14, p < .001. Crucially, a significant main effect of instructions could not be observed, $t(352)$ = 0.01, p = .99, d < 0.01 (Figure 2). Participants who had been instructed to approach Whites and avoid Blacks did not have a significantly different degree of implicit prejudice (M = 0.42, SD = 0.36) compared with participants who had been instructed to avoid Whites and approach Blacks (M = 0.42, SD = 0.43).

The warmth score revealed a preference for Whites (warmth score: M = 0.25, SD = 1.51), $t(353)$ = 3.18, p = .002, whereas the liking score did not reveal such a preference (M = 0.08, SD = 1.27), $t(353)$ = 1.13, p = .26. Importantly, a t-test did not reveal a significant main effect of instructions for the liking score (approach Whites: M = 0.08, SD = 1.33; approach Blacks: M = 0.06, SD = 1.22) or the warmth score (approach Whites: M = 0.27, SD = 1.52; approach Blacks: M = 0.24, SD = 1.50), *ps* > .87, *d*s < 0.02.

Additional analyses

First, additional analyses corroborated that the instruction effect for participants in the Niffites/Luupites condition was significantly larger compared with participants in the Blacks/Whites condition. Second, correlational analyses indicated that the scores on the explicit and implicit measures were significantly correlated for participants in the Niffites/Luupites condition as well as for participants in the Whites/Blacks condition. Third, analyses including participants' answers on the hypothesis awareness questions indicated no impact of hypothesis awareness. A large effect of AA instructions in the Niffites/Luupites condition was observed even if participants did not think that a purpose of the experiment was to change their attitudes. More details on these analyses are available in Appendix.

Discussion

Despite an increased power for detecting also small effects (i.e., power = .60 to detect an effect size of *d* = 0.20), we found no evidence that AA instructions changed evaluations of well-known social groups. In contrast, the data provide convincing further evidence that being instructed to approach or avoid fictitious social groups influences evaluative responses towards members of these groups.

Experiment 4 also provided evidence that the instruction-based AA effect for evaluations of fictitious groups is not dependent on whether people correctly identify the purpose of the experiment. This indicates that, in line with practice-based AA training (e.g., Kawakami et al., 2007), participants do not need to be aware that the manipulation (i.e., the task to approach or avoid stimuli) was supposed to have an impact on their evaluations (i.e., impact awareness; see Gawronski, Hofmann & Wilbur, 2006) for the effect to arise.

EXPERIMENT 5

Most studies investigating AA training effects on implicit evaluations have used the IAT as the measurement tool of implicit evaluations (e.g., Kawakami et al., 2007; Phills et al., 2011). However, when other tasks are used, AA training has not always produced clear effects (e.g., Vandenbosch & De Houwer, 2011). Therefore, it can be argued that AA training effects on implicit evaluations and instruction-based effects in particular, are due to specific properties of the IAT. In Experiment 5, we investigated instruction-based AA effects by using a different task to measure implicit evaluations, namely the evaluative priming task (Fazio, Sanbonmatsu, Powell, & Kardes, 1986).

Method

Participants

Participants were 773 visitors (519 women, mean age = 39.53, *SD* = 14.38) to the Project Implicit research website. Participation was restricted to United States citizens. Seventy participants (9.1%) were either African American or of mixed Black-White race. Their data were not included in the analyses.

Procedure

The procedure of this experiment was identical to that of Experiment 4, with the only exception that participants performed an evaluative priming task instead of an IAT. Before the start of the evaluative priming task, participants were told that words would appear one after the other on the screen. They were instructed that their task was to categorize the words they saw as either "Good" or "Bad" using the 'E' and 'I keys of a computer keyboard and to try to do this as quickly as possible, while making as few mistakes as possible. Participants were instructed that they would see names presented before the words and told that they can watch the names, but that their task was simply to respond to the good and bad words. Then participants were shown a list of the 14 positive and 14 negative words that they would have to categorize. During all trials, the labels '"bad" and "good" appeared in the left and right upper corners of the screen, respectively. A single trial consisted of a fixation cross presented in white for 500 ms, a blank screen for 500 ms, a prime for 200 ms, a post-prime pause for 50 ms and the presentation of a target word in white font for 1500 ms. The inter-trial interval was set to vary randomly between 500 ms and 1500 ms. There were four types of trials in both conditions. Each type of trials was presented on $1/4^{th}$ of the trials. In the Niffites/Luupites condition the types of trials were trials with the word Niffite as prime and a positive word as target, trials with the word Niffite and a negative target, trials with the word Luupite and a positive target, or trials with the word Luupite and a negative target. In the Blacks/Whites condition the primes were the five prototypical names of Black people and the five prototypical names of White people that were used in the previous study. Note that in the Niffites/Luupites condition the prime was always either the word 'Niffite' or the word 'Luupite' and not the exemplars that were used in the previous experiment. We expected that, if participants would not be able to learn the categorization rule from the first instruction, participants would not categorize the exemplar primes as Niffites/ Luupites in the evaluative priming task. As this would make the task an invalid measure of implicit evaluations of the groups Niffites and Luupites, we decided to use the category labels as primes. In contrast, in the Blacks/Whites condition, names of Black people and White people were used as primes. We expected that participants would have no difficulty categorizing these names due to previous experience with these names. We did not use the words ‘Black’ and ‘White’ as prime (i.e., the category labels) because the names ‘Black’ and ‘White’ are very familiar last names in the US that can apply to both Black and White people (e.g., Barry White).

The types of trials consisted of trials with a typical White name as prime and a positive word as target, trials with a White name and a negative target, trials with a Black name and a positive target, or trials with a Black name and a negative target. Participants first completed eight practice trials (two of each of the four types of trials) and then completed 120 trials separated into three blocks of 40 trials, each containing 10 of the four types of trials, presented in random order.

Results

Data preparation

In line with Experiment 4, data from 22 participants (3.2 %) who did not fully complete all tasks and questions were discarded as well as the data of 100 participants (Niffites/Luupites: 13.0 %, Blacks/Whites: 16.6 %) who did not correctly answer the memory questions.

To calculate the evaluative priming score, trials with an incorrect response were dropped (7.7 %) as well as any trials on which reaction times were at least 2.5 standard deviations removed from an individual’s mean for that type of trial. A first difference score was created for each participant in the Niffites/Luupites condition by subtracting mean latencies for Niffites-positive trials from mean latencies for Niffites-negative trials. A second difference score was created in the same way for Luupites-trials such that, in both cases, higher scores indicate more positive evaluations for the group. Finally, the evaluative priming score was constructed by subtracting the difference score for Luupites-trials from the difference score for Niffites trials. A positive score indicates a preference for

Niffites. For participants in the Whites/Blacks condition the same procedure was used to construct an evaluative priming score that indicates a preference for Whites. Data from 37 participants (6.5 %) who made a substantial number of errors (i.e., more than 2.5 standard deviations over the population mean) in the evaluative priming task were dropped. The analyses were performed on the data of 533 participants (i.e., 257 in the Niffites/Luupites condition and 286 in the Whites/Blacks condition). A correlation between the priming scores for the first and second block of 60 trials indicated a split-half reliability of *r*(257) = .38 for participants in the Niffites/Luupites condition and *r*(286) = .29 for participants in the Whites/Blacks condition.

Performance on implicit and explicit measure in Niffite/Luupite condition

Analysis of the evaluative priming scores indicated no significant preference for either Niffites or Luupites (*M* = 3.14, *SD* =72.61), *t*(256) = 0.69, *p* = .49. A between-subjects t-test revealed an effect of instructions, *t(*255) = 4.26, *p* < .001, *d* = 0.59 (Figure 3). When participants had been instructed to approach Niffites and avoid Luupites, the former was preferred (*M* = 22.03, *SD* = 62.77), *t(*126) = 3.96, *p* < .001, and when participants had been instructed to avoid Niffites and approach Luupites, the latter was preferred (*M* = -15.32, *SD* = 76.94), *t(*129) = -2.27, *p* = .025.

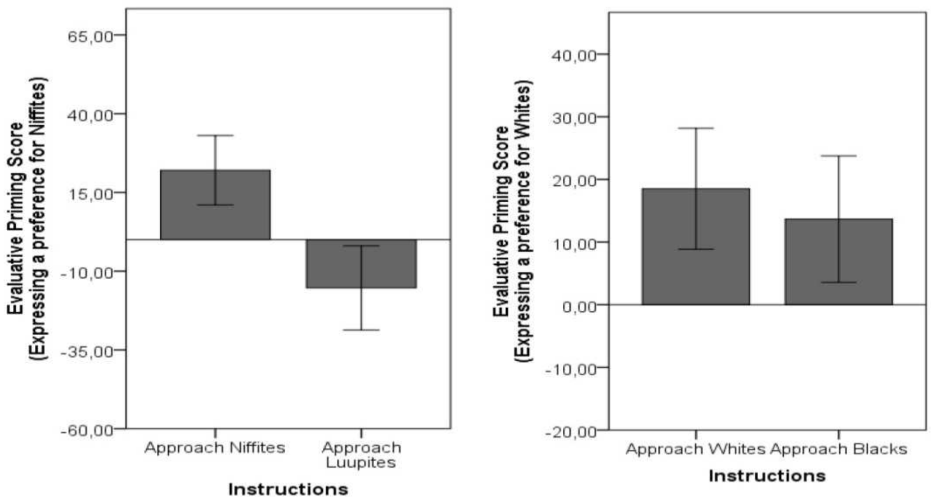

Figure 3. Mean evaluative priming scores indicating an implicit preference for Niffites names over Luupites names or names of Whites over names of Blacks, respectively, as a function of instructions, for participants who correctly remembered instructions in Experiment 5. Error bars represent 95% confidence intervals.

Both the warmth score and the liking score did not reveal a significant preference for any of the two groups (warmth score: M = -0.05, SD = 2.92, t[256] = -0.26, p = .80; liking score: M = -0.08, SD =2.86, t[256] = -0.46, p = .65). However, both the warmth score and the liking score indicated a significant instruction effect (warmth score: approach Niffites: M = 1.13, SD = 2.89; approach Luupites: M = -1.19, SD = 2.46, t[255] = 6.94, p < .001, d = 0.86; liking score: approach Niffites: M = 1.01, SD = 2.76; approach Luupites: M = -1.15, SD = 2.53, t[255] = 6.51, p < .001, d = 0.81).

Performance on implicit and explicit measure in Whites/Blacks condition

Analysis of the evaluative priming task indicated a preference for Whites (M = 16.21, SD = 58.57), t(275) = 4.60, p < .001. The between-subjects t-test did not reveal a significant main effect of instructions, t(274) = 0.69, p = .49, d = 0.08 (Figure 3). Participants who had been instructed to approach Whites and avoid Blacks did not have a significantly different degree of implicit prejudice (M = 18.51, SD = 58.83), compared with participants who had been instructed to avoid Whites and approach Blacks (M = 13.67, SD = 58.39).

The warmth score and liking score did not reveal a significant preference for Whites (warmth score: M = 0.04, SD = 1.32, t[275] = 0.46, p = .65; liking score: M = 0.02, SD = 1.09, t[275] = 0.28, p = .78). Also, a t-test did not reveal a significant main effect of instructions for the liking score (approach Whites: M = 0.09, SD = 1.08; approach Blacks: M = -0.06, SD = 1.10) or the warmth score (approach Whites: M = 0.06, SD = 1.19; approach Blacks: M = 0.01, SD = 1.45), ps > .25, ds < 0.14.

Additional analyses

Correlational analyses and analyses on the impact of hypothesis awareness revealed a similar data pattern as observed in Experiment 4 (see Appendix).

Discussion

The data of this experiment provide evidence that instruction-based AA changes are not restricted to explicit measures and performance on the IAT. As shown in the previous experiments on explicit measures and on IAT data, Experiment 5 shows that also effects in the priming data critically depend on the type of stimulus that was used. Effects were clearly present in evaluations of fictitious groups but not in evaluations of well-known social groups. To further corroborate this conclusion, we conducted a combined MANOVA on implicit and explicit measures of both Experiments 4 and 5, providing us with sufficient power for detecting small effects (i.e., power = .85 to detect an effect size of d = 0.20). This analysis revealed the significant instruction effect on implicit measures and explicit measures in the Niffites/Luupites group, $F(3,576) = 36.93$, $p < .001$, whereas an instruction effect was not observed in the Blacks/Whites group, $F(3,720) = 0.26$, $p = .85$. We conclude that AA instructions influence evaluations of fictitious social groups, but not of well-known social groups.

GENERAL DISCUSSION

In five experiments, we compared evaluations of stimuli that participants were instructed to either approach or avoid. Our data show that typical AA training effects (i.e., a preference for approached stimuli over avoided stimuli) can be observed even if participants do not have to perform the AA actions. Specifically, when participants were instructed to approach nonwords or the names of members of fictitious social groups, their evaluations of these nonwords or social groups were more positive than evaluations of stimuli they were instructed to avoid. These findings were consistently observed across five experiments when evaluations were measured with an explicit self-report measure as well as when two different implicit measures were used (IAT and evaluative priming), suggesting that these effects were not due to measurement-related factors or demand compliance. In addition, our data revealed a clear boundary condition for effects of AA instructions. That is, consistent evidence was found that AA instructions are not sufficient to change evaluations of well-known social groups.

As we noted in the introduction, demonstrating instruction-based AA effects raises a number of issues for future research. Most notably, it highlights the need to carefully separate the effects of instructions and actual practice and to document when and how both instructions and practice exert an impact on stimulus evaluations. However, we believe that the current results already have a number of theoretical and practical implications. In the remainder of this section, we first discuss both the implications of our finding that AA instructions do influence the (implicit) evaluation of novel attitude objects. Afterwards, we discuss the implications of the fact that AA instructions do not seem to influence the evaluation of well-known attitude objects.

The presence of instruction-based AA effects for novel stimuli

Implications for theories of approach-avoidance training

From an embodiment perspective, it has been argued that performing AA actions activates congruent motivational orientations. As such, repeatedly approaching or avoiding a stimulus is assumed to result in the gradual formation of an association between the activated motivational state and the stimulus representation, which leads to a change in the evaluation of the stimulus (Cacioppo et al., 1993). Whereas early motivational accounts assumed that AA behavior activates motivational systems of AA via the proprioceptive feedback that this behavior produces, this strong embodiment view (i.e., the view that embodiment effects require actual proprioceptive feedback from motor behavior) has been largely abandoned (see Neumann & Strack, 2000). Theories of embodied cognition currently assume that AA effects result from an interplay between perceptual and motor representations and representations of the self (Markman & Brendl, 2005). This is in line with recent evidence suggesting that the automatic link from evaluation to AA behavior crucially involves representations that encode distance change (see Krieglmeyer, De Houwer & Deutsch, 2013), rather than the activation of specific muscle groups and that AA training effects have been observed when the AA training task involved repeatedly pressing a button to increase or decrease the distance between a stick figure and the stimulus (e.g., Huijding et al., 2013; Woud et al., 2013).

Performing these distance change responses is assumed to activate the motivational state of approach or avoidance, resulting in associative transfer of valence to the stimulus (Woud et al., 2013). In contrast to this view, our data suggest that AA effects do not necessitate performing any distance change responses nor perceiving distance change. Rather, merely telling participants that they will somehow have to approach or avoid a (type of) stimulus is sufficient to immediately change the evaluation of this stimulus. To account for instruction-based AA effects it seems necessary for motivational accounts to postulate that the mere thought of the action (i.e., the anticipation of the responses' sensory consequences) activates similar representations as the movement itself, thereby triggering the motivational system of AA and causing immediate changes in evaluation.

The most prominent alternative explanation for AA training effects is the common-coding account (Eder & Klauer, 2009; Hommel, Müsseler, Aschersleben & Prinz, 2001), which implies that affective features of objects and actions are cognitively represented by means of identical affective codes. When actions are construed as having a specific evaluative meaning, affective codes are activated, and these affective codes can be transferred to stimuli that are processed in relation to this action. Though there is extensive evidence suggesting that the movement compatibility effect (i.e., the finding that participants perform AA movements faster in response to a stimulus with congruent valence) depends on cognitive representations of the action rather than on actual behavior (e.g., Eder & Rothermund, 2008), our results extend this research by providing evidence that AA training effects might also critically depend on action representations rather than the actual performance of behavior. Thus, finding instruction-based AA effects is compatible with the common-coding account because this theory assumes that affective action codes are activated also during action planning. As a result, merely receiving instructions about the action-stimulus link might lead to similar effects as AA training (i.e., biased affective processing of stimuli).

On the other hand, our findings also put important constraints on a common-coding account. First, our results suggest that typical AA effects can be observed even when there are no repeated occasions on which action and stimulus are presented concurrently. From the common-coding perspective, merely being instructed to approach a certain group would only briefly lead to shared codes between action and stimulus. To explain instruction-based AA effects this account needs to stipulate how these short-term overlaps can transfer to a completely new task. Second, because participants were not informed how they would have to perform this approach and avoidance behavior it is less likely that effects resulted from mentally simulated action (i.e., mental practice). This suggests that, if effects originate from shared valence codes, these valence codes must have been activated via the activation of task representations at an abstract level, rather than via representations of specific actions.

In sum, the observation of instruction-based AA effects challenges current views on AA training and puts important constraints on these theories. It highlights the possibility that (a) motivational or evaluative representations can be activated by the mere anticipation of an AA response rather than the actual execution of the response and (b) associations involving motivational or evaluative representations can be formed instantly as the result of instructions.

Implications for theories of attitude formation

Our results confirm that, in line with instruction-based EC effects, implicit and explicit evaluations can result not only from extended training but also from mere instructions about relations in the environment (e.g., Gast & De Houwer, 2012). These results cannot be easily explained by single-process association formation models or dual-process models that assume that (a) associations underlie implicit evaluations and (b) that these associations can only form gradually as the result of repeated experiences (e.g., Baeyens, Eelen, Crombez & Van den Bergh, 1992; Baeyens, Eelen & Crombez, 1995). Our findings are especially striking given that effects of AA training on implicit evaluations are typically interpreted as stemming from gradual changes in associations (e.g., Phills et al., 2011) that necessitate a substantial amount of training. Our data provide evidence that propositional information (at least partially) influences these effects that are considered prototypical examples of effects that result from automatic processing in an associative system (Strack & Deutsch, 2004).

We propose that our results fit best with single-process propositional models of attitude formation (e.g., De Houwer, 2009; Hughes et al., 2011). In these models, the impact of stimulus pairings on evaluative responses is mediated by the formation of propositions. It is assumed that propositions can be formed not only on the basis of experience, but also on the basis of instructions or inferences (De Houwer, 2009). Again, our findings constrain these models. Most importantly, we observed that AA instructions influenced not only explicit but also implicit evaluation. The latter finding can be explained only by single-process propositional models that allow for the automatic retrieval of propositional knowledge from memory (see Hughes et al., 2011, for a discussion).

Note, however, that our data do not fully exclude the possibility that associations underlie instruction-based AA effects. Some have argued that such a single pairing could induce an association between the stimulus on the one hand and a representation of the action and its evaluative meaning on the other hand (e.g., Fazio, 2007; Field, 2006). Likewise, dual-process models which claim that both associations and propositions underlie evaluations could explain instruction-based AA effects provided that propositional knowledge (as gained via instructions) can lead to the formation of associations (e.g., Gawronski & Bodenhausen, 2011).

In sum, the fact that AA instructions can lead to changes in explicit and implicit evaluation contradicts a significant subset of all possible single-process associative and dual-process models (i.e., those models which postulate that implicit evaluation is mediated by associations that form gradually as the result of many experiences). Our findings also put important constrains on any current or future model of attitude formation.

Implications for the role of impact awareness in attitude change

Our studies provide the first demonstration that instructions about a behavior-stimulus link can impact implicit and explicit evaluations. Importantly, in contrast with EC instructions, which contain clear information that stimuli will be paired with valenced stimuli (De Houwer, 2006), in AA instructions, valence is not explicitly mentioned. In practice-based AA training studies, participants typically state that they were unaware that approaching or avoiding a stimulus was designed to change their evaluation (e.g., Cacioppo et al., 1993; Kawakami et al., 2007). In line with this, in Experiments 4 and 5, we found that only a small number of participants (i.e., less than 21%) identified the link between the instructions and evaluation change. Moreover, when participants did not believe that the experiment was designed to change their attitudes, a robust instruction-based AA effect could still be observed. These data provide evidence that an instruction-based influence on liking does not necessitate impact awareness (i.e., knowledge of the impact that the instructions are supposed to have on their attitude). This is an important addition to previous research on effects of instructions.

The absence of instruction-based AA effects for well-known stimuli

An important limitation of AA instructions seems to be that changes in evaluations of well-known social groups cannot be readily induced through this procedure. In contrast, a number of practice-based AA training studies have demonstrated an effect on implicit evaluations of well-known stimuli (Amir, et al., 2013; Phills et al., 2011; Wiers et al., 2011; Kawakami et al., 2007, Jones et al., 2013). If instructions are indeed insufficient to change established evaluations this would seriously constrain practical implications (e.g., the possibility of changing maladaptive evaluations) of AA instructions. However, to account for this apparent difference between instruction-based and practice-based AA effects, a number of explanations can be proposed. First, the difference could result from procedural details that differed between the current study and practice-based AA training studies, other than that the AA behavior was not performed. For example, in comparison with the study by Kawakami and colleagues, we instructed participants to approach or avoid more abstract stimuli, that is, the names of group members. This type of stimulus might be less relevant for the motivational circuit of AA behavior than the more concrete stimuli that were used in Kawakami's study (i.e., faces of members of the social groups). Note, however, that this explanation is at odds with our observation of large instruction effects for abstract novel stimuli across five experiments.

Second, changes in evaluations of well-known stimuli might be more difficult to obtain than changes in evaluations of novel stimuli (see also Hofmann et al., 2010), and AA instructions might simply not be potent to produce such changes. Association formation models typically suggest that changing evaluations of well-known stimuli is more difficult because association formation is easier than changing previous associations (Gregg et al., 2006). Alternatively, in line with a propositional point of view, it has been argued that adding new propositions to an existing set of propositions might result in less attitude change than creating new propositions (Gawronski & Bodenhausen, 2006). Therefore, it is possible that practice-based AA training does produce effects on evaluations of well-known stimuli because, in addition to propositional knowledge about stimulus-action contingencies, it adds something to the effects (e.g., it gives the new association or proposition more power due to the repeated experience). However, two types of evidence go against this explanation. First, the effect size of instruction-based AA effects on novel stimuli was larger than typical practice-based AA training effects, suggesting that, at least for this type of stimulus, practice-based AA training is not more potent to produce effects. Second, we observed that participants' implicit evaluations were more positive for Luupites' than for Niffites' names in Experiment 4. This observation is consistent with previous evidence suggesting that humans tend to evaluate stimuli automatically and as a result, novel stimuli are very often initially valenced (Perugini, Richetin & Zogmaister, 2012). Consequently, these findings propose that instruction-based AA effects do not necessitate novel stimuli that are fully neutral.

Third, for stimuli that have already acquired a specific valence, such as names of Black and White people, AA instructions might lead to changes in evaluations that are *opposite* to typical AA training effects, in addition to typical effects of AA training. Indeed, AA training can sometimes produce atypical effects (i.e., increased liking for avoided stimuli and reduced liking for approached stimuli), depending on the type of stimuli that participants approach or avoid (see Centerbar & Clore, 2006). These effects have been explained by postulating that approaching a disliked object can result in aversive motivational incongruity, thereby exacerbating the dislike. In the present study, participants who were instructed to approach names of Black people might experience this incongruity, resulting in increased prejudice. Compared with practice-based AA training this process might have been stimulated to a larger extent because participants were explicitly instructed to remember the instructions carefully, thereby making the stimulus-action relation more salient (Jones et al. ,2013). The instruction based manipulation might thus have (partially) counteracted the effect that is typically observed in AA training studies on prejudice, namely that participants who approach the prejudiced group display less prejudice (Kawakami et al., 2007; Phills et al., 2011). This account can also provide an explanation for the observation that, in Experiments 4 and 5, implicit and explicit evaluations had significantly lower correlations for participants instructed to approach Blacks than for participants instructed to approach Whites. Such a pattern could be expected if evaluations were changed in the approach Blacks group in line with their prejudice for some participants, but contrary to their prejudice for other people, and if implicit and explicit measures detect these changes to a different extent. Interestingly, in Experiment 4, a similar correlational pattern was observed for participants in the Niffites/Luupites condition. That is, correlations between implicit and explicit evaluations were non-significant only for participants who were instructed to approach the group that is liked less (i.e., the Niffites). This raises further questions about the boundary conditions of AA training and AA instructions in particular that should be addressed in future studies.

Concluding remarks

In this study we found evidence that instructions to approach or avoid can influence both implicit and explicit evaluations. These findings provide insight into the mechanisms underlying effects of AA training and open up important new questions about when and how avenues for understanding how evaluations can be formed and changed by means of instructions and actual AA training. Our results already reveal an important boundary condition of AA instruction effects, namely that AA instructions seem to have little effect on the evaluation of verbal information. However, explanations for this effect need to take into account that changes in liking for well-known groups were not easily induced with AA instructions in this study. Future research should investigate effects for evaluations of other well-known stimuli and provide a direct comparison between instruction-based and practice-based AA effects to distinguish the mechanisms that underlie effects of AA training.

REFERENCES

Amir, N., Kuckertz, J. M., & Najmi, S. (2013). The effect of modifying automatic action tendencies on overt avoidance behaviors. *Emotion, 13,* 478-484. doi: 10.1037/a0030443

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, 27*, 279-287.

Baeyens, F., Eelen, P., & Crombez, G. (1995). Pavlovian associations are forever: On classical conditioning and extinction. *Journal of Psychophysiology, 9*, 127-141. doi: 10.1016/0005-7967(92)90136-5

Cacioppo, J.T., Marshall-Goodell, B.S., Tassinary, L.G. & Petty, R.E. (1992). Rudimentary determinants of attitudes: Classical conditioning is more effective when prior knowledge about the attitude stimulus is low than high. *Journal of Experimental Social Psychology, 28,* 207-233. doi: 10.1016/0022-1031(92)90053-M

Cacioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary determinants of attitudes. II: Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology, 65*, 5-17. doi: 10.1037//0022-3514.65.1.5

Centerbar, D. & Clore, G.L. (2006). Do approach-avoidance actions create attitudes? *Psychological Science, 17*, 22-29. doi: 10.1111/j.1467-9280.2005.01660.x

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin, 25*, 215-224. doi: 10.1177/0146167299025002007

Cohen, J. (1992). "A power primer". *Psychological Bulletin, 112*, 155–159. doi:10.1037/0033-2909.112.1.155

Cohen, J., & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum.

Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for white Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology, 36*, 316-328, doi: 10.1006/jesp.1999.1418

De Houwer, J. (2006). Using the Implicit Association Test does not rule out an impact of conscious propositional knowledge on evaluative conditioning. *Learning and Motivation, 37*, 176-187. doi: 10.1016/j.lmot.2005.12.002

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

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

Dhont, K., & Van Hiel, A. (2009). We must not be enemies: Interracial contact and the reduction of prejudice among authoritarians. *Personality and Individual Differences 46,* 172-177. doi: 10.1016/j.paid.2008.09.022

Eder, A. B., & Klauer, K. C. (2009). A common-coding account of the bi-directional evaluation-behavior link. *Journal of Experimental Psychology: General, 138,* 218-235. doi: 10.1037/a0015220

Eder, A. B., & Rothermund, K. (2008). When do motor behaviors (mis)match affective stimuli? An evaluative coding view of approach and avoidance reactions. *Journal of Experimental Psychology: General, 137*, 262-281. doi: 10.1037/0096-3445.137.2.262

Elliot, A. J. (2006). The hierarchical model of approach-avoidance motivation. *Motivation and Emotion*, *30*, 111–116. doi: 10.1007/s11031-006-9028-7

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229–238. doi: 10.1037//0022-3514.50.2.229

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, 25, 603-637. doi: 10.1521/soco.2007.25.5.603

Field, A. P. (2006). *Is conditioning a useful framework for understanding the development and treatment of phobias?* Clinical Psychology Review, 26, 857-875. doi: 10.1016/j.cpr.2005.05.010

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

Gast, A., & De Houwer, J. (2013). The influence of extinction and counterconditioning instructions on evaluative conditioning effects. *Learning and motivation, 44*, 312-325. doi: 10.1016/j.lmot.2013.03.003

Gast, A., De Houwer, J., & De Schryver, M. (2012). Evaluative conditioning can be modulated by memory of the CS-US Pairings at the time of testing. *Learning and Motivation, 43*, 116- 126. doi: 10.1016/j.lmot.2012.06.001

Gawronski, B., & Bodenhausen, G. V. (2006). Associate and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*, 692-731. doi: 10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology, 44,* 59-127. doi: 10.1016/B978-0-12-385522-0.00002-0

Gawronski, B., Hofmann, W., & Wilbur, C. (2006). Are "implicit" attitudes unconscious? *Consciousness & Cognition, 15,* 485-499. doi: 10.1016/j.concog.2005.11.007

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*, 1464–1480. doi: 10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of*

Personality and Social Psychology, 85, 197–216. doi: 10.1037/0022-3514.85.2.197

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,* 1-20. doi: 10.1037/0022-3514.90.1.1

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

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences, 8*, 494-500. doi: 10.1016/j.tics.2004.08.007

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24*, 849-878. doi: 10.1017/S0140525X01000103

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

Huijding, J., Muris, P., Lester, K. J., Field, A. P., & Joosse, G. (2011). Training children to approach or avoid novel animals: Effects on self-reported attitudes and fear beliefs and information-seeking behaviors. *Behaviour Research and Therapy, 49*, 606-613. doi: 10.1016/j.brat.2011.06.005

Iyengar, S., Messing, S., Hahn, K. S., Banaji, M., & Dial, C. (2011). Explicit and implicit racial attitudes: A test of their convergent and predictive validity. *APSA 2011 Annual Meeting* Paper. Retrieved from http://ssrn.com/abstract=1901991

Jones, C. R., Vilensky, M. R., Vasey, M. W., & Fazio, R. H. (2013). Approach behavior can mitigate predominately univalent negative attitudes: Evidence regarding insects and spiders. *Emotion, 13*, 989-996. doi: 10.1037/a0033164

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder: Improving implicit racial evaluations and interracial interactions through approach behaviors. *Journal of Personality and Social Psychology, 92*, 957-971. doi: 10.1037/0022-3514.92.6.957

Krieglmeyer, R., De Houwer, J., & Deutsch, R. (2013). On the Nature of Automatically Triggered Approach-Avoidance Behavior. *Emotion Review, 5*, 280–284. doi: 10.1177/1754073913477501

Laham, S. M., Kashima, Y., Dix, J., Wheeler, M., & Levis, B. (in press). Elaborated contextual framing is necessary for action-based attitude acquisition. *Cognition & Emotion.* doi: 10.1080/02699931.2013.867833

Markman, A. B., & Brendl, C. M. (2005). Constraining theories of embodied cognition. *Psychological Science, 16*, 6-10. doi: 10.1111/j.0956-7976.2005.00772.x

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

Neumann, R., Förster, J. & Strack, F. (2003). Motor compatibility: The bidirectional link between behavior and evaluation. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 7-49). Mahwah, NJ: Lawrence Erlbaum.

Neumann, R., & Strack, F. (2000). Approach and avoidance: The influence of proprioceptive and exteroceptive cues on encoding of affective information. *Journal of Personality and Social Psychology, 79*, 39–48. doi: 10.1037/0022-3514.79.1.39

Niedenthal, P.M., Barsalou, L.W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review, 9*, 184-211. doi: 10.1207/s15327957pspr0903_1

Ottaway, S. A., Hayden, D. C., & Oakes, M. A. (2001). Implicit attitudes and racism: Effects of word familiarity and frequency on the implicit association test. *Social Cognition,* 19*,* 97-144. doi: 10.1521/soco.19.2.97.20706

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. doi: 10.1016/j.lmot.2012.02.002

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. doi: 10.1037/a0022159

Priester, J. R., Cacioppo, J. T., & Petty, R. E. (1996). The influence of motor processes on attitudes toward novel versus familiar semantic stimuli. *Personality and Social Psychology Bulletin,* 22, 442-447. doi: 10.1177/0146167296225002

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36. URL http://www.jstatsoft.org/v48/i02/

Slepian, M.L., Young, S.G., Rule, N.O., Weisbuch, M. & Ambady, N. (2012). Embodied impression formation: Social judgments and motor cues to approach and avoidance. *Social Cognition, 30*, 232-240.

Smith, C. T., De Houwer, J., & Nosek, B. (2013). Consider the source: persuasion of implicit evaluations is moderated by source credibility. *Personality and Social Psychology Bulletin,* 39, 193-205. doi: 10.1177/0146167212472374

Solarz, A.K. (1960). Latency of instrumental responses as a function of compatibility with the meaning of eliciting verbal signs. *Journal of Experimental Psychology, 59*, 239–45. doi: 10.1037/h0047274

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*, 220-247. doi: 10.1207/s15327957pspr0803_1

Taylor, C. T., & Amir, N. (2012). Modifying automatic approach action tendencies in individuals with elevated social anxiety symptoms. *Behaviour Research and Therapy, 50*, 529-536. doi: 10.1016/j.brat.2012.05.004

Teige-Mocigemba, S., & Klauer, K. C. (2008). 'Automatic' evaluation? Strategic effects on affective priming. *Journal of Experimental Social Psychology, 44*, 1414-1417. doi: 10.1016/j.jesp.2008.04.004

Vandenbosch, K., & De Houwer, J. (2011). Failures to induce implicit evaluations by means of approach-avoid training. *Cognition and Emotion, 25,* 1311-1330. doi: 10.1080/02699931.2011.596819

Wennekers, A. M. (2013). Embodiment of Prejudice: The Role of the Environment and Bodily States (doctoral dissertation). Retrieved from http://www.annemariewennekers.com

Wiers, R.W., Eberl, C., Rinck, M., Becker, E. & Lindenmeyer, J. (2011). Re-training automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science, 22,* 490-497. doi: 10.1177/0956797611400615

Woud, M. L., Becker, E. S., & Rinck, M. (2008). Implicit evaluation bias induced by approach and avoidance. *Cognition and Emotion, 22*, 1187-1197. doi: 10.1080/02699930801906843

Woud, M. L., Maas, J., Becker, E.S., & Rinck, M. (2013). Make the manikin move: Symbolic approach-avoidance responses affect implicit and explicit face evaluations. Journal of Cognitive Psychology, 25, 738-744. doi: 10.1080/2044

APPENDIX

Additional Analyses for Experiment 4

First, we examined whether the instruction effect that we observed for participants in the Niffites/Luupites condition was significantly larger compared with participants in the Blacks/Whites condition. We performed an Instructions (approach Niffites/Whites vs approach Luupites/Blacks) x Condition (Niffites/Luupites vs Whites/Blacks) Multivariate ANOVA (MANOVA) on IAT and explicit measure scores. In addition to the main effect of Instructions, $F(3,619) = 23.39$, $p < .001$, and Condition, $F(3,619) = 88.87$, $p < .001$, we observed a significant interaction effect, $F(3,619) = 22.96$, $p < .001$. This interaction effect indicated a larger instruction effect for participants in the Niffites/Luupites condition than for participants in the Whites/Blacks condition and was observed on the IAT score and on both explicit measures, *p*s < .001.

Second, we performed a correlational analysis for participants in both conditions. The scores on the explicit and implicit measures were significantly correlated for participants in the Niffites/Luupites condition (i.e., warmth and liking score: $r[271] = .84$, $p < .001$; IAT score and warmth score: $r[271] = .35$, $p < .001$; IAT score and liking score: $r[271] = .39$, $p < .001$) as well as for participants in the Whites/Blacks condition (i.e., explicit measures: $r[354] = .55$, $p < .001$; IAT score and warmth score: $r[354] = .16$, $p = .002$; IAT score and liking score: $r[354] = .15$, $p = .006$). Additional analyses revealed that correlations between implicit and explicit measures were significantly larger for participants who were instructed to approach Whites than for participants who were instructed to approach Blacks and that correlations were significantly larger for participants who were instructed to approach Luupites than participants who were instructed to approach Niffites. Because evidence suggests that AA training impacts implicit prejudice to a different degree for participants who approach the prejudiced group compared to participants who approach the group that participants belong to (e.g., Kawakami et al., 2007; Wennekers, 2013), we performed separate correlations for the participants who had been instructed to approach Whites and participants who had been instructed to approach Blacks. This analysis revealed that implicit and explicit prejudice measures were correlated in the approach Whites condition (i.e., IAT score and warmth score: $r[176] = .23$, $p = .002$; IAT score and liking score: $r[176] = .26$, $p = .001$), but not in the approach Blacks condition (i.e., IAT score and warmth score: $r[178] = .11$, $p = .16$; IAT score and liking score: $r[178] = .04$, $p = .56$). We subsequently compared the correlational coefficients for the two groups (Cohen & Cohen, 1983). The difference between the two groups' correlational coefficients of IAT and liking score was statistically significant, $Z = 2.11$, $p = .035$. The difference between the correlational coefficients of IAT and warmth score was not significant, $Z = 1.15$, $p = .25$. Performing separate correlations for participants instructed to approach Niffites and participants instructed to approach Luupites also revealed a different pattern. Implicit and explicit measures were correlated in the approach Luupites condition (i.e., IAT score and warmth score: $r[131] = .36$, $p < .001$; IAT score and liking score: $r[131] = .40$, $p < .001$), but in the approach Niffites condition only IAT score and liking score were significantly correlated ($r[140] = .18$, $p = .037$), whereas IAT score and warmth score were not ($r[140] = .10$, $p = .23$). The difference between the two groups' correlational coefficients was statistically significant (IAT and warmth score: $Z = 2.25$, $p = .024$; IAT and liking score: $Z = 1.97$, $p = .049$).

Third, we compared the instruction effect for participants who indicated that they thought that a purpose of the experiment was to change their attitudes towards the social groups (Niffites/Luupites: 57.6%, Whites/Blacks: 31.9%) and participants who did not indicate this. Additionally, we included the time when participants first believed that a purpose of the experiment was to change their attitudes (Niffites/Luupites: during the AA instructions: 20.7%, during the IAT: 31.4%, after the IAT or never: 47.9%; Whites/Blacks: during the AA instructions: 16.1%, during the IAT: 12.4%, after the IAT or never: 71.5%) in the analysis as well as participants' ratings about their belief that the experiment could have changed their attitudes (Niffites/Luupites: $M = 2.2$, $SD = 1.1$; Whites/Blacks: $M =$

1.6, *SD* = 0.9). Most importantly, these analyses revealed no impact of the first and second hypothesis awareness factor. For participants in both the Niffites/Luupites and Blacks/Whites conditions, main and interaction effects of the first hypothesis awareness factor (i.e., whether participants thought that a purpose of the experiment was to change their attitudes towards the social groups) and timing factor were not significant, *p*s > .44. Participants in the Niffites/Luupites condition still displayed an instruction effect if they did not think that a purpose of the experiment was to change their attitudes on the IAT score, $t(113) = 5.00$, $p < .001$, $d = 0.97$, and on both explicit measures, *p*s < .001. However, participants' belief ratings (i.e., rating about whether the experiment changed their attitude) were related to the instruction effect, such that the preference for the approached group was larger for participants who had higher belief ratings. This effect was observed only for participants in the Niffites/ Luupites condition, and only on liking ratings, $F(1,268) = 9.42$, $p = .002$, and warmth ratings, $F(1,268) = 6.85$, $p = .009$, but not IAT scores, $F(1,268) = 0.02$, $p = .88$.

Finally, we performed mediational analyses with the lavaan package (version 0.5-16; Rosseel, 2012) to investigate the relationship between implicit and explicit evaluative change. In the Niffites/Luupites condition we observed that changes in implicit evaluations were partly mediated by corresponding changes in explicit evaluations, $Z = 3.55$, $p < .001$. However, the effect of AA instructions on implicit evaluations remained significant after controlling for explicit evaluations, $Z = 6.17$, $p < .001$. Similarly, changes in explicit evaluations were partly mediated by corresponding changes in implicit evaluations, Z = 3.81, p < .001, yet the effect of AA instructions on explicit evaluations remained significant after controlling for implicit evaluations, $Z = 3.98$, $p < .001$. In the Whites/Blacks condition no direct or indirect effects of instructions were observed, *Z*s < 1.06, *p*s > .27.

Additional Analyses for Experiment 5

First, we performed an instructions (approach Niffites/Whites vs. approach

Luupites/Blacks) x condition (Niffites/Luupites vs. Whites/Blacks) MANOVA on evaluative priming and explicit measure scores. In addition to the main effect of instructions, $F(3,527) = 19.26$, $p < .001$, we observed a significant interaction effect, $F(3,527) = 15.32$, $p < .001$. This interaction effect indicated a larger instruction effect for participants in the Niffites/Luupites condition than for participants in the Whites/Blacks condition and was observed on the evaluative priming score and on both explicit measures, *p*s < .005.

Second, a correlational analysis of the implicit and explicit measures revealed that the scores on the implicit and explicit measures were significantly correlated for participants in the Niffites/Luupites condition (i.e., warmth and liking score: $r[257] = .95$, $p < .001$; evaluative priming score and warmth score: $r[257] = .15$, $p = .018$; evaluative priming score and liking score: $r[257] = .13$, $p = .039$). For participants in the Whites/Blacks condition correlations were significant, except for the correlation between warmth score and the implicit measure score (i.e., explicit measures: $r[276] = .57$, $p < .001$; evaluative priming score and warmth score: $r[276] = .06$, $p = .34$; evaluative priming score and liking score: $r[276] = .14$, $p = .023$). In line with Experiment 1, implicit evaluations were significantly correlated with explicit evaluations in the approach Whites condition (i.e., evaluative priming score and warmth score: $r[145] = .22$, $p = .007$; evaluative priming score and liking score: $r[145] = .19$, $p = .021$), but not in the approach Blacks condition (i.e., evaluative priming score and warmth score: $r[131] = -.09$, $p = .29$; evaluative priming score and liking score: $r[131] = .07$, $p = .40$). The difference between the two groups' correlational coefficients was significant for the IAT and warmth score, $Z = 2.58$, $p = .010$, but not for the IAT and liking score, $Z = 1.00$, $p = .32$. Separate correlations for participants instructed to approach Niffites and participants instructed to approach Luupites did not reveal significant differences between correlations, *p*s > .55.

Third, we compared the instruction effect for participants who thought that a purpose of the experiment was to change their attitudes towards the social groups (Niffites/Luupites: 46.9%, Whites/Blacks: 20.7%) and participants who did not believe this. Additionally, we included the time when participants first believed that a purpose of the experiment was to change their attitudes (Niffites/Luupites: during the instructions: 20.6%, during the evaluative priming task: 22.2%, after the evaluative priming task or never: 57.2%; Whites/Blacks: during the instructions: 11.6%, during the evaluative priming task: 7.2%, after the evaluative priming task or never: 81.2%) in the analysis as well as participants' ratings about their belief that the experiment could have changed their attitudes (Niffites/Luupites: M = 2.0 , SD = 1.1; Whites/Blacks: M = 1.5, SD = 0.7). For participants in both the Niffites/Luupites and Blacks/Whites conditions, main and interaction effects including the first two hypothesis awareness factors were not significant, ps > .49. Also, participants in the Niffites/Luupites condition still displayed an instruction effect on the evaluative priming score, $t(133) = 2.62$, p = .010, d = 0.45, and on both explicit measures, ps < .001, if they had indicated that they did not think that a purpose of the experiment was to change their attitudes. However, participants' belief ratings were related to the instruction-based AA effect for participants in the Niffites/ Luupites condition. We observed an effect of belief on liking ratings, $F(1,321) = 3.88$, p = .050, warmth ratings, $F(1,321) = 7.44$, p = .007, and evaluative priming scores, $F(1,254) = 7.99$, p = .005.

Finally, mediational analyses indicated that changes in implicit evaluations in the Niffites/Luupites condition were not significantly mediated by corresponding changes in explicit evaluations, Z = 0.81, p = .42. The effect of AA instructions on implicit evaluations remained significant after controlling for explicit evaluations, Z = 3.60, p < .001. Similarly, changes in explicit evaluations weren't significantly mediated by corresponding changes in implicit evaluations, Z = 0.80, p = .42, and the effect of AA instructions on explicit evaluations remained significant after controlling for implicit evaluations, Z = 6.53, p < .001. In the Whites/Blacks condition no direct or indirect effects of instructions were observed, Zs < 0.67, ps > .50.

CHAPTER 3

INSTRUCTING IMPLICIT PROCESSES: WHEN INSTRUCTIONS TO APPROACH OR AVOID INFLUENCE IMPLICIT BUT NOT EXPLICIT EVALUATION[1]

Previous research has shown that linking approach or avoidance actions to novel stimuli through mere instructions causes changes in the implicit evaluation of these stimuli even when the actions are never performed. In two high-powered experiments (total N = 1147), we examined whether effects of approach-avoidance instructions on implicit evaluations are mediated by changes in explicit evaluation. Participants first received information about the evaluative properties of two fictitious social groups (e.g., Niffites are good; Luupites are bad) and then received instructions to approach one group and avoid the other group. We observed an effect of approach-avoidance instructions on implicit but not explicit evaluations of the groups, even when these instructions were incompatible with the previously obtained evaluative information. These results indicate that approach-avoidance instructions allow for unintentional changes in implicit evaluations. We discuss implications for current theories of implicit evaluation.

---

[1] Based on Van Dessel, P., De Houwer, J., Gast, A., Smith, C. T., & De Schryver, M. (2016). Instructing Implicit Processes: When Instructions to Approach or Avoid Influence Implicit but not Explicit Evaluation. *Journal of Experimental Social Psychology, 63*, 1-9. doi: 10.1016/j.jesp.2015.11.002

INTRODUCTION

The way in which humans evaluate stimuli as good or bad has long been a central research topic in various sub-disciplines of psychology (Allport, 1935). In contemporary research on evaluations, researchers often contrast deliberate, explicit evaluations and spontaneous, implicit evaluations (see De Houwer, 2009a; Gawronski & Bodenhausen, 2011). Typically, theorists have postulated distinct underlying processes, with explicit evaluations resulting from belief-based processes that involve the validation of propositional information, and implicit evaluations being the product of processes involving the automatic activation of associations in memory (Gawronski & Bodenhausen, 2006, 2011, 2014).

Given the unique relation between implicit evaluations and behavior (Greenwald, Poehlman, Uhlmann, & Banaji, 2009), it is vital to understand how implicit stimulus evaluations are acquired and can be changed. Because implicit evaluation is traditionally attributed to the activation of associations between representations in memory and because associations are typically thought to develop gradually over many experiences, it is sometimes assumed that implicit evaluations of stimuli arise exclusively as the result of repeated experiences, such as recurrent pairings of physical stimuli (Rydell & McConnell, 2006). Evaluative conditioning (EC) research provides ample evidence that changes in the implicit evaluation of a stimulus (conditioned stimulus; CS) occur when it is paired with a valenced stimulus (unconditioned stimulus; US; for a review see Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). Moreover, research on approach and avoidance (AA) training has suggested that changes in implicit evaluations can be obtained by pairing a stimulus with a valenced action (i.e., approach or avoidance). Typically, the repeated approaching of one stimulus and avoiding of another stimulus leads to more positive implicit evaluations for the former stimuli (e.g., Kawakami, Phills, Steele, & Dovidio, 2007; Woud, Maas, Becker, & Rinck, 2013; but see Vandenbosch & De Houwer, 2011).

Recent research has, however, shown that implicit evaluations change even when pairings are not experienced directly, but are implied by the verbal presentation of relational information via instructions. For instance, studies on instructed EC have shown that changes in the implicit evaluation of a CS occur when verbal instructions link a CS with a valenced US even when the CS-US pairings are not experienced directly (De Houwer, 2006; Gast & De Houwer, 2012). Similarly, in a recent study we observed typical AA training effects when participants did not actually perform AA actions, but were merely instructed that they would later have to perform these actions (Van Dessel, De Houwer, Gast, & Smith, 2015). That is, participants who received instructions to approach one fictitious social group (e.g., Niffites) and avoid another fictitious social group (e.g., Luupites) showed a preference for the former group both on implicit measures (i.e., the Implicit Association Test, IAT, Greenwald, McGhee, & Schwarz, 1998; and the evaluative priming task, Fazio, Sanbonmatsu, Powell, & Kardes, 1986) and explicit measures of evaluation.

These findings pose a challenge to associative and dual-process models of evaluation which assume that implicit evaluations result from the gradual formation of associations in memory as the result of actual pairings (Smith & DeCoster, 2000; Rydell & McConnell, 2006). In contrast, contemporary dual-process models in which association formation processes can interact with propositional learning processes allow one to explain effects of instructions on implicit evaluations. For instance, the Associative-Propositional Evaluation (APE) model (Gawronski & Bodenhausen, 2006; 2011; 2014) postulates that associations may sometimes arise as the result of the generation and validation of propositions. More specifically, when people determine in a propositional manner that a stimulus is either positive or negative this may instigate the proactive construction of new associations between representations of the stimulus and representations of positivity or negativity. As a result, any information that allows participants to consciously entertain the proposition that a stimulus is positive or negative may influence implicit evaluations. In line with this idea, changes in implicit evaluations have been observed when participants are provided with information about the valenced properties of a stimulus (Castelli, Zogmaister, Smith & Arcuri, 2004; Gregg, Seibt, & Banaji, 2006; Whitfield & Jordan, 2009; Cone & Ferguson, 2015).

Importantly, these models predict a specific pattern of mediation such that instruction effects on explicit evaluation should mediate effects on implicit evaluation (see Gawronski & Bodenhausen, 2006; Case 4). That is, instructions should first influence whether participants consider a stimulus positive or negative (which is reflected in explicit evaluations) before this may lead to the formation of novel associations (which is reflected in implicit evaluations). Support for this idea was found by Whitfield and Jordan (2009), who observed that receiving information about the behavior of unknown individuals caused changes in implicit evaluations of these individuals that were fully mediated by changes in explicit evaluations.

Contrasting this result, our previous study on AA instruction effects provided evidence that changes in explicit evaluations do not fully mediate effects of AA instructions on implicit evaluations. Statistical mediation analyses indicated that the impact of AA instructions on implicit evaluations was partly mediated by changes in explicit evaluations, but an effect remained after controlling for changes in explicit evaluation (Van Dessel et al., 2015). This is an intriguing finding because it suggests that mere (AA) instructions may sometimes cause unintentional changes in (implicit) stimulus evaluations. Instructions may have a direct effect on implicit evaluation (i.e., unmediated by changes in explicit evaluation) and may therefore cause changes in implicit evaluations even when participants do not consider the instructions as a valid basis for their (explicit) evaluation.

However, on the basis of the available evidence it is premature to conclude that AA instructions can influence implicit evaluation without any mediation by changes in explicit evaluation. Most importantly, our earlier AA instruction study (Van Dessel et al., 2015) included only statistical analyses of mediation. This measurement-of-mediation approach, however, is ultimately correlational in nature, and is thus problematic for establishing a causal chain (Spencer, Zanna, &

Fong, 2005). This is especially the case when examining patterns of mediation between implicit and explicit evaluations. When a manipulation affects both implicit and explicit measures of evaluation, the particular direction of the obtained mediation pattern is strongly influenced by the internal consistency of the employed measure (Gawronski & Bodenhausen, 2011). Moreover, when implicit and explicit evaluations are strongly correlated (as was the case in our previous study), this creates multicollinearity which inflates the standard error of all variables in the mediation model and compromises the estimation of the indirect effect (Alin, 2010). Hence, when examining mediation of implicit and explicit evaluations, it is strongly recommended to supplement statistical mediation analyses with experimental manipulations (De Houwer, Gawronski, & Barnes-Holmes, 2013). This is particularly true if, as in our case, a theoretical debate requires the precise understanding of the causal relation.

In the current studies, we used both a statistical and an experimental approach to test the extent to which the impact of AA instructions on implicit evaluation is mediated by changes in explicit evaluation. We manipulated the proposed mediating variable (i.e., changes in explicit evaluation) by providing participants with 'trait instructions' that should prevent an impact of AA instructions on explicit evaluation. In line with Gregg et al. (2006), we asked participants to imagine that the members of one fictitious social group had very positive traits and the members of another fictitious social group had very negative traits (e.g., Niffites are peaceful, civilized, benevolent, and law-abiding; Luupites are violent, savage, malicious, and lawless). Subsequently, participants received instructions to approach or avoid these social groups. Whereas trait instructions directly specify the evaluative properties of the social group, AA instructions only provide evaluative information if participants infer that the task to approach or avoid members of a group tells something about the evaluative properties of that group. Participants might rely on this inference when they have no other information about the evaluative properties of the group, but even then they will probably be aware that this inference rests on shaky grounds. Prior research indeed suggests that participants are likely to refrain from using information that has a low diagnostic validity (such as AA instructions) when more valid information (such as instructions about evaluative traits) is available (Lynch, 2005; Cone & Ferguson, 2015). For these reasons, we expected that participants who received trait instructions would not take the AA instructions into account when explicitly evaluating the stimuli. We examined whether, under these circumstances, AA instructions would still cause changes in implicit evaluation. That is, we examined whether an AA instruction effect on implicit evaluation would be observed not only in the absence of mediation by changes in explicit evaluation, but even when there is no impact on explicit evaluation. The latter result would not only confirm that AA instructions can have a direct effect on implicit evaluation (because mediation via changes in explicit evaluation can occur only if there are changes in explicit evaluation) but would also support the novel conclusion that this direct effect can arise even when participants do not have the intention to use the AA instructions for evaluating the stimuli.

If we would find that AA instructions influence implicit evaluation in the absence of (mediation by) changes in explicit evaluation, this is bound to have important theoretical implications. First, it would strongly constrain current and future models of (implicit) evaluation. For instance, it would contradict dual-process models that assume that (1) only directly experienced repeated pairings can influence implicit evaluations (Smith & DeCoster, 2000), and it would contradict dual-process models that assume that (2) instructions can only influence implicit evaluation via the mediation of explicit evaluation (Gawronski & Bodenhausen, 2006). To accommodate these findings, dual-process accounts would need to make additional assumptions (e.g., that strong associations can form as the result of a single pairing of a valenced word and a stimulus even in the absence of changes in explicit evaluation).

Finding an impact of AA instructions on implicit evaluation but not on explicit evaluation would also constrain single-process propositional models of evaluation (De Houwer, 2009b; 2014; Mitchell, De Houwer, & Lovibond, 2009). These models postulate that both implicit and explicit evaluations arise exclusively as the result of propositional processes. Prima facie, these models seem less equipped to explain dissociations between implicit and explicit evaluations (e.g., a change in implicit evaluation in the absence of a similar change in explicit evaluation). However, dissociations do not necessarily mean that different processes underlie these different types of evaluation. Rather, dissociations may arise because implicit and explicit measures of evaluation are differentially sensitive to the truth evaluation of propositional information. For example, when participants are told that a specific stimulus has to be approached, they might consider the possibility that this stimulus is good because it has to be approached. If this newly formed proposition can be activated automatically (e.g., in the sense of unintentional) then it may influence implicit evaluation even when the proposition is not considered valid (De Houwer, 2014). In contrast, explicit evaluation may be more contingent on the outcome of truth validation processes.

Second, finding an AA instruction effect on implicit but not explicit evaluation would provide valuable information about the mechanisms that specifically underlie the acquisition of evaluations by means of AA training, that is, by means of the repeated actual performance of approach and avoidance responses. Currently, there is ample evidence that training-based effects involve changes in implicit evaluation that are not mediated by changes in explicit evaluations (Gawronski & LeBel, 2008; Whitfield & Jordan, 2009). These findings have typically been interpreted as evidence that training directly influences processes of association formation. However, these effects might also reflect the acquisition of propositional information that specifically influences implicit evaluation (e.g., because it allows for the automatic activation of propositions) but not explicit evaluation (e.g., because the information is not considered a valid basis for evaluation). If we observe an impact of AA instructions on implicit but not explicit evaluations, this would support the idea that propositional information can indeed influence implicit evaluations independently of changes in explicit evaluation.

We conducted two experiments to investigate whether the impact of AA instructions on implicit evaluations is mediated by changes in explicit evaluation. In Experiment 1, half of the participants first received instructions that specified the traits of the fictitious social groups. Subsequently, participants received instructions to approach the names of members of one of the social groups and avoid the names of members of the second social group. For half of the participants, these AA instructions were supplemented with actual AA training. We then assessed implicit and explicit evaluations of the social groups. With this design, two tests are possible of the hypothesis that AA instructions allow for a direct influence on implicit evaluation. First, it can be tested whether AA instructions influence implicit evaluations even after statistically controlling for changes in explicit evaluations. Second, it can be tested if AA instructions influence implicit evaluations even if trait instructions prevent the effects of AA instructions on explicit evaluations. To investigate this issue, we supplemented standard significance tests with Bayesian analyses. Bayesian analyses were performed according to the procedures outlined by Rouder, Speckman, Sun, Morey, and Iverson (2009). These procedures provide a Bayes Factor (BF) that gives an indication of how strongly the data support either the null hypothesis (BF_0; reflecting the absence of a significant effect) or the alternative hypothesis (BF_1; reflecting the presence of a significant effect). BFs smaller than 1, between 1 and 3, between 3 and 10, respectively designate ‘no evidence’, ‘anecdotal evidence’, and ‘substantial evidence’, for either the null or the alternative hypothesis (Jeffreys, 1961). We examined whether, in the presence of trait instructions, AA instructions do not cause changes in explicit evaluation (i.e., analyses provide substantial evidence for the null hypothesis, $BF_0 > 3$) yet still cause changes in implicit evaluation (i.e., analyses provide substantial evidence for the alternative hypothesis, $BF_1 > 3$).

In Experiment 2, all participants received trait instructions and subsequently received either AA instructions that were compatible with these instructions (e.g., instructions to approach Niffites when participants had been asked to imagine that Niffites have positive traits), AA instructions that were incompatible with these instructions (e.g., instructions to avoid Niffites when participants had been asked to imagine that Niffites have positive traits), or no AA instructions. We examined whether changes in implicit evaluations arise in the absence of changes in explicit evaluations when AA instructions are compatible with the trait instructions (and thus strengthen the previously acquired evaluations) or when they are incompatible with the trait instructions (and thus revise the previously acquired evaluations).

EXPERIMENT 1

Method

Participants and Design

In Experiment 1, 1121 English-speaking volunteers participated online via the Project Implicit research website (https://implicit.harvard.edu). We employed a 2 (Presence of Trait instructions: yes, no) x 2 (Content of AA Instructions: approach Niffites, approach Luupites) x 2 (Presence of AA Training: yes, no) between-subjects design (Table 1). Data-exclusion involved removing participants who (a) did not fully complete all questions and tasks (257 participants; i.e., 22.9%), or (b) made at least one error on the memory questions that probed memory for valence or AA instructions (189 participants; i.e., 21.9 %).[1] After removing participants based on the previous two criteria, there were no additional participants who needed to be removed because of IAT error rates above 30% across the entire task, or above 40% for any one of the four critical blocks (Smith, De Houwer, & Nosek, 2013). Analyses were performed on the data of 675 participants (440 women, mean age = 32, *SD* = 13).

---

[1] We excluded participants with incorrect memory because we expected that, in line with previous results (Van Dessel et al., 2015), instructions would impact evaluations only if participants correctly remembered these instructions. Importantly, including the data from all participants in the analyses weakened the main effect of Content of AA Instructions and the main effect of Content of Trait instructions on implicit and explicit evaluations, but did not result in any shift in significance for any of the reported effects.

Table 1.

Experimental Design of Experiment 1.

3 between participants variables:			
Trait Instructions		AA instructions	AA training
Present (50 %)	Niffites are good (25%)	Approach Niffites (12.5%)	Yes (6.25%)
			No (6.25%)
		Approach Luupites (12.5%)	Yes(6.25%)
			No (6.25%)
	Niffites are bad (25 %)	Approach Niffites (12.5%)	Yes (6.25%)
			No (6.25%)
		Approach Luupites (12.5%)	Yes (6.25%)
			No (6.25%)
Absent (50%)		Approach Niffites (25%)	Yes (12.5%)
			No (12.5%)
		Approach Luupites (25%)	Yes (12.5%)
			No (12.5%)

Procedure

All participants were first familiarized with the two fictitious social groups (i.e., Luupites and Niffites). They read that all the names of Luupites have two consecutive vowels in them and end with "lup". Then they were shown two examples of Luupites' names (i.e., Loomalup, Ageelup). Subsequently, participants read that all the names of Niffites would contain two consecutive consonants and end with "nif." This statement was followed by two Niffites names (i.e., Borrinif, Kennunif).

Half of the participants were then given trait instructions. Similar to Gregg et al. (2006), participants were asked to imagine that these two social groups actually exist and to suppose that the two groups have very different characters. They were instructed that one group 'are very good people; they are peaceful, civilized, benevolent, and law-abiding, whereas the other group 'are very bad people; they are violent, savage, malicious, and lawless.' Participants were also instructed to suppose that the two groups consistently behave in ways that justified these descriptions when they interact with each other and with other groups. Participants were asked to try and keep clear in their minds which group is which and which group possesses which characteristics as they would later be asked questions about the groups. Half of the participants who received trait instructions learned that Niffites are good and Luupites are bad, whereas the other half received instructions that conveyed the idea that Luupites are good and Niffites are bad.

Subsequently, all participants received AA instructions. Half of the participants were told that they would have to approach each name of a Luupite and avoid each name of a Niffite. The other participants were given the opposite instruction. These AA instructions were followed by the information that we would later on explain exactly how they would be able to perform these actions, but that for now it was very important to remember which action they would have to perform with each type of name as they would need this information to complete the task successfully.

Following the AA instructions, only half of the participants actually performed the AA training task. This manipulation was orthogonal to (1) the manipulation of the content of trait instructions (Niffites are good and Luupites are bad / Niffites are bad and Luupites are good) and (2) the content of AA instructions (approach Niffites and avoid Luupites / avoid Luupites and approach Niffites). Participants in the AA training condition performed 80 trials of the AA training task in which 4 Niffites' names (i.e., Cellanif, Eskannif, Lebbunif, Zallunif) and 4 Luupites' names (i.e., Meesolup, Naanolup, Omeelup, Wenaalup) were each presented ten times. Participants pushed away names by pressing the up arrow on the keyboard (i.e., avoided) and pulled names towards them by pressing the down arrow on the keyboard (i.e., approached). A zoom effect enhanced the visual experience of approaching or avoiding; names that were avoided became smaller and moved off into the perceptual distance, whereas names that were approached became larger and appeared to move toward the participant. Only actions that were in line with the AA instructions were registered as correct and resulted in the zoom effect. Incorrect responses were not registered. Participants always had to perform the correct response to proceed to the following trial. The other half of the participants did not receive AA training and they were instructed that they would complete a reaction time task which would last approximately 10 minutes before they could start the AA task.

The reaction time task that followed was an IAT in which participants categorized positive words, negative words, and the names of members of both social groups into one of four categories: positive, negative, Niffites, or Luupites. The IAT followed the procedure described in more detail in Van Dessel et al. (2015). It consisted of three practice blocks and two experimental blocks. Participants began the IAT with 20 practice trials sorting the names of Niffites and Luupites and 20 practice trials sorting positive and negative stimuli. Next, participants completed 56 trials in which stimuli related to Niffites and positive shared a single response key and stimuli related to Luupites and negative shared a single response key (half of the participants completed the IAT in this way, while the other participants began by sorting Luupites and positive with the same key). Participants then practiced sorting Niffites and Luupites names with the response key assignment reversed for 40 trials and finally participants completed a second set of 56 trials in which Niffites shared a response key with negative and Luupites shared a response key with positive (or vice versa). If the participant made an error in categorizing, a red “X” appeared on the screen and the participant corrected their mistake in order to continue. Latencies were recorded until a correct response was made. IAT-scores were calculated using the D2-algorithm (Greenwald, Nosek, & Banaji, 2003) so that positive scores indicate a preference for Niffites over Luupites. The Spearman-Brown corrected split-half reliability of the IAT score, calculated on the basis of an odd-even split, was $r(675) = .84$.

After the implicit evaluation task, participants rated their liking of each of the social groups by answering two questions: “To what extent do you like Niffites/Luupites?” and “To what extent do you have warm feelings for Niffites and Luupites?”. Participants gave their ratings by selecting an option on a 9-point Likert scale (1= not warm/liked at all; 9 = completely warm/liked). Rating scores (i.e., warmth scores and liking scores) were calculated by subtracting the score rating for Luupites from the corresponding score rating for Niffites so that positive scores indicate a preference for Niffites over Luupites. Because of high internal consistency (Cronbach’s Alpha = .94), we collapsed these score ratings into one explicit evaluation score by averaging the respective scores. This explicit evaluation score correlated significantly with the IAT score, $r(673) = .43$, $p < .001$.

Finally, participants completed two types of manipulation check questions. The first question was completed only by participants who had received trait instructions. Participants were asked to remember which trait instructions were presented at the start of the study and to answer by selecting an option on a dropdown menu with “That Niffites are good and Luupites are bad”, “That Luupites are good and Niffites are bad”, and “I don’t remember” as possible answers. The next two questions asked what action they would have to perform (or had performed in the case of actual training) according to the instructions when the name of a Niffite/Luupite was presented. Participants answered by selecting an option on a dropdown menu with “Approach”, “Avoid” and “I don’t remember” as possible answers.

Results

We split up the analyses for participants who did not receive trait instructions and participants who did receive trait instructions to separately address (1) whether AA instruction and AA training effects on implicit evaluations are fully mediated by changes in explicit evaluations, and (2) whether AA instructions and AA training cause changes in implicit evaluations even when trait instructions are provided.

No trait instructions condition

We performed a 2 (Content of AA Instructions: approach Niffites, approach

Luupites) x 2 (Presence of AA Training: yes, no) analysis of variance (ANOVA) on the IAT scores. Because there was an unequal number of participants per condition (no AA training: $N = 96$ for approach Niffites, $N = 97$ for approach Luupites; AA training: $N = 84$ for approach Niffites, $N = 87$ for approach Luupites), we used type III sums of squares in this and all subsequent statistical analyses. The ANOVA revealed a main effect of Content of AA Instructions, $F(1,360) = 135.93$, $p < .001$. Participants who had been instructed to approach Niffites and avoid Luupites ($M = 0.13$, $SD = 0.43$) preferred Niffites more than participants who had been instructed to approach Luupites and avoid Niffites ($M = -0.27$, $SD = 0.55$), $d = 1.22$, 95% confidence interval (CI) [1.00, 1.45], $BF_1 > 10000$. Neither the main effect of Presence of AA Training nor the interaction with Content of AA Instructions was significant, Fs < 0.93, ps > .33.

An ANOVA on the explicit rating scores revealed a similar pattern. We observed only a main effect of Content of AA Instructions, $F(1,360) = 52.49$, $p < .001$, indicating that participants who had been instructed to approach Niffites and avoid Luupites preferred Niffites ($M = 0.52$, $SD = 1.63$) more than participants who had been instructed to avoid Niffites and approach Luupites ($M = -0.99$, $SD = 2.29$), $d = 0.76$, 95% CI [0.54, 0.97], $BF_1 > 10000$. We observed no main or interaction effects involving the Presence of AA Training factor, Fs < 1.33, ps > .24.

To investigate the extent to which changes in implicit evaluation are mediated by changes in explicit evaluations we performed mediation analyses with the LAVAAN package (version 0.5-16; Rosseel, 2012). We used the bootstrap method to estimate standard errors for the effects. Results indicated that changes in implicit evaluations were mediated by corresponding changes in explicit evaluations, both when participants received only AA instructions ($Z = 2.31$, $p = .021$), and when they received AA instructions and subsequent AA training ($Z = 2.03$, $p = .042$). Importantly, however, the AA effect on implicit evaluations remained significant after controlling for changes in explicit evaluations for participants without ($Z = 5.65$, $p < .001$) and with actual training ($Z = 8.78$, $p < .001$). Regression coefficients of the performed mediation analyses are provided in Appendix.

Trait instructions condition

To examine AA effects in the context of trait instructions we performed a 2 (Content of AA Instructions: approach Niffites, approach Luupites) x 2 (Presence of AA Training: yes, no) x 2 (Content of Trait Instructions: Niffites are good, Luupites are good) ANOVA on the IAT scores of participants who had received trait instructions. We included the Content of Trait Instructions factor to estimate the effect of trait instructions on evaluations and control for the variance attributable to this factor. We observed a main effect of Content of Trait Instructions, $F(1,303) = 183.27$, $p < .001$, indicating that participants preferred Niffites more when Niffites were presented as positive and Luupites as negative ($M = 0.23$, $SD = 0.48$) than when Niffites were presented as negative and Luupites as positive ($M = -0.45$, $SD = 0.46$), $d = 1.45$, 95% CI [1.20, 1.70], $BF_1 > 10000$. This analysis also revealed a main effect of Content of AA Instructions, $F(1,303) = 36.78$, $p < .001$, but this effect was qualified by an interaction effect of Content of AA Instructions x Presence of AA Training, $F(1,303) = 5.02$, $p = .026$ (Table 2). Importantly, a significant effect of Content of AA Instructions was observed for participants who had merely received AA instructions, $F(1,150) = 7.44$, $p = .007$, $d = 0.44$, 95% CI [0.12, 0.77], $BF_1 = 5.36$. This effect was larger for participants who had received additional AA training, $F(1,153) = 33.48$, $p < .001$, $d = 0.74$, 95% CI [0.41, 1.06], $BF_1 = 1961.39$. Finally, an interaction effect of Content of AA Instructions and Content of Trait Instructions, $F(1,303) = 13.22$, $p < .001$, indicated that the effect of Content of AA Instructions was stronger when trait instructions conveyed that Niffites are good and Luupites are bad than when trait instructions conveyed the opposite information.[2]

[2] This finding relates to the observation that, even in the absence of trait instructions, participants preferred Luupites over Niffites, and may indicate that AA effects are reduced if participants have clearly univalent positive or negative implicit evaluations (e.g., because they find Luupites' names more appealing and they learned that Luupites are positive). Please consult Jones, Vilensky, Vasey, and Fazio (2013), and Woud, Becker, Lange, and Rinck (2013) for reasons why stimuli that have a non-ambivalent valence might be less susceptible to AA effects.

Table 2.

Mean IAT and Explicit Scores in Experiment 1 for participants who received trait instructions as a function of Content of Trait Instructions and Content of AA Instructions.

	Content of Trait Instructions			
	Niffites good and Luupites bad		Niffites bad and Luupites good	
IAT score:	Approach Niffites	Approach Luupites	Approach Niffites	Approach Luupites
No AA training	0.42 (0.37)	0.05 (0.47)	-0.40 (0.48)	-0.40 (0.39)
AA training	0.49 (0.37)	-0.09 (0.46)	-0.39 (0.53)	-0.62 (0.39)
Explicit score:				
No AA training	3.20 (3.08)	2.27 (3.49)	-3.99 (3.51)	-3.11 (3.00)
AA training	1.64 (2.89)	1.49 (2.63)	-3.21 (3.51)	-2.79 (3.38)

Note. Standard deviations are in parentheses. Scores reflect a relative preference for Niffites over Luupites.

An ANOVA on the explicit rating scores revealed a main effect of Content of Trait Instructions, $F(1,303) = 222.10$, $p < .001$. This effect was qualified by an interaction effect with Presence of AA Training, $F(1,303) = 5.60$, $p = .019$, which indicated that the effect of trait instructions was smaller for participants who received AA training, $d = 1.47$, 95% CI [1.11, 1.83], than for participants who received no AA training, $d = 1.91$, 95% CI [1.53, 2.30].[3] Most importantly, we observed no main effect of Content of AA instructions, $F(1,303) = 0.01$, $p = .90$, $d = 0.05$, 95% CI [-0.18, 0.28]. The BF score provided substantial evidence in favor of *the null hypothesis* ($BF_0 = 7.19$). We also observed no other main or interaction effects, *F*s < 2.66, *p*s > .10.

Mediation analyses showed that changes in implicit evaluations were not

[3] One possible explanation for this is that participants who received actual training may have been distracted from the trait instructions (e.g., because there was a longer delay between receiving these instructions and completing the evaluative rating task) and therefore used these trait instructions to a lesser extent for their evaluative ratings. Receiving the trait instructions, however, still discouraged participants from considering the AA information as a valid source of evaluative information.

significantly mediated by corresponding changes in explicit evaluations, both for participants who received only AA instructions, $Z = 0.70$, $p = .49$, and participants who received AA instructions in addition to AA training, $Z = -0.07$, $p = .95$. The effect of AA instructions on implicit evaluations remained significant after controlling for changes in explicit evaluations (no training: $Z = 2.66$, $p = .008$; training: $Z = 5.71$, $p < .001$).

Discussion

Experiment 1 provided both correlational and experimental evidence that the impact of AA instructions on implicit evaluation is not fully mediated by changes in explicit evaluation. First, correlational analyses show that changes in explicit evaluation only partly mediated the effects of AA instructions on implicit evaluation. That is, AA instructions (and AA training) caused effects on implicit evaluations that remained significant after controlling for the mediating impact of explicit evaluations. This finding corroborates the correlational results of Van Dessel et al. (2015). Second, and most importantly, we found an experimental dissociation on implicit and explicit evaluations with regard to the impact of AA instructions (and AA training). More specifically, when trait instructions were presented, AA instructions and AA training caused changes in implicit but not explicit evaluations. Participants who received information about the evaluative traits of the social groups did not take the AA instructions or training into account when expressing their explicit evaluation, yet still exhibited an implicit preference for the approached group. This resembles previous findings of changes in implicit, but not explicit evaluations as a result of the repeated pairing of stimuli (e.g., Gawronski & LeBel, 2008) and indicates that both AA instructions and AA training can cause changes in implicit evaluation even when participants do not consider this information as a valid source of evaluative information. Given the well-known limitations of correlational mediation analyses, our experimental results provide important new evidence for the conclusion that AA instructions can influence implicit evaluations directly, that is, without first changing explicit evaluations. These findings contradict the idea that instructions influence implicit evaluations only if these instructions are considered a valid basis for evaluation and, hence, are incorporated in explicit evaluations (Gawronski & Bodenhausen, 2006; Whitfield & Jordan, 2009).

In addition to showing that instructions can influence implicit evaluations even when they are not considered a valid basis for evaluation, the present findings also provide information about another important research question that has informed research on the nature of implicit evaluation. Specifically, they inform us on whether the formation and change of implicit evaluations can occur rapidly. In line with Van Dessel et al. (2015) and other studies (e.g., De Houwer, 2006; Peters & Gawronski, 2011) our findings challenge the widespread assumption (e.g., Rydell & McConnell, 2006) that implicit evaluations are slow to *build*. Additionally, and more importantly, these findings indicate that existing implicit evaluations can also be *altered* rapidly, as the result of AA instructions. When participants' evaluations were biased in favor of one of the two social groups as the result of trait instructions, subsequent AA instructions still caused changes in the implicit evaluation of these groups. This contrasts with previous findings suggesting that, once established, implicit evaluations cannot be easily changed (Gregg et al., 2006; Rydell, McConnell, Strain, Claypool, & Hugenberg, 2007). Recently, however, research has shown that new valenced information about a stimulus can lead to a rapid revision of implicit evaluations, but only when this information is considered highly diagnostic about the evaluative properties of this stimulus (Mann, Cone & Ferguson, 2015; Cone & Ferguson, 2015). Our findings go beyond this previous research by showing that rapid alterations in implicit evaluations can occur in the absence of changes in explicit evaluations. AA instructions may thus rapidly alter existing implicit evaluations even if these instructions are not considered diagnostic of the valence of the stimulus.

Experiment 1, however, did not include a control condition to estimate effects of trait instructions and AA instructions separately. Hence, although the results of Experiment 1 confirm our main hypothesis that instructions can cause changes in implicit evaluations in the absence of changes in explicit evaluations, they do not demonstrate conclusively that instructions can also counteract existing implicit evaluations directly. For instance, because of a lack of a control condition, it is theoretically possible that our results were due to the fact that compatible AA instructions *strengthened* the previously acquired implicit evaluations rather than that incompatible AA instructions *revised* them. To examine this question and to ascertain that the finding of a direct influence of AA instructions on implicit evaluations in Experiment 1 was not a chance finding, we performed Experiment 2.

EXPERIMENT 2

In Experiment 2 we further explored AA instruction effects in the context of trait instructions. The aim of this experiment was two-fold. First, we aimed to replicate the finding that AA instructions cause a direct influence on implicit evaluation in the absence of changes in explicit evaluation. In contrast to Experiment 1, we counterbalanced the order of the IAT and the explicit rating task to exclude the possibility that performing the implicit evaluation task first, changed the effects on explicit evaluations (see Perugini, Richetin & Zogmaister, 2014). Second, we extended the previous findings by addressing whether AA instructions cause changes in implicit evaluation when AA instructions are compatible or incompatible with the trait instructions. To this end, participants were provided with either compatible AA and trait instructions, incompatible AA and trait instructions or only trait instructions. Including a condition with only trait instructions allowed us to estimate the effect of trait instructions on evaluations (i.e., the preference for the group that is presented as positive) and examine whether compatible or incompatible AA instructions moderate this effect.

Method

Participants

Participants were 823 English-speaking volunteers who participated online via the Project Implicit research website. Data-exclusion involved removing 195 participants who did not complete all tasks (23.7%), and 156 participants who did not correctly answer the memory questions (24.8%), leaving data from 472 participants (307 women, mean age = 38, *SD* = 13). None of the participants had previously participated in Experiment 1.

Procedure

Experiment 2 was identical to Experiment 1 except for the following points. First, participants were randomly assigned to start with the IAT and then perform the explicit rating task or to perform tasks in the opposite order. Second, participants never received actual AA training. Third, all of the participants received trait instructions. Fourth, not all of the participants received AA instructions. Participants were randomly assigned to receive either (1) no AA instructions, (2) instructions to approach Niffites and avoid Luupites, or (3) instructions to approach Luupites and avoid Niffites. Hence, this experiment employed a 2 (Content of Trait Instructions: Niffites are good, Luupites are good) x 3 (AA Instructions: approach Niffites, approach Luupites, no AA instructions) between-subjects design (Table 3).

Table 3.

Experimental Design of Experiment 2.

2 between participants variables:	
Trait Instructions	AA instructions
	Compatible: Approach Niffites (16.7%)
Niffites are good (50 %)	Incompatible: Approach Luupites (16.7%)
	No AA instructions (16.7%)
	Incompatible: Approach Niffites (16.7%)
Luupites are good (50 %)	Compatible: Approach Luupites (16.7%)
	No AA instructions (16.7%)

Split-half reliability of the IAT score was $r(472) = .92$. Internal consistency of the explicit evaluation score was high (Cronbach's Alpha = .96), and this score correlated significantly with the IAT score, $r(470) = .59$, $p < .001$.

Results

A 3 (AA Instructions: Approach Niffites, Approach Luupites, no AA instructions) x 2 (Content of Trait Instructions: Niffites are good, Luupites are good) ANOVA on the IAT scores revealed a main effect of Content of Trait Instructions, $F(1,466) = 377.50$, $p < .001$, indicating that participants preferred Niffites more when Niffites were presented as positive and Luupites as negative ($M = 0.26$, $SD = 0.48$) than when Niffites were presented as negative and Luupites as positive ($M = -0.54$, $SD = 0.39$), $d = 1.82$, 95% CI [1.60, 2.04], $BF_1 > 10000$. Most importantly, we also observed a main effect of AA Instructions, $F(2,466) = 4.59$, $p = .011$ (Table 4). In line with Experiment 1, participants who had been instructed to approach Niffites and avoid Luupites ($M = -0.02$, $SD = 0.58$) preferred Niffites more than participants who had been instructed to approach Luupites and avoid Niffites ($M = -0.27$, $SD = 0.55$), $F(1,309) = 9.24$, $p = .003$, $d = 0.44$, 95% CI [0.21, 0.66], $BF_1 = 131.22$. Compared to participants who had not received AA instructions ($M = -0.08$, $SD = 0.62$), participants who had received instructions to approach Luupites preferred Luupites more, $F(1,313) = 4.98$, $p = .026$, but we observed no significant difference for participants who had received approach Niffites instructions, $F(1,310) = 0.41$, $p = .52$.

Table 4.

Mean IAT and Explicit Scores in Experiment 2 as a function of Content of Trait Instructions and AA Instructions.

	Content of Trait Instructions					
	Niffites good and Luupites bad			Niffites bad and Luupites good		
	Approach Niffites	Approach Luupites	No AA instructions	Approach Niffites	Approach Luupites	No AA instructions
IAT score:	0.31 (0.47)	0.15 (0.46)	0.31 (0.50)	-0.46 (0.41)	-0.60 (0.36)	-0.54 (0.40)
Explicit score:	3.07 (3.13)	2.38 (4.02)	2.75 (3.04)	-3.40 (3.80)	-3.34 (3.41)	-3.29 (3.10)

Note. Standard deviations are in parentheses. Scores reflect a relative preference for Niffites over Luupites.

To examine whether compatible or incompatible AA instructions cause changes in evaluations we performed planned tests comparing the main effect of Content of Trait Instructions for participants who received no AA instructions, participants who received compatible AA instructions and participants who received incompatible AA instructions. Importantly, the main effect of Content of Trait Instructions was reduced when AA instructions were incompatible with the trait instructions, *d* = 1.39, 95% CI [1.01, 1.78], compared to when no AA instructions were provided, *d* = 1.85, 95% CI [1.47, 2.23], $F(1,291) = 5.24$, $p = .023$, indicating that incompatible AA instructions influenced implicit evaluations. In contrast, the main effect of Content of Trait Instructions was not significantly different for participants who received compatible AA instructions, *d* = 2.15, 95% CI [1.78, 2.53] compared to participants who received no AA instructions, $F(1,332) = 0.34$, $p = .56$.

The 3 x 2 ANOVA on explicit ratings revealed only the main effect of Content of Trait Instructions, $F(1,466) = 370.73$, $p < .001$, *d* = 1.80, 95% CI [1.58, 2.01], indicating a larger preference for Niffites when they were presented as positive (*M* = 2.76, *SD* = 3.37) than when they were presented as negative (*M* = -3.34, *SD* = 3.42). We did not observe a significant main effect of AA Instructions, $F(1,466) = 0.36$, $p = .70$, nor an interaction effect with Content of Trait Instructions, $F(1,466) = 0.47$, $p = .63$. Also, the main effect of Content of Trait Instructions did not differ significantly between participants who received compatible, incompatible or no AA Instructions, *F*s < 0.37, *ps* > .54, BF_0s > 7.00.

AA instructions condition

In line with Experiment 1, mediation analyses on the data of participants who received both AA and trait instructions showed that AA instruction effects on implicit evaluations were not significantly mediated by corresponding changes in explicit evaluations, $Z = 1.87$, $p = .062$. The effect of AA instructions on implicit evaluations remained significant after controlling for explicit evaluations, $Z = 2.92$, $p = .003$.

Discussion

Results from Experiment 2 provide further support for the idea that the impact of AA instructions on implicit evaluations is not fully mediated by changes in explicit evaluations. Replicating the pattern obtained in Experiment 1, participants who received AA instructions exhibited an implicit, but not an explicit preference for the approached group over the avoided group when prior instructions specified the valence of these groups. Mediation analyses indicated that AA instruction effects on implicit evaluation were not fully mediated by changes in explicit evaluation in the context of trait instructions.

Additionally, results indicated that AA instructions caused changes in implicit evaluation even when the valence implied by the approach or avoidance action was *incompatible* with the evaluative information provided in the trait instructions. This suggests that AA instructions can (partly) undo recently established implicit evaluations, even in the absence of changes in explicit evaluations. This contrasts evidence that implicit evaluations are more difficult to change than explicit evaluations with verbally presented counter-attitudinal information (Gregg et al., 2006). We found no evidence that AA instructions caused changes in implicit evaluations when these instructions were *compatible* with the trait instructions. This is consistent with previous findings that AA training causes changes in implicit evaluations of social groups only when the training is incompatible with participants' evaluations (Kawakami et al., 2007). It suggests that AA effects may be strongly reduced when participants have clearly univalent positive or negative implicit evaluations and corroborates previous evidence that the effectiveness of instructions to approach or avoid a stimulus may critically depend on specific stimulus properties (e.g., whether a stimulus is novel or well-known; see Van Dessel et al., 2015).

GENERAL DISCUSSION

In two experiments, we observed that instructions to approach or avoid members of a fictitious group impact implicit evaluations of these groups. Our results indicate that these changes in implicit evaluation are not fully mediated by changes in explicit evaluations. Experiment 1 provided evidence that participants who merely received AA instructions and participants who received additional AA training exhibited a direct effect on implicit evaluations. Moreover, both procedures caused changes in implicit evaluations even when trait instructions clearly specified the valence of the groups which canceled any AA effect on explicit evaluative ratings. Experiment 2 corroborated that AA instructions influenced implicit, but not explicit evaluations in the context of trait instructions and extended these findings by showing that AA instructions caused changes in implicit evaluations when AA instructions were incompatible with the trait instructions.

These findings have meaningful theoretical and practical implications. We first discuss implications for theories on the mental processes that underlie implicit evaluation. Afterwards, we discuss implications for mental process theories that account for AA instruction and AA training effects. Finally, we discuss practical implications of the present research.

Implications for theories of implicit evaluation

The current experiments provide important information that constrains current and future models of implicit evaluation. First, the observation that AA instructions have a direct influence on implicit evaluation (i.e., independent of changes in explicit evaluation) is difficult to reconcile with associative and dual-process models of evaluation that only allow for evaluative associations to form (1) gradually as the result of many pairings (e.g., Smith & DeCoster, 2000; Rydell & McConnell, 2006) or (2) rapidly when consciously entertaining the proposition that a stimulus is positive or negative (Gawronski & Bodenhausen, 2006). However, dual-process models can accommodate these findings if they allow for the immediate formation of associations even on the basis of information that is not considered to be valid. Also propositional single-process accounts of evaluation can account for our results if they assume that the automatic activation of propositional information underlies implicit evaluation (De Houwer, 2014). More specifically, receiving AA instructions may allow participants to consider the proposition that the approached social group is positive. A dissociation between implicit and explicit evaluation will arise when this proposition is judged to be invalid (and thus dismissed when making an explicit evaluation) but still automatically retrieved when the social group is implicitly evaluated.

Second, the observation that incompatible AA instructions reduce effects of trait instructions on implicit, but not on explicit evaluations suggests that implicit evaluations can be updated rapidly. It provides direct evidence against the often entertained idea that implicit evaluations are more difficult to change than explicit evaluations via counter-attitudinal information (Gregg et al., 2006; Rydell & McConnell, 2006). Rather, changes in explicit evaluation seem to critically depend on the perceived validity of the obtained evaluative information (Peters & Gawronski, 2011). When information directly contradicts previous valence information, this causes an immediate reversal of participants' explicit liking of the stimulus (Gregg et al., 2006). Because AA instructions do not invalidate the more diagnostic evaluative trait information, they do not influence explicit evaluation when they contradict trait instructions. In contrast, changes in implicit evaluations may arise as the result of any information that links a stimulus with a specific valence, such as information about its relation with another valenced stimulus (see Zanon, De Houwer, Gast, & Smith, 2014) or with a valenced action (Van Dessel et al., 2015). Immediate changes in implicit evaluation may occur, even when participants do not consider the obtained information as valid.

Note that the present findings do not contradict the idea that the impact of counter-attitudinal information strongly depends on the diagnosticity of this information (Cone & Ferguson, 2015). In fact, our data also suggest that AA instructions have a stronger influence on implicit evaluation if they are more diagnostic. This can be inferred from the fact that we observed a bigger AA instruction effect in the absence of trait instructions, that is, when the AA instructions were the most diagnostic piece of information that was available to the participants. However, our results extend the previous research by showing that changes in implicit evaluations may occur as the result of instructions even when these instructions provide information that is not considered highly diagnostic of the evaluative properties of the stimulus and therefore do not influence explicit evaluations. This effect is automatic in the sense that, in all likelihood, our participants did not intend to use this information for their evaluation.

In sum, the current findings provide important information for theories that explain how implicit evaluations arise and can be changed. Although our results cannot distinguish between the broad class of single-process propositional and the broad class of dual-process models, they do force these models into adopting specific assumptions without which they cannot account for our effects. In general, we believe that it is difficult, if not impossible, to distinguish between broad classes of models like dual-process or single-process models that have such a high degree of flexibility. Therefore, we believe that, in order to further advance research on evaluation, it is necessary to (1) define specific models (e.g., propositional or association formation models of AA effects) that make testable predictions and (2) perform research to test these predictions. The data produced by such research will allow us to further constrain these models and to have greater confidence in the assumptions that survive this process.

Implications for accounts of AA instruction and AA training effects

First, the current findings indicate that instructions that link a valenced action and a fictitious social group cause unintentional changes in the implicit evaluation of these groups. This extends knowledge about the effects of AA instructions by showing that these effects are not necessarily the result of controlled, non-automatic processes that involve the intentional use of this information for evaluation (e.g., as the result of demand compliance) (Van Dessel et al., 2015).

Second, our results also constrain ideas about the processes that underlie AA training effects. More specifically, they reveal important similarities between the effects of AA training and those of AA instructions. Not only can both interventions lead to changes in implicit evaluations, they both can have direct effects on implicit evaluations, that is, effects that are not mediated by changes in explicit evaluations. Although these similarities do not prove that both types of effects are due to the same mental processes (e.g., the formation and activation of propositions), they are in line with this idea and hence undermine the position that AA training effects can be due only to low level processes such as the gradual, performance-driven formation of associations in memory (e.g., Woud et al., 2013; Phills, Kawakami, Tabi, Nadolny, & Inzlicht, 2011). Future studies are required to establish whether instructions and pairings are also similar regarding other features, for example regarding uncontrollability (see Gawronski, Balas, & Creighton, 2014).

Finally, our findings suggest that actually performing AA behavior may, under certain conditions, add to the effect of AA instructions on implicit evaluations. Experiment 1 included a direct comparison of AA instruction and AA training effects on implicit and explicit evaluations. For participants who did not receive trait instructions, additional AA training did not have an added effect even though we had sufficient statistical power to detect even a small effect (power = .77 to detect an effect size of d = 0.25). In contrast, participants who received trait instructions exhibited a stronger AA effect on implicit evaluation when AA instructions were supplemented with AA training. Whether this added effect of AA training involves the strengthening of the previously obtained knowledge structures (i.e., associations or propositions) or the acquisition of entirely different knowledge structures requires further research.

Practical Implications

AA training is considered an important procedure for the modification of pathological biases in cognitive functioning (see Woud & Becker, 2014). Repeatedly performing AA movements in response to specific stimuli has proven effective in a number of therapeutic contexts such as the treatment of alcohol addiction (Wiers, Eberl, Rinck, Becker, & Lindenmeyer, 2011), social anxiety (Taylor & Amir, 2012), or contamination-related fear (Amir, Kuckertz, & Najmi, 2013). Given the important relation between implicit evaluation and the dysfunctional behavioral responses under investigation (see Houben, Havermans, & Wiers, 2010), it can be argued that changes in implicit evaluation may (partly) underlie therapeutic effects of AA training. Following this reasoning, our current results may indicate that AA instructions could play an important role in these AA training effects. Preliminary evidence supporting this idea was found in a recent study by Wiers et al. (2014) where therapeutic effects of 'avoid alcohol' training at one month follow-up were more robust if participants had received explicit instructions to push alcohol away in addition to the re-training procedure. Future research might consider whether replacing or complementing AA training with AA instructions may improve the therapeutic effectiveness of AA training.

Concluding remarks

In sum, the present results extend past findings that verbal instructions influence implicit evaluation by showing that AA instruction effects on implicit evaluations occur in the absence of mediation by changes in explicit evaluation. These findings provide insight into the mechanisms underlying implicit evaluation and open up important new avenues for changing implicit evaluations.

REFERENCES

Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*, 370-374. doi: 10.1002/wics.84.

Allport, G. (1935). Attitudes. In Murchison, C. (Ed.). *A Handbook of Social Psychology* (pp. 789-843). Worcester, MA: Clark University Press.

Amir, N., Kuckertz, J. M., & Najmi, S. (2013). The Effect of Modifying Automatic Action Tendencies on Overt Avoidance Behaviors. *Emotion, 13,* 478-484. doi: 10.1037/a0030443

Castelli, L., Zogmaister , C., Smith, E.R., Arcuri, L. (2004). On the automatic evaluation of social exemplars. *Journal of Personality and Social Psychology, 86*, 373-387. doi: 10.1037/0022-3514.86.3.373

Cone, J., & Ferguson, M. J. (2015). He Did what?: The Role of Diagnosticity in Revising Implicit evaluations. *Journal of Personality and Social Psychology, 108*, 37-57. doi: 10.1037/pspa0000014

De Houwer, J. (2006). Using the Implicit Association Test does not rule out an impact of conscious propositional knowledge on evaluative conditioning. *Learning and Motivation, 37*, 176-187. doi: 10.1016/j.lmot.2005.12.002

De Houwer, J. (2009a). How do people evaluate objects? A brief review. *Social and Personality Psychology Compass, 3*, 36-48. doi: 10.1111/j.1751-9004.2008.00162.x

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

De Houwer, J. (2014). A Propositional Model of Implicit Evaluation. *Social and Personality Psychology Compass, 8*, 342-353. doi: 10.1111/spc3.12111

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013).A functional- cognitive framework for attitude research. *European Review of Social Psychology, 24*, 252–287. doi: 10.1080/10463283.2014.892320

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229–238. doi: 10.1037//0022-3514.50.2.229

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

Gawronski, B., Balas, R., & Creighton, L. A. (2014). Can the formation of conditioned attitudes be intentionally controlled? *Personality and Social Psychology Bulletin, 40*, 419-432. doi: 10.1177/0146167213513907

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692-731. doi: 10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology, 44,* 59-127. doi:10.1016/B978-0-12-385522-0.00002-0

Gawronski, B., & Bodenhausen, G. V. (2014). Implicit and explicit evaluation: A brief review of the associative-propositional evaluation model. *Social and Personality Psychology Compass*, *8*, 448-462. doi: 10.1111/spc3.v8.8

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*, 1355-1361. doi: 10.1016/j.jesp.2008.04.005

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*, 1464–1480. doi: 10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of*

Personality and Social Psychology, 85, 197–216. doi: 10.1037/0022-3514.85.2.197

Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. Journal of *Personality and Social Psychology, 97*, 17–41. doi: 10.1037/ a0015575

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,* 1-20. doi: 10.1037/0022-3514.90.1.1

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

Houben, K., Havermans, R., & Wiers, R. W. (2010). Learning to dislike alcohol: Conditioning negative implicit attitudes towards alcohol and its effect on drinking behavior. *Psychopharmacology, 211*, 79-86. doi: 10.1007/s00213-010-1872-1

Jeffreys, H. (1961). Theory of Probability. Oxford: Oxford University Press.

Jones, C. R., Vilensky, M. R., Vasey, M. W., & Fazio, R. H. (2013). Approach behavior can mitigate predominately univalent negative attitudes: Evidence regarding insects and spiders. *Emotion, 13*, 989-996. doi: 10.1037/a0033164

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder: Improving implicit racial evaluations and interracial interactions through approach behaviors. *Journal of Personality and Social Psychology, 92*, 957-971. doi: 10.1037/0022-3514.92.6.957

Lynch, J. G. Jr. (2005). Accessible but Nondiagnostic Memories about Memory and Consumer Choice. In Griffin, A. & Ottnes, C.C. (Eds.). *16th Paul D. Converse Symposium* (88-115). Chicago: American Marketing Association.

Mann, T., Cone, J., & Ferguson, M. J. (2015). Social-psychological evidence for the effective updating of implicit attitudes. *Behavioral and Brain Sciences.* 38: e15. doi: 10.1017/S0140525X14000223

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

Peters, K. R., Gawronski, B. (2011). Are We Puppets on a String? Comparing the Impact of Contingency and Validity on Implicit and Explicit Evaluations. *Personality and Social Psychology Bulletin, 37*, 557-569. doi: 10.1177/0146167211400423

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

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. doi: 10.1037/a0022159

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. URL: http://www.jstatsoft.org/v48/i02/

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237. doi: 10.3758/PBR.16.2.225

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*, 995- 1008. doi: 10.1037/0022-3514.91.6.995

Rydell, R., McConnell, A. R., Strain, L. M., Claypool, H. M., & Hugenberg, K. (2007). Implicit and explicit attitudes respond differently to increasing amounts of counterattitudinal information. European Journal of Social Psychology, 37, 867–878. doi: 10.1002/ejsp.393

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology, 89*, 845–851. doi:10.1037/0022-3514.89.6.845

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review,* 4, 108-131. doi: 10.1207/S15327957PSPR0402_01

Smith, C. T., De Houwer, J., & Nosek, B. (2013). Consider the Source: Persuasion of Implicit Evaluations Is Moderated by Source Credibility. *Personality and Social Psychology Bulletin,* 39, 193-205. doi: 10.1177/0146167212472374

Taylor, C. T., & Amir, N. (2012). Modifying automatic approach action tendencies in individuals with elevated social anxiety symptoms. *Behaviour Research and Therapy, 50*, 529-536. doi: 10.1016/j.brat.2012.05.004

Vandenbosch, K., & De Houwer, J. (2011). Failures to induce implicit evaluations by means of approach-avoid training. *Cognition and Emotion, 25,* 1311-1330. doi: 10.1080/02699931.2011.596819

Van Dessel, P., De Houwer, J., Gast, A., & Smith, C. T. (2015). Instruction-Based Approach–Avoidance Effects: Changing Stimulus Evaluation via the Mere Instruction to Approach or Avoid Stimuli. *Experimental Psychology, 62*, 161-169. doi: 10.1027/1618- 3169/a000282

Whitfield , M., & Jordan , C. H. (2009). Mutual influences of explicit and implicit attitudes. *Journal of Experimental Social Psychology, 45*, 748–759. doi: 10.1016/j.jesp.2009.04.006

Wiers, R.W., Eberl, C., Rinck, M., Becker, E. & Lindenmeyer, J. (2011). Re-training automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science, 22,* 490-497. doi:10.1177/0956797611400615

Wiers, R.W., Houben, K., Fadardi, J.S., van Beek, P., Rhemtulla, M., & Cox, W.M. (2014) Alcohol Cognitive Bias Modification training for problem drinkers over the web. *Addictive Behaviors.* 40, 21-26. doi: 10.1016/j.addbeh.2014.08.010

Woud, M. L., & Becker, E.S. (2014). Editorial for the Special Issue on Cognitive Bias Modification Techniques: An Introduction to a Time Traveler's Tale.

Cognitive Therapy and Research, 38, 83-88. doi: 10.1007/s10608-014-9605-0

Woud, M. L., Becker, E. S., Lange, W. G., & Rinck, M. (2013). To like or not to like: The effect of approach-avoidance training on neutral, angry and smiling face stimuli. Psychological Reports, 113, 199-216. doi: 10.2466/21.07.PR0.113x10z1

Woud, M. L., Maas, J., Becker, E.S., & Rinck, M. (2013). Make the manikin move: Sumbolic approach-avoidance responses affect implicit and explicit face evaluations. *Journal of Cognitive Psychology, 25*, 738-744. doi: 10.1080/20445911.2013.817413

Zanon, R., De Houwer, J., Gast, A., & Smith, C. T. (2014). When does relational information influence evaluative conditioning? *Quarterly Journal of Experimental Psychology, 67*, 2105-2122. doi: 10.1080/17470218.2014.907324

APPENDIX

Mediation Analyses Experiment 1

No trait instructions condition

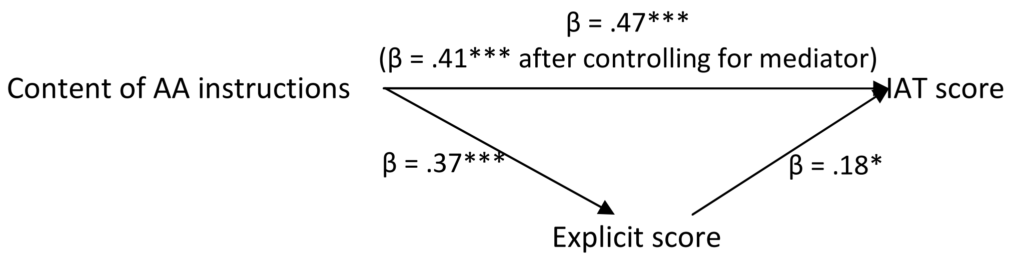

Figure A1. Standardized Estimates of mediation coefficients for participants in Experiment 1 who received no trait instructions and no AA training. * p < .05 ** p < .01 *** p < .001.

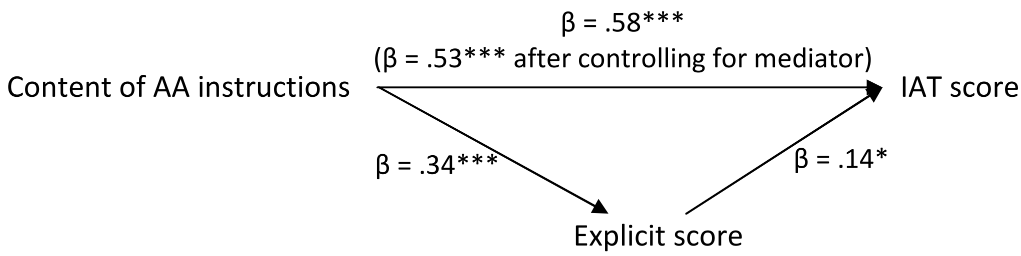

Figure A2. Standardized Estimates of mediation coefficients for participants in Experiment 1 who received no trait instructions and received AA training. * p < .05 ** p < .01 *** p < .001.

Trait instructions condition.

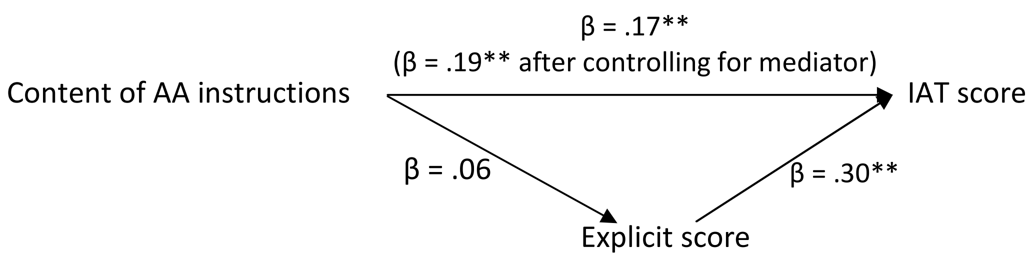

Figure A3. Standardized Estimates of mediation coefficients for participants in Experiment 1 who received trait instructions and no AA training. * p < .05 ** p < .01 *** p < .001.

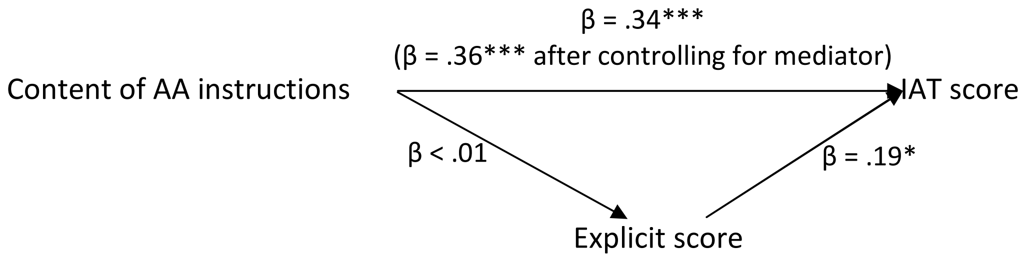

Figure A4. Standardized Estimates of mediation coefficients for participants in Experiment 1 who received trait instructions and AA training.
**** p < .05 ** p < .01 *** p < .001.***

Mediation Analyses Experiment 2.

AA instruction condition.

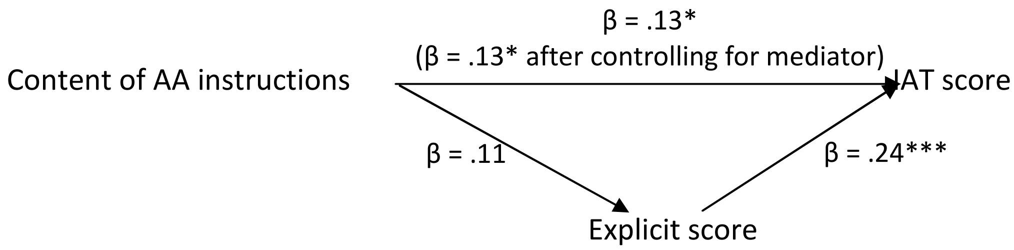

Figure A5. Standardized Estimates of mediation coefficients for participants in Experiment 2 who received AA instructions.
**** p < .05 ** p < .01 *** p < .001.***

CHAPTER 4

APPROACH–AVOIDANCE TRAINING EFFECTS ARE MODERATED BY AWARENESS OF STIMULUS-ACTION CONTINGENCIES[1]

Prior research suggests that repeatedly approaching or avoiding a stimulus changes the liking of that stimulus. In two experiments, we investigated the relationship between, on the one hand, effects of approach-avoidance (AA) training on implicit and explicit evaluations of novel faces and, on the other hand, contingency awareness as indexed by participants' memory for the relation between stimulus and action. We observed stronger effects for faces that were classified as contingency aware and found no evidence that AA training caused changes in stimulus evaluations in the absence of contingency awareness. These findings challenge the standard view that AA training effects are (exclusively) the product of implicit learning processes, such as the automatic formation of associations in memory

---

[1] Based on Van Dessel, P., De Houwer, J., & Gast, A. (2016). Approach-Avoidance Training effects are Moderated by Awareness of Stimulus-Action Contingencies. *Personality and Social Psychology Bulletin, 42*, 81-93. doi: 10.1177/0146167215615335

INTRODUCTION

Actions of approach and avoidance (AA) are assumed to be closely linked to the evaluation of a stimulus as good or bad. First, evaluative stimuli are thought to automatically evoke approach (in the case of positive stimuli) or avoidance responses (in the case of negative stimuli; e.g., Chen & Bargh, 1999; but see Rotteveel et al., 2015). Second, AA movements have also been used to change the evaluation of stimuli. For instance, Kawakami, Phills, Steele, and Dovidio (2007) demonstrated that participants who repeatedly approached photographs of Black people exhibited more positive evaluations of Black people on the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) than participants who repeatedly avoided photographs of Black people. Recent research indicates that AA training can cause changes in the evaluations of a variety of well-known stimuli, such as familiar alcoholic drinks (Wiers, Eberl, Rinck, Becker, & Lindenmeyer, 2011), insects and spiders (Jones, Vilensky, Vasey, & Fazio, 2013), or contamination-related objects (Amir, Kuckertz, & Najmi, 2013).

Researchers have considered whether AA training procedures can also be used to establish evaluations of novel stimuli. Woud, Becker, and Rinck (2008) reported that participants who repeatedly performed AA movements in response to pictures of faces with neutral emotional expressions exhibited a preference for approached faces on an implicit measure of evaluation (the evaluative priming task; Fazio, Sanbonmatsu, Powell, & Kardes, 1986). A growing number of studies have provided evidence that AA training causes changes in implicit (e.g., Woud, Maas, Becker, & Rinck, 2013) and explicit evaluations (e.g., Huijding, Muris, Lester, Field, & Joosse, 2011; Laham, Kashima, Dix, Wheeler, & Levis, 2014) of novel stimuli. Vandenbosch and De Houwer (2011), however, failed to find any evidence for AA training effects on evaluations of novel faces in five experiments and failed to reproduce the effect reported by Woud et al. (2008) when reanalyzing their data, suggesting that AA training effects may be subject to subtle boundary conditions or moderators that yet have to be identified.

The aim of the present study was to investigate the role of one possible moderator, namely contingency awareness. The role of contingency awareness has been an important topic in research on the acquisition of preferences via conditioning procedures. In evaluative conditioning (EC) studies, neutral conditioned stimuli (CSs) are repeatedly paired with positive or negative unconditioned stimuli (USs), resulting in changes in liking of the CSs. Some have argued that EC differs from other variants of conditioning in that EC can occur in the absence of conscious awareness of the contingency between CS and US (e.g., Baeyens, Eelen, & van den Bergh, 1990; Olson & Fazio, 2001). However, a number of studies have challenged this view and provided evidence that EC effects can be observed only when participants are able to report the contingency between CS and US (e.g., Pleyers, Corneille, Luminet, & Yzerbyt, 2007; Gast, De Houwer, & De Schryver, 2012). Though there is still debate about the necessity of the awareness of the CS-US contingencies in EC effects (see for instance, Hütter, Sweldens, Stahl, Unkelbach, & Klauer, 2012), there is general consensus that contingency awareness is an important moderator of EC effects (Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010; Sweldens, Corneille, & Yzerbyt, 2014).

AA training effects resemble EC effects in that a change in liking is observed that results from a contingency between a neutral stimulus and a valenced event. Whereas in EC studies, the valenced event is typically conceived of as the presentation of a stimulus (De Houwer, 2007; but see Gast & Rothermund, 2011), in AA studies the valenced event corresponds to the execution of a valenced action (De Houwer, 2007). Hence, the role of contingency awareness can be studied also in AA training research. Examining this issue is bound to have important theoretical implications. Most importantly, traditional associative theories of AA training imply that contingency awareness might not be critical. These accounts postulate that AA behavior activates specific AA motivational orientations (Cacioppo, Priester, & Berntson, 1993; Markman & Brendl, 2005;

Neumann & Strack, 2000). AA training effects are thought to arise as a result of the gradual formation of an association between the activated motivational state and the stimulus representation (Woud et al., 2013). In line with the idea that approach and avoidance are primitive behavioral tendencies that are tightly linked with an impulsive, associative system (Strack & Deutsch, 2004), these accounts attribute AA training effects to automatic associative processes. From this perspective, the processes underlying AA training effects might differ from those involved in EC in that they provide a more low-level route to changing stimulus evaluations. Attesting to the dominance of this low-level associative view on AA training, factors that indicate the involvement of high-level controlled processes, such as participants' awareness of the contingency between the AA action and the stimulus, have received little or no attention in AA training research.

There are, however, reasons to believe that contingency awareness is a key factor in establishing AA training effects. First, AA training procedures that allow participants to become aware of the stimulus-action contingencies typically produce more robust AA training effects. Effects of AA training are consistently reported when participants receive instructions about the crucial stimulus-action contingencies (e.g., Kawakami et al., 2007) or when the target stimuli or stimulus features (i.e., the stimuli or stimulus features whose valence is registered and targeted for change) are specified in the instructions (e.g., Wiers et al., 2011). In contrast, when targets are not specified in the instructions, effects sizes tend to be small at best. In the studies by Vandenbosch and De Houwer (2011), for instance, AA training consisted of the repeated approach or avoidance of individual stimuli (i.e., pictures of 12 novel faces). Importantly, these faces had a subtle brown or red filter placed over them and participants were instructed to approach or avoid on the basis of the color of the presented face. Unbeknownst to the participants, some stimuli were always presented in the to-be-approached color whereas other stimuli were always presented in the to-be-avoided color. The fact that the target feature of the stimuli (i.e., stimulus identity) was not specified in the instructions probably reduced the chance that participants realized that there were specific stimulus-action contingencies and thus that they identified those contingencies (e.g., approach Face 1, avoid Face 2, …). The lack of contingency awareness could have been responsible for the lack of AA training effects. Finally, in a recent study by Laham et al. (2014), participants repeatedly performed AA actions in response to unfamiliar shapes, which resulted in a preference for the approached stimuli. Importantly, this effect was observed only if participants performed the AA movements in a motivating context (i.e., collecting or discarding fruits in a foraging context). As a possible explanation for their results, the authors suggested that elaborated framing instructions increased the likelihood that participants became aware of the stimulus-action contingencies.

A second line of research that points at the importance of contingency awareness in AA training focused on the effects of instructions. In a recent set of studies that were conducted at our laboratory, we observed typical AA training effects even when participants did not perform AA actions but were merely instructed that they would later on have to perform these actions (Van Dessel, De Houwer, Gast, & Smith, 2015). Participants who received instructions to approach one fictitious social group (e.g., the group Niffites) and avoid another fictitious social group (e.g., the group Luupites) exhibited a preference for the former group both on implicit and explicit measures of evaluation. These findings suggest that the acquisition of conscious propositional knowledge about stimulus-action regularities can cause changes in liking. Although these findings do not allow for the conclusion that all AA training effects are based on conscious propositional knowledge, they at least support the idea that the acquisition of contingency information can be an important factor also during regular AA training.

In the present studies, we used a variant of the procedure introduced by Woud et al. (2008) that allowed us to investigate effects of AA training on implicit and explicit evaluations of novel faces. More specifically, we used both a correlational and an experimental approach to address the importance of contingency awareness in AA training effects. First, we tried to capture participants' awareness of the experienced face-action contingencies by measuring participants' memory about the relation between faces and actions. We compared AA training effects for faces that were correctly linked to the action they were paired with, faces that were linked to the incorrect action, and faces for which participants did not remember the correct action. Second, in our first study, we manipulated contingency awareness by providing one group of participants with instructions that specified the face-action contingencies whereas a second group did not receive these instructions. We examined whether this manipulation caused changes in contingency awareness and whether these changes affected AA training effects.

EXPERIMENT 1

Method

Participants

Sixty-three native Dutch-speaking undergraduates (51 women) participated in exchange for a monetary reward of 5 euros. All participants had normal or corrected-to-normal vision and were naive with respect to the purpose of the experiment. We excluded the data from one participant whose error rate in the evaluative priming task was more than 2.5 standard deviations above the population mean (population mean = 5.41 %, *SD* = 4.01%).

Apparatus and Materials

Eight photographs of faces (four men and four women) served as stimuli[1]. Pictures were selected from the Radboud Faces Database on the basis of a validation study conducted by Langner et al. (2010) in which participants indicated the emotional expression of the face by choosing between nine possible expressions and provided ratings for valence of the face on a five-point

[1] To increase the possibility that participants would identify some of the stimulus-action contingences we used a smaller number of evaluative stimuli than Woud et al. (2008).

Likert scale ranging from "1" (negative) to "5" (positive). Two selection criteria were used. First, the emotional expression of the face was correctly identified as neutral (i.e., more than 85 % correct identifications) in the validation study. Second, the mean rating for the valence of the face was near the nominal midpoint of the rating scale (range: 2.85 – 3.25).

In the evaluative priming task, four positive words (the Dutch words for HAPPY, HONEST, NICE, and SINCERE), and four negative words (the Dutch words for MEAN, BRUTAL, AGGRESSIVE, and FAKE) were presented as target stimuli. The eight faces were used as primes. All words were presented in uppercase letters in Arial Black font with font size 36.

The experiment was programmed in C-language and presented using the C-library Tscope package (Tscope 1.0.171.) on a Tori PC with a 19-inch monitor (80 Hz refresh rate), a keyboard and a joystick (Wingman Attack 2) attached to it.

Procedure

After participants had given informed consent, they were seated in front of a computer screen on which instructions for the AA training task appeared. Participants were informed that they would see pictures of different faces and that they would have to make a certain action each time a picture was presented, depending on the color of the frame that was presented around the picture. Half of the participants were told that they would have to approach photos that had a blue frame by pulling the joystick towards them and avoid photos with a green frame by pushing the joystick away from them. The other half received the opposite instructions. Orthogonal to this manipulation, participants were randomly assigned to receive instructions about stimulus-action contingencies or no contingency instructions. Participants in the 'contingency instructions group' were shown the eight faces they would see during the task and were told which four faces they would approach and which four faces they would avoid. They were asked to make sure that they would not forget which action belonged with each face. Participants assigned to the 'no contingency instructions group' were merely shown the eight faces they would see during the AA training task without any information about face–action contingencies.

The AA training task consisted of two blocks of 96 trials. During each training block, each of the eight faces was presented on 12 trials and was always presented with either a blue or a green frame, indicating that it had to be approached or avoided. For each participant four faces were always approached and four faces were always avoided. We randomized the assignment of faces to the approach or avoidance action. Each trial started with the presentation of a white fixation cross presented in the centre of the screen. After 500 ms the fixation cross was replaced by the picture of a face surrounded by a colored frame. This picture was randomly presented in four different sizes (i.e., 6.08 cm high x 4.56 cm wide; 6.40 cm high x 4.80 cm wide; 6.72 cm high x 5.04 cm wide; 7.04 cm high x 5.28 cm wide) to prevent participants from performing the task by focussing on a specific point on the screen in order to process the color of the frame, thereby limiting picture content processing (see Huijding & De Jong, 2005). The face disappeared as soon as participants responded correctly with the joystick by performing a vertical movement towards the screen or away from the screen.[2] After 200 ms the next trial started. Note that Woud et al. (2008) included a phase in which participants performed both approach and avoidance movements in response to each stimulus. This provided an additional index of learning because it allowed Woud et al. to check whether performance was better on trials that respected the initial stimulus-action contingencies. We decided not to include such a phase in order to allow participants to experience a perfect contingency between face and action and to increase the possibility that participants would identify these contingencies.

In the evaluative priming task, participants categorized target words as either "positive" or "negative" using the ‘E’ and ‘I keys of a computer keyboard. The assignment of the response keys to either the positive or negative category

[2] In contrast to the procedure by Woud et al. (2008), performing the AA action resulted in the immediate disappearance of the stimulus. We decided not to include a zoom effect (i.e., an effect where pictures become smaller while pushing and larger while pulling the joystick) on the basis of an initial study with 20 participants where we observed a significant AA training effect only when the training did not involve a zoom effect.

was counterbalanced across participants. Participants were instructed to perform this categorization task as quickly as possible, while making as few mistakes as possible. Participants were further told that they would see pictures of faces presented before the words and that they could look at these pictures, but that their task was simply to respond on the basis of the valence of the positive or negative word. A single trial consisted of a fixation cross presented for 500 ms, a blank screen for 500 ms, a face for 200 ms (i.e., prime), a post-prime pause of 50 ms, and the target word in white font for 1500 ms or until participants had given a response. Error feedback was presented on the screen (i.e., the Dutch word for ‘Wrong’ presented in red font) for 250 ms if participants made an error. The inter-trial interval was set to vary randomly between 500 ms and 1500 ms. Participants completed 128 trials separated into four blocks of 32 trials, each containing two trials with each of the faces as prime and a positive or negative word as target presented in random order.

After the priming task, we registered explicit evaluations of the faces. Participants indicated whether they liked the person in the photo and whether they thought the person in the photo was friendly on two nine-point Likert scales (0 = not liked at all/not friendly at all; 8 = liked a lot/very friendly). For each face, we collapsed these score ratings into one explicit rating score by averaging the respective ratings. The internal consistency of this measure was moderate (mean Cronbach’s Alpha = .62, *SD* = 0.12).

Participants then completed questions assessing awareness of the stimulus-action contingencies. Each of the faces was presented in a random order. Participants were asked to indicate what action they had performed most often in response to this picture by choosing from three options (i.e., approach’, ‘avoid’, or ‘both actions the same number of times’). Participants were asked to report their confidence in each of their answers on a 3-point Likert scale ranging from 1 (i.e., unsure) to 3 (i.e., very sure).

Results

Contingency awareness

On average, participants in the contingency instructions group selected the correct action for 77% of the faces (*SD* = 27%). Participants who had not received contingency instructions indicated the correct action for fewer faces (*M* = 54%, *SD* = 27%), $t(60) = 3.34$, $p = .001$. In contrast, participants in the no contingency instructions group indicated more often that they had performed both actions an equal number of times in response to the face stimulus (contingency instructions: *M* = 8%, *SD* = 12%; no contingency instructions: *M* = 26%, *SD* = 27%), $t(60) = -3.44$, $p = .001$. We observed no significant differences in the number of times participants chose the incorrect action (contingency instructions: *M* = 15%, *SD* = 23%; no contingency instructions: *M* = 20%, *SD* = 15%), $t(60) = -0.91$, $p = .37$. Importantly, participants in both the contingency instructions and the no contingency instructions group correctly identified the action more often than they chose the incorrect action, *ts* > 5.68, *ps* < .001, indicating that they were able to identify some of the stimulus-action contingencies.

Linear mixed effects models analysis

The analyses of the explicit rating scores and evaluative priming task data were performed with item-based linear mixed effects models (multilevel model analysis) as implemented in R package lme-4 (Bates, Maechler, Bolker, & Walker, 2014). Linear mixed effects models allow us to base the analyses on items (rather than participants' means) and simultaneously control for random effects of participants and items while assessing relevant (fixed) factors of interest (Baayen, Davidson, & Bates, 2008; Hoffman & Rovine, 2007; Locker, Hoffman, & Bovaird, 2007). Linear mixed effects regression (lmer) analyses are preferred over standard analyses of variance (ANOVA) in studies that use item-based analyses of awareness because they are better able to deal with unbalanced data (see also Gast et al., 2012). In our study, they prevent the substantial data loss that would result from analysing the influence of the contingency awareness factor with a repeated measures ANOVA because for many participants at least one of the cells involving the interaction between contingency awareness and AA action was empty.

Evaluative priming task

In line with standard procedures for analyzing evaluative priming reaction time data (e.g., Spruyt, De Houwer, & Hermans, 2009), trials with an incorrect response were dropped (4.8 %) as well as any trials in which reaction times (RTs) were at least 2.5 standard deviations removed from an individual's mean (2.8 %). To perform the lmer analysis on evaluative priming task RTs we defined a model with the grouping variables *Participant* and *Target Word* as random factors. The random effect of *Face* was not included in the model because including this factor did not significantly improve model fit, $p > .99$.

To find out whether a standard AA training effect was obtained, we tested a model that contained Prime Face Type (approached, avoided), Target Type (positive, negative) and Contingency Instructions (yes, no) as fixed factors. We observed a main effect of Target Type, $\chi^2(1) = 4.82$, $p = .028$, indicating that participants were faster to respond to positive target words ($M = 560$, $SD = 140$) than to negative target words ($M = 587$, $SD = 154$). More importantly, this main effect was qualified by an interaction effect of Prime Face Type and Target Type, $\chi^2(1) = 8.62$, $p = .003$. RTs on trials with a positive target and approached face prime ($M = 556$, $SD = 141$) were faster than RTs on trials with a positive target and avoided face prime ($M = 563$, $SD = 139$), $\chi^2(1) = 3.36$, $p = .067$, 95 % confidence interval (CI) = [-14.16, 0.47], whereas RTs on trials with a negative target were slower when the prime was an approached face ($M = 592$, $SD = 159$; avoided face: $M = 582$, $SD = 150$), $\chi^2(1) = 5.40$, $p = .020$, 95% CI = [1.48, 17.37]. We observed no main or interaction effects involving the Contingency Instructions factor, $\chi^2s < 0.59$, $ps > .44$.

To investigate the role of contingency awareness, we added a Contingency Awareness factor to our model. For each participant, we classified each face as contingency aware, contingency indiscriminate (i.e., faces for which participants had indicated they had performed both actions an equal number of times) or contingency reversed[3]. This analysis corroborated the main effect of Target Type,

[3] This classification was based on participants' responses to the question what action they

$\chi2(1) = 5.18$, $p = .023$. However, we did not observe a significant interaction effect of Prime Face Type and Target Type, $\chi^2(1) = 1.78$, $p = .18$. Importantly, the predicted three-way interaction effect Prime Face Type x Target Type x Contingency Awareness was significant, $\chi^2(2) = 8.31$, $p = .016$ (Table 1). Further inspection of this interaction, showed that the two-way interaction effect of Prime Face Type and Target Type was significant only for trials with contingency aware faces, $\chi^2(1) = 10.91$, $p < .001$, indicating faster RTs on positive target trials with approached primes ($M = 546$, $SD = 134$) than with avoided primes ($M = 554$, $SD = 133$), $\chi^2(1) = 4.53$, $p = .033$, 95 % CI = [-17.99, -0.74], and slower RTs on negative target trials with approached primes ($M = 586$, $SD = 161$) than with avoided primes ($M = 569$, $SD = 144$), $\chi^2(1) = 4.63$, $p = .031$, 95 % CI = [0.96, 20.39]. This interaction effect was not observed on trials with contingency indiscriminate faces, $\chi^2(1) = 2.01$, $p = .16$, or contingency reversed faces, $\chi^2(1) = 1.72$, $p = .19$. For contingency reversed faces, however, we did observe a Prime Face Type x Target Type x Contingency Instructions interaction effect, indicating that only participants who had received contingency instructions exhibited a Prime Face Type x Target Type interaction effect, $\chi^2(1) = 4.75$, $p = .029$. However, post-hoc tests showed that for these faces participants who had received contingency instructions were significantly faster on *negative* target trials with approached primes ($M = 614$, $SD = 153$) compared to avoided primes ($M = 671$, $SD = 202$), $\chi^2(1) = 5.49$, $p = .019$, 95 % CI = [-82.70, -7.34], but not on positive target trials with approached primes ($M = 620$, $SD = 187$; avoided primes: $M = 629$, $SD = 163$), $\chi^2(1) = 0.85$, $p = .36$, 95 % CI = [-58.41, 21.84], indicating a preference for avoided faces. This Prime Face Type x Target Type x Contingency Instructions interaction effect was observed only for contingency reversed faces, and produced a four-way interaction effect, $\chi^2(2) = 11.06$, $p = .004$. All other main or interaction effects were non-significant, $\chi^2s < 0.59$, $ps > .44$.

---

had performed most often in response to a face. Note that this question actually registers participants' memory of co-occurrences between action and stimuli, which is merely an indication of participants' contingency awareness (see Gast et al., 2012).

Table 1.

Mean RTs and AA effects (in ms) in the evaluative priming task in Experiment 1 as a function of Target Type, Prime Face Type and Contingency Awareness.

	Positive Target		Negative Target		AA effect	
	Approached Face	Avoided Face	Approached Face	Avoided Face		
Contingency aware	546 (134)	554 (133)	586 (161)	569 (144)	13	*p* < .001
Contingency reversed	581 (160)	587 (150)	588 (138)	606 (166)	-6	*p* = .19
Contingency indiscriminate	568 (145)	576 (144)	619 (164)	609 (149)	9	*p* = .16

Note. Standard deviations are in parentheses. The AA effect was calculated by subtracting the mean latency of congruent trials (i.e., trials with positive target and approached prime or trials with negative target and avoided prime) from the mean latency of incongruent trials (i.e., trials with positive target and avoided prime or trials with negative target and approached prime).

Explicit rating scores

We defined a model with the grouping variables *Participant* and *Face* as random factors. To find out whether a standard AA training effect was obtained, we tested a model that contained only Face Type (approached, avoided) and Contingency Instructions (yes, no) as fixed factors. This revealed a main effect of Face Type, $\chi^2(1) = 14.30$, $p < .001$, indicating that participants preferred approached faces ($M = 4.03$, $SD = 1.28$) over avoided faces ($M = 3.68$, $SD = 1.27$), 95 % CI = [0.17, 0.55]. Similar to the results for evaluative priming RTs, we observed no main or interaction effects involving the Contingency Instructions factor, χ^2s < 0.07, *ps* > .79.

Analyses on the model that included the Contingency Awareness factor did not corroborate the main effect of Face Type, $\chi^2(1) = 0.09$, $p = .76$, but did show a significant interaction effect of Contingency Awareness and Face Type, $\chi^2(2) = 17.74$, $p < .001$ (Table 2). To investigate this interaction, we performed separate analyses for faces in each of the three different awareness categories. These analyses revealed a significant main effect of Face Type for contingency aware faces, $\chi^2(1) = 25.34$, $p < .001$, indicating a preference for approached faces ($M = 4.26$, $SD = 1.31$) over avoided faces ($M = 3.47$, $SD = 1.33$), 95 % CI = [0.40, 0.90].

The main effect of Face Type was not significant for contingency indiscriminate faces, $\chi^2(1) = 0.38$, $p = .54$. For contingency reversed faces we observed a marginally significant main effect of Face Type, $\chi^2(1) = 3.62$, $p = .057$. In contrast to the effect for contingency aware faces, participants preferred the avoided faces (*M* = 4.16, *SD* = 1.04) over approached faces (*M* = 3.35, *SD* = 1.19), 95 % CI = [-0.81, 0.01]. We observed no other main or interaction effects, χ^2s < 1.08, *p*s > .29. [4]

Table 2.

Mean explicit rating scores and AA effects in Experiment 1 as a function of Face type and Contingency Awareness.

	Approached Face	Avoided Face	AA effect	
Contingency aware	4.26 (1.31)	3.47 (1.33)	0.79	*p* < .001
Contingency reversed	3.35 (1.19)	4.16 (1.04)	- 0.81	*p* = .057
Contingency indiscriminate	3.89 (0.98)	4.03 (1.06)	- 0.14	*p* = .54

Note. Standard deviations are in parentheses. The AA effect was calculated by subtracting explicit rating scores for avoided faces from explicit rating scores for approached faces.

Discussion

Experiment 1 provided clear evidence that AA training caused changes in implicit and explicit evaluations of novel faces. Most importantly, however, our data indicate that contingency awareness is an important moderator of these AA training effects. Specifically, item-based analyses of awareness showed that participants exhibited a preference for approached faces over avoided faces only

[4] Both the evaluative priming task RTs and explicit rating scores were also analyzed with a standard repeated measures ANOVA. The results were similar to the reported effects with the exception that the interaction effect of Face Prime Type x Target Type x Contingency Awareness was not significant for the evaluative priming task RTs. In contrast to the lmer-analyses, we observed a marginally significant AA training effect for contingency indiscriminate faces. However, this analysis involved the data of only 25 participants because most participants did not have both an approached and an avoided face for which they indicated that both actions were performed an equal number of times. Because lmer analyses include all available data, while controlling for by-subject and by-item variation, we believe that these analyses provide more reliable information about the absence or presence of AA training effects and the factors that moderate these effects.

if they were able to correctly identify what action the stimulus had been paired with. In contrast, our manipulation of contingency awareness, which involved providing participants with information about the stimulus-action contingencies, failed to produce any evidence that contingency awareness influenced AA training effects even though the manipulation did influence measured contingency awareness.

Because our item-based analyses indicated that contingency awareness moderated AA training effects, it seems strange that AA training effects were not enhanced for participants who received contingency instructions. Some aspects of our data provide us with information that may help explain this data pattern. First, participants who did not receive contingency information correctly identified the face-action contingency above chance level. This indicates that, even in AA training studies where the target feature of the stimuli (e.g., identity) is task-irrelevant, participants can identify some of the stimulus-action contingencies. Consequently, contingency awareness may have influenced AA training effects even for participants who did not receive contingency information. The between-subjects analyses may simply have lacked the power to identify an added effect of instructions. In contrast, our contingency awareness analyses allowed us to gain more power because awareness was based on items rather than on participants (see Pleyers et al., 2007, for an argumentation why such item-based contingency awareness analyses are methodologically more sound than participant-based analyses). Second, participants who received contingency instructions exhibited an implicit preference for avoided faces over approached faces if they had incorrectly remembered the face-action contingencies. Because this contrast effect reduces the overall AA training effect and because this contrast effect occurred only if participants receive contingency instructions, this may have impeded the detection of a stronger overall AA training effect for participants who received contingency instructions compared to participants who did not receive contingency instructions.

In Experiment 1, we observed a preference for approached faces only if participants indicated correct awareness for the stimulus-action contingencies. This suggests that AA training effects occur only in the presence of contingency awareness. However, an alternative explanation is that our item-based contingency awareness measure was biased towards the conclusion that AA training requires awareness because participants relied on their liking of the stimulus to answer the contingency awareness questions (see Hütter et al., 2012). Contingency awareness questions asked participants to indicate whether they most often performed (a) approach actions, (b) avoid actions or (c) both actions an equal number of times in response to a face stimulus. The questions did not include a response option with which participants could indicate that they did not remember or had not identified the stimulus-action contingencies. In the absence of contingency awareness, this may have encouraged participants to search for other information that could help them answer these questions, including their liking of the stimuli. Importantly, what participants like could have been influenced by the (unconscious effects of) AA training. Hence, participants would select the correct response on the contingency awareness questions if they would select the response that has the same valence as the stimulus (i.e., select "approach" for liked stimuli and "avoid" for disliked stimuli). These responses would, however, not indicate actual contingency awareness but unconscious effects of AA training on liking (see Bar-Anan, De Houwer, & Nosek, 2010, and Hütter et al., 2012, for a similar argument in the context of EC).

EXPERIMENT 2

In Experiment 2, we sought to extend the findings in Experiment 1 in three ways. First, to reduce the possibility that participants base their answers to the contingency awareness questions on other information, we provided participants with the opportunity to indicate that they did not know the stimulus-action contingency. Second, we counterbalanced the order of the evaluative priming task and the explicit rating task to exclude the possibility that performing the implicit evaluation task first, changed the effects on explicit evaluations (see

Perugini, Richetin, & Zogmaister, 2014). Third, to focus and allocate test power to the question whether contingency awareness moderates AA training effects even if participants are never told that contingencies between stimuli and actions exist, none of the participants received any contingency information.

Method

Participants

A total of 64 native Dutch-speaking undergraduates participated in Experiment 2 (51 women). The data from two participants were discarded because their error rate in the evaluative priming task was more than 2.5 standard deviations above the population mean (population mean = 4.44 %, *SD* = 3.01%).

Procedure

Experiment 2 was identical to Experiment 1 except for the following changes. First, none of the participants received instructions specifying the stimulus-action contingencies. Second, subsequent to performing the AA training task, half of the participants first performed the evaluative priming task and then completed the explicit ratings. The other participants completed the explicit rating task before the evaluative priming task. Third, participants could choose between four response options for answering the contingency awareness questions (i.e., 'approach', 'avoid', 'I don't know', or 'both actions the same number of times').

Results

Contingency awareness

In line with Experiment 1, participants selected the correct action (M = 49%, SD = 27%) more often than the incorrect action (M = 19%, SD = 17%), $t(61) = 6.58$, $p < .001$. On average, participants indicated they did not know the correct action for 9% of the faces (SD = 14%) and indicated they had performed both actions an equal number of times for 22% of the faces (SD = 27%).

Evaluative priming task

We first performed an lmer analysis on RTs in the evaluative priming task. We defined a base model that included Target Type, Prime Face Type, and Task Order as fixed factors and *target word* and *subject* as random effects. In line with Experiment 1, we observed a main effect of Target Type, $\chi^2(1) = 4.72$, $p = .030$, indicating that participants were faster to respond to a positive target, as well as the crucial interaction effect of Prime Face Type x Target Type, $\chi^2(1) = 5.75$, $p = .017$. Participants were faster on trials with positive target and approached face ($M = 531$, $SD = 115$) than on trials with positive target and avoided face ($M = 538$, $SD = 125$), $\chi^2(1) = 4.72$, $p = .030$, 95% CI = [-12.97, -0.67], whereas no significant differences were found for trials with negative targets (approached face: $M = 563$, $SD = 129$, avoided face: $M = 560$, $SD = 129$), $\chi^2(1) = 1.77$, $p = .18$, 95% CI = [-2.17, 11.33]. We observed no other effects, $\chi^2 s < 1.29$, *p*s > .25.

In our second model, we included the Contingency Awareness factor (contingency aware, contingency reversed, contingency indiscriminate). Faces for which participants had indicated that both actions were performed an equal number of times and faces for which they had indicated that they did not know which action the face had been paired with, were collapsed in these analyses to reduce the number of empty cells for 'contingency indiscriminate' faces. The main effect of Target Type remained significant, $\chi^2(1) = 5.00$, $p = .025$, whereas the interaction effect of Target Type and Prime Face Type, $\chi^2(1) = 3.07$, $p = .080$, was only marginally significant. We observed an interaction effect of Target Type and Contingency Awareness, $\chi^2(2) = 6.24$, $p = .044$, indicating that the main effect of Target Type was larger for contingency reversed and contingency doubt faces than for contingency aware faces. Most importantly, we also found a marginally significant three-way interaction effect Target Type x Prime Face Type x Contingency Awareness, $\chi^2(2) = 5.17$, $p = .075$ (Table 3). Similar to Experiment 1, separate analyses revealed a significant Target x Prime interaction effect for contingency aware faces, $\chi^2(1) = 9.74$, $p = .002$, indicating faster RTs on positive target trials with approached primes ($M = 529$, $SD = 116$) than with avoided primes ($M = 539$, $SD = 131$), $\chi^2(1) = 6.41$, $p = .011$, 95% CI = [-22.10, -2.82], and slower RTs on negative target trials with approached primes (*M* = 551, *SD* = 119) than with avoided primes (*M* = 561, *SD* = 121), $\chi^2(1) = 2.64$, $p = .10$, 95% CI = [-1.66, 17.77]. We did not observe a significant interaction effect for contingency indiscriminate faces, $\chi2(1) = 0.20$, $p = .65$, or contingency reversed faces, $\chi^2(1) = 0.55$, $p = .46$. We observed no other main or interaction effects, $\chi^2s < 0.50$, *p*s > .47.

Table 3.

Mean RTs and AA effects (in ms) in the evaluative priming task in Experiment 2 as a function of Target Type, Prime Face Type and Contingency Awareness.

	Positive Target		Negative Target		AA effect	
	Approached Face	Avoided Face	Approached Face	Avoided Face		
Contingency aware	529 (116)	539 (131)	561 (121)	551 (119)	0	$p = .002$
Contingency reversed	534 (116)	529 (113)	563 (126)	550 (113)	4	$p = .46$
Contingency indiscriminate	532 (113)	541 (122)	565 (142)	581 (149)	-4	$p = .65$

Note. Standard deviations are in parentheses. The AA effect was calculated by subtracting the mean latency of congruent trials from the mean latency of incongruent trials.

Explicit rating scores

The base model for analyzing participants' explicit rating scores (internal consistency: mean Cronbach's Alpha = .84, *SD* = 0.04) included *Face* and *Participant* as random factors and Face Type (approached, avoided) and Task Order (evaluative priming task first, explicit rating task first) as fixed factors. This revealed only a main effect of Face Type, $\chi^2(1) = 7.41$, $p = .007$. Participants liked approached faces (*M* = 3.87, *SD* = 1.31) better than they liked avoided faces (*M* = 3.61, *SD* = 1.27), 95% CI = [0.08, 0.46]. No other effects were observed, χ^2s < 0.96, *ps* > .32.

When we added the Contingency Awareness factor to the model, the main effect of Face Type was not significant, $\chi^2(1) = 1.46$, $p = .23$. Importantly, we again observed a significant interaction effect of Face Type and Contingency Awareness, $\chi^2(2) = 23.13$, $p < .001$ (Table 4). A significant effect of Face Type was observed for contingency aware faces, $\chi^2(1) = 25.00$, $p < .001$, showing that approached faces were preferred ($M = 4.28$, $SD = 1.29$) over avoided faces ($M = 3.44$, $SD = 1.31$), 95% CI = [0.47, 1.07], but not for faces that participants did not indicate a specific action for, $\chi^2(1) = 1.66$, $p = .20$. In line with Experiment 1, a (non-significant) trend for a contrast effect was found for faces participants had indicated the incorrect action for, $\chi^2(1) = 2.64$, $p = .10$. We also observed a main effect of Contingency Awareness, $\chi^2(2) = 6.13$, $p = .047$, indicating that contingency aware faces were liked more than contingency reversed or contingency indiscriminate faces. No other effects were observed, χ^2s < 2.61, *ps* > .27.

Table 4.

Mean explicit rating scores and AA effects in Experiment 2 as a function of Face type and Contingency Awareness.

	Approached Face	Avoided Face	AA effect	
Contingency aware	4.28 (1.29)	3.44 (1.31)	0.84	$p < .001$
Contingency reversed	3.54 (1.31)	3.93 (1.33)	- 0.39	$p = .10$
Contingency indiscriminate	3.45 (1.16)	3.70 (1.15)	- 0.25	$p = .20$

Note. Standard deviations are in parentheses. The AA effect was calculated by subtracting explicit rating scores for avoided face from explicit rating scores for approached face.

Discussion

Experiment 2 corroborated that training to approach or avoid novel faces causes changes in implicit and explicit evaluations of these faces, and that these effects are strongly related to participants' awareness of the stimulus-action contingencies. Even though participants never received any information that the AA training procedure involved specific stimulus-action contingencies, they detected these contingencies at an above chance level. More importantly, AA training effects were observed only for contingency aware stimuli. This data pattern was observed even though participants had the opportunity to indicate that they did not know the contingencies, which renders it less likely that these effects are the result of biases in our contingency awareness measure.

In line with Experiment 1, we found no evidence that AA training effects can arise in the absence of awareness. To further corroborate this, we performed an lmer analysis on the data from both Experiment 1 and Experiment 2. The overall analysis showed the Prime Face Type x Target Type x Contingency Awareness interaction effect for evaluative priming task RTs, $\chi^2(2) = 8.25$, $p = .016$, and the Face Type x Contingency Awareness interaction effect for explicit rating scores, $\chi^2(2) = 38.53$, $p < .001$. To examine these interactions, we calculated indices of the AA effects for contingency aware, contingency reversed and contingency indiscriminate faces. For each type of face an index of the AA effect on implicit evaluation was calculated by subtracting participants' mean response latency for congruent trials (i.e., trials with positive target and approached prime or trials with negative target and avoided prime) from their mean response latency for incongruent trials (i.e., trials with positive target and avoided prime or trials with negative target and approached prime) and an index of AA effects on explicit evaluation was calculated by subtracting participants' mean explicit rating score for avoided faces from their mean explicit rating score for approached faces. In order to test whether these indices differed significantly from zero, we performed one-sample t-tests supplemented with Bayesian analyses. Bayesian analyses were performed according to the procedures outlined by Rouder, Speckman, Sun, Morey, and Iverson (2009). They provide a Bayes Factor that gives an indication of how strongly the data support either the null hypothesis (BF_0; reflecting the absence of a significant effect) or the alternative hypothesis (BF_1; reflecting the presence of a significant effect). BFs smaller than 1, between 1 and 3, between 3 and 10, and larger than 10, respectively designate 'no evidence', 'anecdotal evidence', 'substantial evidence', and 'strong evidence' for either the null or the alternative hypothesis (Jeffreys, 1961). We observed a significant AA effect for contingency aware faces on implicit, $t(103) = 3.97$, $p < .001$, $d = 0.39$, $BF_1 = 164$, and explicit evaluations, $t(103) = 6.52$, $p < .001$, $d = 0.64$, $BF_1 = 3861013$, and a contrast AA effect for contingency reversed faces on explicit, $t(47) = -3.08$, $p = .003$, $d = -0.44$, $BF_1 = 11$, but not implicit evaluations, $t(48) = 0.06$, $p = .95$, $d = 0.01$, $BF_0 = 5$. Importantly, we observed no significant AA effect for contingency indiscriminate faces on implicit, $t(48) = 1.07$, $p = .29$, $d = 0.15$, $BF_0 = 3$, nor explicit evaluations, $t(48) = -0.72$, $p = .48$, $d = -0.10$, $BF_0 = 4$, even though we had sufficient statistical power (power > .80) to detect small to medium effect sizes ($d > 0.35$).[5]

GENERAL DISCUSSION

In two experiments, we observed more positive evaluations of novel faces that were approached compared to faces that were avoided. We consistently found that contingency awareness, assessed by participants' memory for stimulus-action relations, moderated AA training effects on implicit and explicit evaluations. Both participants who received information specifying the stimulus-action contingencies and participants who did not receive contingency information preferred approached faces over avoided faces only if they were able to correctly identify what action they had performed most often in response to the specific face. Providing participants with contingency information via instructions, however, did not significantly influence AA training effects.

The role of contingency awareness in AA training effects

Our data provide the first evidence that contingency awareness moderates effects of AA training. Even though an imperfect measure was used to estimate contingency awareness (i.e., participants' memory for the stimulus-action contingencies), AA training effects were larger when participants reported awareness of the contingencies. This strongly resembles findings in the EC

[5] We report power estimates for the t-test analyses and not the lmer analyses because for mixed model analyses the required analytical tools for calculating the sampling distributions in situations where the null hypothesis is false are currently lacking. Because lmer analyses included the data of a larger number of participants who had expressed doubt about at least one of the stimulus-action contingencies (N = 82) and because these analyses control for by-subject and by-item variation, these analyses had more power to detect a significant effect of small effect size. These analyses also did not reveal a significant AA effect for contingency indiscriminate faces, $\chi 2s < 1.26$, $ps > .26$.

literature that contingency awareness is a potent moderator of EC (see Hofmann et al., 2010; Sweldens et al., 2014). We also found no evidence for unaware effects of AA training. When participants did not identify the stimulus-action contingencies correctly or expressed doubt about the contingencies we did not find a preference for approached faces over avoided faces even though our statistical tests had sufficient power to detect a small to medium sized effect. Bayesian analyses indicated that our data provided substantial evidence that AA training effects do not arise for contingency indiscriminate faces. Nevertheless, caution is warranted when drawing conclusions about the role of contingency awareness in AA training on the basis of our results. Importantly, our evidence is essentially correlational in nature. The liking of a stimulus was influenced by AA training only if participants could report the action that they performed in the presence of that stimulus during training. However, an experimental manipulation of contingency awareness did not influence AA training effects in the expected direction. Correlations do not reveal the direction of causality. Thus, although it is possible that AA training was related to contingency awareness because contingency awareness mediates the impact of training on liking, it is also possible that learning mediates the impact of training on contingency awareness. We will now consider two ways in which AA training could have mediated the changes in contingency awareness.

First, it is possible that the relation between contingency awareness and AA training effects in our study arose because our contingency awareness measure was influenced by the effects of AA training on liking. In EC research, there is a lot of discussion about the usefulness of contingency awareness measures and correlational approaches in general to address questions about contingency awareness (Gawronski & Walther, 2012; Dedonder, Corneille, Bertinchamps, & Yzerbyt, 2014). Most importantly, contingency awareness measures may be contaminated by reconstructive memory processes. That is, participants may complete these measures on the basis of other information, such as their liking of the stimulus (Hütter et al., 2012). Although we cannot completely exclude this alternative explanation, we did try to minimize guessing in Experiment 2 by giving participants the opportunity to indicate that they do not know the contingency. Importantly, we observed a strong relation between AA training effects and contingency awareness also in Experiment 2.

Second, even if the contingency awareness measure did capture contingency awareness rather than changes in liking, it is possible that contingency awareness itself was produced by the same processes that lead to changes in liking. For instance, one could assume that AA training leads to the formation of associations that in their turn produce both changes in liking and contingency awareness (see Lovibond & Shanks, 2002). In such a scenario, any factor that leads to variations in the strength of associations would lead to corresponding changes in both liking and contingency awareness, thus resulting in a correlation between AA training effects and contingency awareness. For instance, participants might have differed in the extent to which they attended the identity of the faces. Participants were asked to respond to the color of the frame surrounding face pictures. Hence, the task did not require that they processed face identity. Assuming that the formation of associations requires attention to the elements that are paired (e.g., Mackintosh, 1975; Wagner, 1981), participants who did pay attention to face identity might have formed stronger face-action associations in memory than participants who did not attend face identity. If association strength determines both changes in liking and contingency awareness, then inter-individual differences in attention to face identity might have resulted in a correlation between AA training effects and contingency awareness. Note, however, that this explanation is at odds with the observation that contingency instructions, which draw attention to the identity of the stimuli, did not cause stronger AA training effects in Experiment 1. Also, Vandenbosch and De Houwer (2011) included a manipulation designed to draw attention to the identity of the faces in four AA training studies, but did not observe an overall effect of training, suggesting that AA training effects critically depend on other boundary conditions.

Although alternative explanations are possible, the fact remains that our results are compatible with the view that AA training effects are mediated by contingency awareness. Contingency awareness might not only be *sufficient* for AA effects (as indicated by the fact that AA instructions alone can produce changes in liking; Van Dessel et al., 2015), it might also be necessary (as indicated by the fact that, in the current study, we observed changes in evaluations only when participants were able to consciously report the relation between face and action). Nevertheless, the idea that contingency awareness is necessary for AA training effects seems to contradict earlier studies in which AA training effects were observed when AA actions were performed in response to subliminally presented stimuli (Kawakami et al., 2007; Jones et al., 2011). However, in a recent attempt to replicate and extend these findings we failed to find any evidence for subliminal AA training effects (Van Dessel, De Houwer, Roets, & Gast, in press). Moreover, Bayesian analyses indicated that the original studies provided only anecdotal evidence for subliminal AA training effects while our results consistently provided substantial evidence for the absence of subliminal AA training effects. However, given the paucity of studies on this matter, more research is warranted to establish if and under what circumstances AA training causes changes in evaluation in the absence of conscious knowledge of stimulus-action contingencies

Importantly, we did not observe an overall effect of our experimental manipulation of contingency awareness on stimulus evaluations which seems difficult to reconcile with the interpretation of our results as evidence that contingency awareness causes AA training effects. As we previously contended, however, the absence of an effect of contingency instructions may have resulted from (a) a lack of power to detect such an effect in the between-subjects analysis, or (b) the fact that a larger proportion of participants revealed a reversed AA training effect in the contingency instruction group than in the no contingency instructions group. Another explanation might be that contingency instructions have effects other than making participants aware of the contingencies, effects that actually reduce the impact of AA training. For instance, some participants (e.g., participants with high levels of psychological reactance) may follow the goal to control against influences of these contingencies on evaluations. It seems possible that reactant responses might be more common after such instructed compared to merely observed contingencies.

Implications for mental process theories of AA training

The observation that contingency awareness moderates AA training effects seems to contradict the idea that AA training effects depend exclusively on implicit learning processes and does not fit well with current associative accounts of AA training effects. According to these accounts, the gradual formation of associations during action performance influences stimulus liking (e.g., Woud et al., 2013; Phills, Kawakami, Tabi, Nadolny, & Inzlicht, 2011). Because association formation is often considered an automatic process, there is no reason to assume that AA training effects should depend on participants' awareness of the contingencies (Kawakami et al., 2007). To accommodate our results, these traditional accounts of AA training would need to make a number of additional assumptions. First, in addition to automatic association formation, it seems necessary to postulate that another process, which critically depends on contingency awareness, also contributes to AA training effects. Demand compliance may serve as a likely candidate. However, the observation that contingency awareness moderated changes also in implicit evaluations, suggests that participants acquired a genuine preference for the approached stimuli which required contingency awareness (but see De Houwer, Beckers, & Moors, 2007). Second, it seems necessary to assume that the specific AA training procedure we used did not activate the implicit learning process sufficiently (and therefore impeded the detection of AA training effects for stimuli participants did not know the correct action for). For instance, AA training may cause changes in evaluations for contingency indiscriminate stimuli only when the training involves a sufficiently large number of training trials (e.g., because associative learning is a slow and gradual process; Rydell & McConnell, 2006). Support for this was found by Woud et al. (2011) who observed that AA training effects were stronger the more often faces were trained. However, because the addition of training trials may increase the likelihood that participants become aware of the stimulus-action contingencies, these findings could also reflect that AA training effects critically depend on contingency awareness. Also note that in our studies, we used a number of training trials that was comparable to that used in previous studies.

Alternatively, the strong impact that contingency awareness seems to have on AA training effects may be more easily explained by an alternative, propositional account of AA training effects. In line with propositional models of EC (De Houwer, 2009; Hofmann et al., 2010; Mitchell, De Houwer, & Lovibond, 2009), AA training may influence the liking of a stimulus only after participants acquired conscious propositional knowledge about the relation between action and stimulus. Participants may elaborate on this information and infer that the approached stimulus is positive (because they typically approach good things). Once this proposition is formed, this may influence both explicit and implicit stimulus evaluation (see De Houwer, 2014). From this perspective, AA training effects are driven by the acquisition of propositional information rather than by a 'training' mechanism that changes evaluations by gradually installing action tendencies to approach or avoid. It is important to note, however, that the current study used only neutral, unfamiliar faces as stimuli. In contrast to these novel stimuli, tendencies to approach or avoid specific well-known stimuli may have been acquired over a long learning history (e.g., spider phobics may have ample experience in avoiding spiders). To change evaluations of well-known stimuli, it may therefore be necessary to repeatedly perform AA actions in response to the stimulus such that the acquired tendencies are gradually re-trained (see Eberl et al., 2014). In line with this idea, we previously found that AA instruction effects were restricted to novel stimuli (Van Dessel et al., 2015). We hope that future research will explore whether contingency awareness is a critical factor also for studies that use AA training as a means of changing stimulus evaluation (e.g., in the treatment of alcohol addiction, Wiers et al., 2011; or social anxiety, Taylor & Amir, 2012).

Concluding remarks

This study indicates that AA training is an important procedure for the acquisition of evaluations of novel stimuli and provides the first evidence that contingency awareness is an important moderator of AA training effects. This conclusion contradicts the prevailing view concerning the automaticity of AA training effects and challenges theories that attribute AA training effects to the automatic acquisition of associations. These results add to recent work showing that various evaluative learning effects which were traditionally assumed to rely on automatic processes, strongly depend on awareness (e.g., EC: Hofmann et al., 2010; the mere exposure effect: de Zilva, Vu, Newell, & Pearson, 2013; mimetic desires: Bry, Treinen, Corneille, & Yzerbyt, 2011). They provide support for recent theoretical accounts that question the involvement of an automatic association formation mechanism in evaluative learning (e.g., De Houwer, 2009, Mitchell et al., 2009). It should be clear, however, that the issue of (evaluative) learning in the absence of awareness is still far from settled. In order for progress to be made, it is important to continue to carefully validate and replicate findings that do seem to provide evidence for unaware learning (e.g., Rydell, McConnell, Mackie, & Strain, 2006; Hu, Antony, Creery, Vargas, Bodenhausen, & Pallar, 2015). Once reliable evidence has been observed across multiple labs, efforts can start to identify the moderators that determine if and when awareness moderates learning.

However, it is important to repeat that our results provide only correlational evidence for a relation between contingency awareness and AA training. As such, it would be premature to make any conclusive statements about the causal role of contingency awareness in AA training effects. We hope, however, that our findings pave the way for additional AA training studies on the role of contingency awareness in AA training effects. Moreover, our findings point at the possibility that AA training effects are non-automatic in ways other than the need for contingency awareness. For instance, it would be interesting to examine the extent to which AA training effects can be controlled or depend on the availability of attentional resources. Future research on the automaticity features of AA training will provide important new information about the moderators of AA training effects and the mental processes that mediate those effects.

REFERENCES

Amir, N., Kuckertz, J. M., & Najmi, S. (2013). The effect of modifying automatic action tendencies on overt avoidance behaviors. *Emotion, 13,* 478-484. doi: 10.1037/a0030443

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390-412. doi: 10.1016/j.jml.2007.12.005

Baeyens, F., Eelen, P., & Van den Bergh, O. (1990). Contingency awareness in evaluative conditioning: A case for unaware affective evaluative learning. *Cognition & Emotion, 4*, 3-18. doi: 10.1080/02699939008406760

Bar-Anan, Y., De Houwer, J., & Nosek, B. (2010). Evaluative conditioning and conscious knowledge of contingencies: A correlational investigation with large samples. *Quarterly Journal of Experimental Psychology, 63,* 2313-2335. doi: 10.1080/17470211003802442

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-6. http://CRAN.R-project.org/package=lme4

Bry, C., Treinen, E., Corneille, O., & Yzerbyt, V. (2011). Eye'm lovin' it! The role of gazing awareness in mimetic desires. *Journal of Experimental Social Psychology, 47*, 987–993. doi: 10.1016/j.jesp.2011.03.023

Cacioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary determinants of attitudes. II: Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology, 65*, 5-17. doi: 10.1037//0022-3514.65.1.5

Chen, M., & Bargh, J. A. (1999). Nonconscious approach and avoidance behavioral consequences of the automatic evaluation effect. *Personality and Social Psychology Bulletin, 25,* 215-224. doi: 10.1177/0146167299025002007

Dedonder, J., Corneille, O., Bertinchamps, D., & Yzerbyt, V. (2014). Overcoming Correlational Pitfalls: Experimental Evidence Suggests That Evaluative

Conditioning Occurs for Explicit But Not Implicit Encoding of CS–US Pairings. *Social Psychological and Personality Science, 5*, 250-257. doi: 10.1177/1948550613490969

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

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

De Houwer, J. (2014). A Propositional Model of Implicit Evaluation. *Social and Personality Psychology Compass, 8*, 342-353. doi: 10.1111/spc3.12111

De Houwer, J., Beckers, T., & Moors, A. (2007). Novel attitudes can be faked on the Implicit Association Test. *Journal of Experimental Social Psychology, 43*, 972-978. doi: 10.1016/j.jesp.2006.10.007

de Zilva, D., Vu, L., Newell, B. R., & Pearson, J. (2013). Exposure Is Not Enough: Suppressing Stimuli from Awareness Can Abolish the Mere Exposure Effect. *PLoS ONE, 8(10)*, e77726. doi:10.1371/journal.pone.0077726

Eberl C., Wiers R. W., Pawelczack, S., Rinck, M., Becker, E. S. , Lindenmeyer J. (2014). Implementation of approach bias re-training in alcoholism-how many sessions are needed? *Alcoholism: Clinical and experimental research, 38,* 587-594. doi: 10.1111/acer.12281.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229–238. doi: 10.1037//0022-3514.50.2.229

Gast, A., De Houwer, J., & De Schryver, M. (2012). Evaluative conditioning can be modulated by memory of the CS-US Pairings at the time of testing. *Learning and Motivation, 43*, 116-126. doi: 10.1016/j.lmot.2012.06.001

Gast, A., & Rothermund, K. (2011). What you see is what will change: Evaluative conditioning effects depend on a focus on valence. *Cognition and Emotion, 25*, 89-110. doi: 10.1080/02699931003696380

Gawronski, B., & Walther, E. (2012). What do memory data tell us about the role of contingency awareness in evaluative conditioning? *Journal of*

Experimental Social Psychology, 48, 617-623. doi: 10.1016/j.jesp.2012.01.002

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*, 1464–1480. doi: 10.1037/0022-3514.74.6.1464

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods, 39,* 101-117. doi: 10.3758/bf03192848

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

Hu, X. Q., Antony, J. W., Creery, J. D., Vargas, I. M., Bodenhausen, G. V., & Pallar, K. A. (2015). Unlearning implicit social biases during sleep. *Science, 348 (6238)*, 1013-1015. doi: 10.1126/science.aaa3841

Huijding, J, & de Jong, P.J. (2005). A Modified Extrinsic Affective Simon Task (EAST) to assess the affective value of pictorial stimuli: No influence of age and educational level. *Psychologica Belgica, 45*, 241–255. doi: 10.5334/pb-45-4-241.

Huijding, J., Muris, P., Lester, K. J., Field, A. P., & Joosse, G. (2011). Training children to approach or avoid novel animals: Effects on self-reported attitudes and fear beliefs and information-seeking behaviors. *Behaviour Research and Therapy, 49*, 606-613. doi: 10.1016/j.brat.2011.06.005

Hütter, M., Sweldens, S., Stahl, C., Unkelbach, C., & Klauer, K. C. (2012). Dissociating contingency awareness and conditioned attitudes: Evidence of contingency-unaware evaluative conditioning. *Journal of Experimental Psychology: General, 141*, 539–557. doi: 10.1037/a0026477

Jeffreys, H. (1961). Theory of Probability. Oxford: Oxford University Press.

Jones, C. R., Vilensky, M. R., Vasey, M. W., & Fazio, R. H. (2013). Approach behavior can mitigate predominately univalent negative attitudes: Evidence regarding insects and spiders. *Emotion, 13*, 989-996. doi: 10.1037/a0033164

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder: Improving implicit racial evaluations and interracial interactions through approach behaviors. *Journal of Personality and Social Psychology, 92*, 957-971. doi: 10.1037/0022-3514.92.6.957

Laham, S. M., Kashima, Y., Dix, J., Wheeler, M., & Levis, B. (2014). Elaborated contextual framing is necessary for action-based attitude acquisition. *Cognition & Emotion, 28*, 1119-1126. doi: 10.1080/02699931.2013.867833

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion, 24,* 1377-1388. doi: 10.1080/02699930903485076

Locker, L., Hoffman, L., & Bovaird, J. A. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavior Research Methods, 39,* 723-730.doi: 10.3758/BF03192962

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes, 28*, 3–26. doi: 10.1037/0097-7403.28.1.3.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82,* 276–98. doi: 10.1037/h0076778

Markman, A. B., & Brendl, C. M. (2005). Constraining theories of embodied cognition. *Psychological Science, 16*, 6-10. doi: 10.1111/j.0956-7976.2005.00772.x

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

Neumann, R., & Strack, F. (2000). Approach and avoidance: The influence of proprioceptive and exteroceptive cues on encoding of affective information. *Journal of Personality and Social Psychology, 79*, 39–48. doi: 10.1037/0022-3514.79.1.39

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*, 413–417. doi: 10.1111/1467-9280.00376

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

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. doi: 10.1037/a0022159

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,* 130–144. doi: 10.1037/0278-7393.33.1.130

Rotteveel, M., Gierholz, A., Koch, G., van Aalst, C., Pinto, Y., Matzke, D., Steingroever, H., Verhagen, J., Beek, T.F., Selker, R., Sasiadek, A., & Wagenmakers, E.-J. (2015). On the automatic link between affect and tendencies to approach and avoid: Chen and Bargh (1999) revisited. *Frontiers in Psychology: Cognition*, *6*, 335. doi: 10.3389/fpsyg.2015.00335.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237. doi: 10.3758/PBR.16.2.225

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*, 995- 1008. doi: 10.1037/0022-3514.91.6.995

Rydell, R. J., McConnell, A. R., Mackie, D.M., & Strain, L.M. (2006). Of two minds: Forming and changing valence inconsistent implicit and explicit attitudes. *Psychological Science, 17*, 954–958.

Spruyt, A., De Houwer, J., & Hermans, D. (2009). Modulation of semantic priming by feature-specific attention allocation. *Journal of Memory and Language, 61,* 37 - 54. doi: 10.1016/j.jml.2009.03.004

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*, 220-247. doi: 10.1207/s15327957pspr0803_1

Sweldens, S., Corneille, O. & Yzerbyt, V. (2014). The Role of Awareness in Attitude Formation Through Evaluative Conditioning. *Personality and Social Psychology Review, 18*, 187-209. doi: http://dx.doi.org/10.1177/1088868314527832

Taylor, C. T., & Amir, N. (2012). Modifying automatic approach action tendencies in individuals with elevated social anxiety symptoms. *Behaviour Research and Therapy, 50*, 529-536. doi: 10.1016/j.brat.2012.05.004

Vandenbosch, K., & De Houwer, J. (2011). Failures to induce implicit evaluations by means of approach-avoid training. *Cognition and Emotion, 25,* 1311-1330. doi: 10.1080/02699931.2011.596819

Van Dessel, P., De Houwer, J., Gast, A., & Smith, C. T. (2015). Instruction-Based Approach–Avoidance Effects: Changing Stimulus Evaluation via the Mere Instruction to Approach or Avoid Stimuli. *Experimental Psychology, 62*, 161-169. doi: 10.1027/1618- 3169/a000282

Van Dessel, P., De Houwer, J., Roets, A., & Gast, A. (in press). Failures to Change Stimulus Evaluations by means of Subliminal Approach and Avoidance Training. *Journal of Personality and Social Psychology*.

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In: N.E. Spear & R.R. Miller (Eds.), In*formation processing in animals: Memory mechanisms* (pp. 5–47). Hillsdale, NJ: Erlbaum.

Wiers, R.W., Eberl, C., Rinck, M., Becker, E. & Lindenmeyer, J. (2011). Re-training automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science, 22,* 490-497. doi: 10.1177/0956797611400615

Woud, M. L., Becker, E. S., & Rinck, M. (2008). Implicit evaluation bias induced by approach and avoidance. *Cognition and Emotion, 22*, 1187-1197. doi: 10.1080/02699930801906843

Woud, M. L., Becker, E. S., & Rinck, M. (2011). Induction of implicit evaluation biases by approach-avoidance training: A commentary on Vandenbosch and De Houwer. *Cognition and Emotion, 25*, 1331-1338. doi: 10.1080/02699931.2011.628300

Woud, M. L., Maas, J., Becker, E.S., & Rinck, M. (2013). Make the manikin move: Symbolic approach-avoidance responses affect implicit and explicit face evaluations. *Journal of Cognitive Psychology, 25*, 738-744. doi: 10.1080/20445911.2013.817413

CHAPTER 5

FAILURES TO CHANGE STIMULUS EVALUATIONS BY MEANS OF SUBLIMINAL APPROACH AND AVOIDANCE TRAINING[1]

Previous research suggests that the repeated performance of approach and avoidance (AA) actions in response to a stimulus causes changes in stimulus evaluations. Kawakami, Phills, Steele, and Dovidio (2007) and Jones, Vilensky, Vasey, and Fazio (2013) provided evidence that these AA training effects occur even when stimuli are presented only subliminally. We also examined whether reliable AA training effects can be observed with subliminal stimulus presentations but added more sensitive checks of perceptual stimulus discriminability. Three experiments, including a direct replication of the study by Kawakami et al. (2007), failed to provide any evidence for effects of subliminal AA training on implicit or explicit evaluations. Bayesian analyses indicated that our data provide robust evidence that subliminal AA training does not cause changes in evaluations. In contrast, we observed changes in evaluations when participants were provided with (either correct or incorrect) information about the stimulus-action contingencies in the subliminal AA training task and when participants performed a supraliminal AA training task that allowed participants to detect these contingencies. These findings support the idea that contingency awareness is necessary for the occurrence of AA training effects.

---

[1] Based on Van Dessel, P., De Houwer, J., Roets, A., & Gast, A. (2016). Failures to Change Stimulus Evaluations by means of Subliminal Approach and Avoidance Training. *Journal of Personality and Social Psychology*, *110,* e1-e15. doi: 10.1037/pspa0000039

INTRODUCTION

It has been recognized for many decades that a person's behavior is largely determined by his/her likes or dislikes (Allport, 1935). Accordingly, understanding how preferences are formed and how they can be influenced is a fundamental research area in psychological science. Prior research showed that the repeated execution of approach and avoidance (AA) actions in response to a stimulus can cause changes in stimulus evaluations. When participants repeatedly approach one stimulus and avoid another stimulus, one typically observes a preference for the approached stimulus over the avoided stimulus (e.g., Kawakami, Steele, Cifa, Phills, & Dovidio, 2008; Laham, Kashima, Dix, Wheeler, & Levis, 2014, Woud, Maas, Becker, & Rinck, 2013; but see Vandenbosch & De Houwer, 2011). The first demonstration of this effect was provided by Kawakami, Phills, Steele, and Dovidio (2007). They found that participants who repeatedly approached photographs of Black people and avoided photographs of White people exhibited more positive evaluations of Black relative to White people on the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998).

In two of the four experiments reported by Kawakami et al. (2007), participants performed the AA movements in response to the words 'approach' or 'avoid'. Immediately preceding the presentation of these words, photographs of Black and White people were presented under conditions that limited the conscious detection of these stimuli (i.e., presentation in between two masking stimuli and for a duration of 23 ms). When photographs of Black people were consistently paired with approach movements and photographs of White people with avoidance movements, participants reported more positive implicit evaluations of Black relative to White people (Experiment 2) and showed more immediacy and openness for communication when interacting with a Black confederate (Experiment 4). The authors concluded that AA training effects may occur outside of participants' awareness of the contingencies in the training task.

Recently, a study by Jones, Vilensky, Vasey, and Fazio (2013) provided corroborative evidence for effects of AA training with subliminally presented stimuli. They observed changes in participants' explicit evaluations of insects when approach behaviors, performed in response to the presentation of the word 'TOWARD', were repeatedly paired with masked images of insects, presented for 13 ms (Experiment 1). Moreover, when participants' repeatedly performed approach behaviors in response to subliminally presented images of spiders, they exhibited more positive implicit evaluations of spiders (Experiment 2) and reported reduced anxiety ratings when encountering live spiders (Experiment 3).

The observation of subliminal AA training effects is in line with the idea that AA training effects are due to automatic association formation mechanisms. This idea entails that the repeated pairing of an AA action and a stimulus automatically results in the gradual formation of associative links between the representations of the stimulus and positively or negatively valenced representations (Phills, Kawakami, Tabi, Nadolny, & Inzlicht, 2011; Woud et al., 2013).

Recent evidence has, however, challenged the idea that AA training effects are (exclusively) the result of automatic associative learning processes. In two studies that were conducted at our laboratory, we provided evidence that AA training effects are moderated by awareness of the stimulus-action contingencies (Van Dessel, De Houwer, & Gast, 2015). When participants repeatedly performed AA actions in response to novel face stimuli, they exhibited a preference for the approached stimuli on implicit and explicit measures of evaluation – but only when they were able to correctly identify what action they had performed most often in response to the stimuli. We obtained no evidence that AA training caused changes in stimulus evaluations in the absence of contingency awareness.

The observation that contingency awareness moderates AA training effects does not fit well with the idea that automatic association formation underlies these effects. To accommodate these findings it seems necessary that traditional association formation accounts make a number of additional assumptions (e.g., that the formation of associations depends on specific boundary conditions such as whether sufficient attention is attributed to the identity of the stimulus; see Van Dessel, De Houwer, & Gast, 2015). In contrast, alternative, propositional accounts of AA training may more easily explain these results. These accounts might, for instance, entail that participants who acquire stimulus-action contingency information may elaborate on this information and infer that the approached stimulus is positive (because they typically approach good things). Once this proposition is formed, this may influence both explicit and implicit stimulus evaluation (see De Houwer, 2014). In line with this account, a recent set of studies showed that typical AA training effects occurred even when participants did not perform AA actions but were merely instructed that they would later on have to perform these actions (Van Dessel, De Houwer, Gast, & Smith, 2015). Participants who received instructions to approach one fictitious social group (e.g., Niffites) and avoid another fictitious social group (e.g., Luupites) exhibited a preference for the former group both on implicit and explicit measures of evaluation.

Although AA learning via instructions suggests that conscious knowledge of stimulus-action contingencies is sufficient for acquiring AA effects, it does not exclude the possibility that, under certain circumstances, AA training effects may occur in the absence of contingency awareness (e.g., as the result of the automatic formation of associations). Therefore, it is important to actively establish evidence or seek to confirm potential evidence for AA training effects that occur in the absence of contingency awareness. Demonstrating that AA training effects can occur after subliminal stimulus presentations provides the strongest case that contingency awareness is not a necessary condition for these effects to occur. Hence, subliminal AA training effects are theoretically important because they provide strong evidence for automatic association formation models of AA training. Such evidence would strongly constrain propositional accounts of AA training. For instance, to explain subliminal AA training effects, these accounts would have to assume that propositions (e.g., ‘I’d like to approach Black people') can be formed even if one is unaware of the perceptual stimulation that initiated the formation of this proposition (e.g., photographs of Black people).

As we noted above, two sets of studies provided support for the existence of subliminal AA training effects (Kawakami et al., 2007; Jones et al., 2013). If the conclusion drawn from these studies (i.e., that AA training effects can occur after subliminal presentations) is valid, these studies present a very strong case for the possibility of AA training effects in the absence of contingency awareness and hence for automatic association formation as an underlying mechanism. In our opinion, however, there are two reasons why this conclusion can currently not be drawn.

First, both studies suffer from an important methodological limitation: because they did not include objective measures of stimulus visibility they could not assure that every presentation of the stimulus was indeed presented below the threshold of conscious awareness ('sub-liminally'). To establish whether participants were aware of the rapidly presented stimuli, Kawakami et al. (Experiments 2 and 4), and Jones et al. (Experiment 1) used a funnel debriefing procedure to question participants for stimulus awareness after the experiment. When participants expressed any awareness of the identity of the stimuli, their data were excluded from the analyses. In the subliminal perception literature, however, there is a lot of debate about whether such subjective self-report measures provide an accurate and reliable method to determine awareness (for a summary of this discussion, see Snodgrass, Bernat, & Shevrin, 2004). For instance, a simple lack of confidence may contribute to someone's reluctance to report having been aware of a particular stimulus, despite having some subjective experience of it (Cleeremans, 2001). Jones et al. did not use any awareness tests in Experiments 2 and 3, but they did conduct a pilot test in which 12 participants first completed the approach spiders training and then indicated for 40 stimuli, 10 of which were spider pictures, whether they had seen these stimuli during the training task. None of the participants reported seeing any of the spider pictures, suggesting that the pictures had not been perceived consciously. This procedure, however, may suffer from the same limitations as the funnel debriefing procedure (e.g., issues related to memory recall or a lack of confidence). Moreover, it does not take into account that there may be individual differences in people's ability to detect rapidly presented pictures. The authors address this limitation in their discussion section where they state that they "cannot rule out the possibility that some presentations of images were correctly identified or that any awareness that occurred contributed to the effect on attitude change." (p.995).

Second, even if one assumes that the method is suitable to exclude conscious awareness of the stimuli, the empirical support provided by the two studies might not be sufficient to yield substantial evidence for the hypothesis that AA training is indeed possible with subliminal stimulus presentation. In order to investigate this issue, we performed Bayesian analyses. Bayesian analyses provide a Bayes Factor that denotes the weight of evidence provided by the data for competing hypotheses. As such, Bayes Factors can provide an indication of how strongly the data support either the null hypothesis (BF_0; reflecting the absence of a significant effect) or the alternative hypothesis (BF_1; reflecting the presence of a significant effect). BF scores can be computed for both the null and the alternative hypothesis. BF scores smaller than 1, between 1 and 3, and between 3 and 10, respectively designate 'no evidence', 'anecdotal evidence', and 'substantial evidence' for either the null or the alternative hypothesis (Jeffreys, 1961). For instance, when $BF_0 = 10$ (and $BF_1= 1/10 = .10$) the observed data are 10 times as likely to have occurred under the null hypothesis than under the alternative hypothesis, providing substantial evidence for the null hypothesis. When $BF_1 = 2.5$ (and $BF_0 = 1/2.5 = .40$) the observed data are 2.5 times as likely to have occurred under the alternative hypothesis than under the null hypothesis, providing anecdotal evidence for the alternative hypothesis. We reanalyzed the critical t-tests reported in Jones et al. (2011), and Kawakami et al. (2007) for all experiments that showed effects of subliminal AA training on stimulus evaluations by computing a Bayes factor (BF) according to the procedures outlined by Rouder, Speckman, Sun, Morey, and Iverson (2009).

Though the obtained BF scores always provided evidence in favor of the alternative hypothesis (that subliminal AA training causes significant changes in implicit evaluations), the evidence was only anecdotal (Kawakami et al. Experiment 2: BF_1 = 2.31; Jones et al. Experiment 1: BF_1 = 1.64; Jones et al. Experiment 2: BF_1 = 1.26). This attests that replication is warranted to establish the robustness of the subliminal AA training effect. We also assessed the available evidence across these experiments by performing a Bayes factor meta-analysis (Rouder & Morey, 2011). This analysis did provide strong evidence for subliminal AA training effects, BF_1 = 46.46 (Table 1). However, this value may be artificially inflated because no unsuccessful attempts have ever been reported, possibly due to the file-drawer problem.

Table 1.

Subliminal AA training Experiment Results.

Experiment	Evaluation object	Evaluation task	N	Test statistic	Bayesian t-test
Kawakami et al. (2007, Exp. 2)	Black people	IAT	50	$t(47) = 2.28$, $p = .027$	BF_1 = 2.31, Anecdotal *(H_1)*
Jones et al. (2011, Exp. 1)	Insects	Rating task	42	$t(40) = 2.08$, $p = .044$	BF_1 = 1.64, Anecdotal *(H_1)*
Jones et al. (2011, Exp. 2)	Spiders	Personalized IAT	118	$t(116) = 2.04$, $p = .044$	BF_1 = 1.26, Anecdotal *(H_1)*
All previous experiments			210		BF_1 = 46.46, Strong *(H_1)*
Experiment 1	Black people	IAT	62	$t(60) = -0.21$, $p = .83$	BF_0 = 4.49, Substantial *(H_0)*
		Rating task		$t(60) = 0.19$, $p = .85$	BF_0 = 3.34, Substantial *(H_0)*
Experiment 2	Unfamiliar faces	Priming Task	76	$t(74) = -0.17$, $p = .87$	BF_0 = 6.49, Substantial *(H_0)*
		Rating task		$t(74) = -0.28$, $p = .78$	BF_0 = 7.00, Substantial *(H_0)*
Experiment 3	Non-words	IAT	96	$t(95) = -0.52$, $p = .60$	BF_0 = 9.14, Substantial *(H_0)*
		Rating task		$t(95) = -0.91$, $p = .37$	BF_0 = 11.38, Strong *(H_0)*

All current experiments	Implicit evaluation	234	BF_0 = 10.34, Strong *(H_0)*
	Explicit evaluation		BF_0 = 10.82, Strong *(H_0)*
All subliminal AA experiments	Implicit evaluation	352	BF_0 = 3.64, Substantial *(H_0)*
	Explicit evaluation	326	BF_0 = 6.51, Substantial *(H_0)*

We performed the current research in order to gain more information on whether AA training effects can be observed with subliminal stimulus presentations. Most importantly, we included more sensitive checks of perceptual stimulus discriminability that are based on participants' forced-choice decisions regarding the identity of the stimulus by using d' measures. These measures are widely used in studies with subliminal stimulus presentations because they provide an objective assessment of stimulus discriminability (Merikle, Smilek, & Eastwood, 2001). If we would find robust AA training effects also when these measures indicate that participants are unable to discriminate the stimuli, this would constitute important evidence that AA training can change evaluations in the absence of awareness of the stimuli and thus stimulus-action contingencies. If, on the other hand, AA training effects strongly depend on participants' ability to discriminate the stimuli, this would be consistent with the idea that AA training effects necessarily involve stimulus awareness.

In this paper, we report three experiments. Experiments 2 and 3 examined subliminal AA training effects for evaluations of novel stimuli, whereas Experiment 1 was an exact replication of the experiment by Kawakami et al. (2007, Experiment 2) who found subliminal AA training effects on evaluations of Black and White social groups. This experiment was selected for replication because it provided the 'strongest' evidence in favor of subliminal AA training effects (BF_1 = 2.31). We used Kawakami et al.'s exact procedures and materials and extended upon Kawakami et al. by (1) including a d' measure of stimulus perceptibility to examine whether stimulus presentations were indeed subliminal, (2) including also explicit measures of evaluations to assess subliminal

AA training effects, and (3) supplementing the data analytic strategies that Kawakami et al. used for the investigation of the AA training effects with Bayesian analyses.

EXPERIMENT 1

Method

Participants

Sixty-two native Dutch-speaking undergraduates (49 women) participated in exchange for a monetary reward of 7 euros. This sample size was determined by performing a power analysis according to the procedures recommended by Cohen (1988) with the aid of G-Power software (version 3.1.). We ensured that the power to obtain an effect size of d = 0.67 (the effect size observed by Kawakami et al., 2007) was greater than .80 (achieved power = .84). All participants had normal or corrected-to-normal vision and were naive with respect to the purpose of the experiment. Participants were randomly assigned to an approach Blacks training condition where they received training to approach subliminally presented photographs of Black people and avoid subliminally presented photographs of White people or to a control condition where they received training to make a leftward or rightward movement in response to the subliminal presentation of photographs of Black or White people[1].

Apparatus and Materials

The experiment was programmed and presented using the Direct RT Empirisoft Software package (DirectRTv2012) on a Tori PC with a 19-inch monitor (85 Hz refresh rate). The stimuli for the AA training task and the IAT (i.e., black-

[1] In contrast to the procedure by Kawakami et al. (2007), we decided not to include a condition where participants received training to avoid photographs of Black people and approach photographs of White people. This was done out of ethical concerns and because participants in the control condition and avoid Blacks condition typically do not exhibit significant changes in their evaluations (see Kawakami et al.; Phills et al., 2011).

and-white photographs of 30 Black faces, 30 White faces, and 48 moonscapes) and the script for the subliminal AA training task were provided to us by one of the authors of Kawakami et al. (2007).

Procedure

Participants were seated at a desk in an individual cubicle in front of a computer with a keyboard and joystick (Logitech Wingman) attached to it. After participants had given informed consent, they were informed that they would perform a series of unrelated tasks. Participants in the approach Blacks training condition were instructed to pull the joystick toward themselves when the word 'approach' was presented and to push the joystick away from themselves when the word 'avoid' was presented. Participants in the control condition were instructed to push the joystick to the right when presented with the word 'right' and to push the joystick to the left when presented with the word 'left'.

The AA training task consisted of ten blocks of 48 trials. On each trial, a forward mask consisting of a photograph of a moonscape was presented for 300 ms and followed by a photograph of a Black or White person's face presented for 23.52 ms (two refresh cycles). A backward moonscape mask was then presented for 35.29 ms (three refresh cycles) and followed by the word "approach" or the word "avoid" (in the approach Blacks training condition) or the word "left" or the word "right" (in the control condition)[2]. In the approach Blacks training condition, a Black face was always followed by the word approach and a White face was followed by the word avoid. In the control condition, for half the participants, the word "left" followed a Black face and the word "right" followed a White face. For the other half of the participants of the control condition, the contingencies of left/right word and Black/White face were reversed. The word remained on screen until participants responded by moving the joystick. After participants made a correct response, a blank screen was presented for 1000 ms before the start of the next trial. After incorrect responses, participants were

[2] We ensured by careful pre-testing that presentation times of photographs and backward masks, as registered by the Direct RT program, were exactly 23.52 ms and 35.29 ms on each presentation.

presented with a blank screen for 100 ms, followed by the presentation of a red X in the middle of the screen for 800 ms. Another blank screen was presented for 100 ms before the start of the next trial.

Participants then performed an IAT where they categorized photographs of six Black and six White faces which had not been used in the AA training task along with positive and negative words. In accordance with standard IAT procedures (Greenwald et al., 1998), participants were presented with five blocks of trials. Participants started with two practice blocks, one in which they categorized the photographs into the categories Black or White and one in which they categorized words as positive or negative. Categorization was done by pushing a left ('Q') or right key ('M') on an AZERTY keyboard. The practice blocks were followed by a critical block in which participants categorized both words and photographs. The practice block with Black and White faces was then repeated, but the response assignments were reversed. Finally, the critical block was repeated with the reversed response assignment for the Black and White faces. Each critical block consisted of 60 trials in which a word or photograph was presented in the center of the screen until the participant pressed one of the two valid keys. If the response was correct, the word disappeared and the next trial started 400 ms later. If the response was incorrect, the word was replaced by a red "X" for 400 ms. The next trial started 400 ms after the red "X" was removed from the screen.

After the IAT, we assessed participants' explicit evaluations of Black and White people. First, participants completed liking ratings and thermometer ratings of self-reported feelings of warmth towards Black and White people on two 9-point Likert scales (1 = not liked/warm at all; 9 = completely liked/warm). Second, participants completed a ten-item version of Pettigrew and Meertens' (1995) questionnaire assessing subtle and blatant racial prejudice, adapted to a Belgian context (Van Hiel & Mervielde, 2005).

Subsequently, a funnel debriefing procedure was used to question participants for awareness of the stimuli, stimulus-action contingencies and research hypotheses. Participants indicated: (1) what the purpose of the experiment and of the joystick task was, (2) what the relationship was between the joystick task and the categorization task, (3) whether they had noticed anything suspicious about the background in the joystick task, and (4) what the specific content of the background flashes was. We also asked participants to indicate what the stimulus-action contingencies were, but only if they had correctly identified the nature of the stimuli.

Finally, participants performed a perceptibility task in which they categorized the photographs of Black and White persons' faces masked under the same conditions as in the AA training task. The task consisted of two blocks of 48 trials. Trials were identical to the AA training trials with the exception that (1) target words were replaced with strings of 'XXXX' and (2) participants did not perform AA actions but responded to the photographs that were presented before these strings. They responded by pushing the 'Z' key when the photograph depicted the face of a Black person and the 'B' key when the photograph depicted the face of a White person. Participants did not receive feedback about the accuracy of their responses.

Results

Awareness and Perceptibility

Thirteen participants (21 %) indicated that photographs of faces had been presented during the AA training task. Seven participants (11%), 6 of which were in the approach Blacks training condition, indicated that these were faces of Black and White people. Of these participants, six (10%) expressed suspicion that the purpose of performing the AA training task was related to the purpose of performing the IAT and four (6%), all in the approach Blacks training condition, indicated that the purpose of the AA training task was to target racial prejudice by approaching faces of Black people.

The overall accuracy in the perceptibility task was 54% (*SD* = 6%). We computed the signal detection sensitivity measure d' with hits defined as 'Black' responses on trials with photographs of Blacks, and false alarms as 'Black' responses on trials with photographs of Whites. The d' score indicated that participants' detection performance was (marginally significantly) above chance level (M = 0.17, SD = 0.67), $t(61)$ = 1.98, p =.053. Twenty-two participants (35%), 12 in the approach Blacks training condition, including 6 of the participants who indicated they had seen faces of Black and White people, had an individual d' score above the 95% confidence interval (d' > 0.34), which indicates a potential capability to see the masked prime (Macmillan & Creelman, 2005). Performance of two of these participants was very high (d' > 1) even though they had not reported awareness of the stimuli in the debriefing questions.

Implicit evaluation

IAT D4 scores were calculated following the procedure by Greenwald, Nosek and Banaji (2003). Positive scores reflect a preference for White people over Black people. Similar to Kawakami et al. (2007), a one-sample t-test revealed that participants displayed a strong implicit preference for White people (M = 0.23, SD = 0.40), $t(61)$ = 4.47, p < .001, d = 0.58, BF_1 = 595.87. However, contrasting Kawakami et al.'s findings, a between-subjects t-test did not reveal a significant difference in IAT scores for participants who approached Blacks and avoided Whites (M = 0.24, SD = 0.38) compared to participants in the left/right control condition (M = 0.22, SD = 0.42), $t(60)$ = -0.21, p =.83, d = 0.05. The BF score provided substantial evidence in favor of *the null hypothesis* (BF_0 = 4.49). We performed an additional t-test excluding the data of the 13 participants who expressed awareness of the presentation of subliminal presentation of faces. Again, we observed no subliminal AA training effect, $t(47)$ = 0.29, p =.78, d = 0.07, BF_0 = 3.01. We also did not observe significantly reduced IAT scores for those participants in the approach Blacks training condition who had reported seeing photographs of Black and White people, $t(5)$ = 0.83, p =.78, BF_0 = 1.70, or who had an individual d' score above the 95% confidence interval, $t(11)$ = 0.47, p =.68, BF_0 = 2.46. Note that the latter analyses revealed only anecdotal evidence in favour of the null hypothesis, which is not surprising given that they included only a very small sample of participants.

Explicit evaluation

We calculated liking rating scores (M = 0.19, SD = 0.33) and warmth rating scores (M = 0.15, SD = 0.29) by subtracting each participants' liking and warmth ratings for Black people from the corresponding ratings for White people. A score for subtle racism (M = 3.77, SD = 0.86, Cronbach's α = 0.51) and a score for blatant racism (M = 1.80, SD =0.69, Cronbach's α = 0.73) was calculated by summing the ratings for the items in the racism scales and dividing this by the number of items. The four resulting explicit evaluation scores correlated significantly (blatant racism – subtle racism: r[60] = 0.57; blatant racism – liking rating: r[60] = 0.32; blatant racism – warmth rating: r[60] = 0.47; subtle racism – liking rating: r[60] = 0.39; subtle racism – warmth rating: r[60] = 0.32; liking rating – warmth rating: r[60] = 0.70), ps < .012. The IAT score correlated significantly with the liking rating score, r(60) = 0.29, p =.020, and warmth rating score, r(60) = 0.30, p =.016, but not with the blatant racism scale score, r(60) = 0.08, p =.54, or the subtle racism scale score, r(60) = 0.18, p =.17. Similar to the IAT score, the explicit rating scores indicated a significant preference for White people over Black people. Participants' liking and warmth ratings were higher for White people (liking: M = 6.85, SD = 1.34; warmth: M = 6.61, SD = 1.40) than for Black people (liking: M = 6.11, SD = 1.48; warmth: M = 6.02, SD = 1.59), ts > 4.03, ps <.001, BF_1s> 153.60. Most importantly, however, none of the explicit evaluation scores revealed significant differences between the approach Blacks training group and the control group, ts < 0.32, ps >.74, ds < 0.01, BF_0s > 2.99.

Discussion

Despite using the same subliminal AA training procedure as Kawakami et al. (2007), we were unable to replicate the effect on implicit evaluations (or provide evidence for an additional effect on explicit evaluations). To account for these discrepant results, a number of explanations can be considered. First, though power calculations were used to determine the sample size and ascertain that we had sufficient power to detect the effect size observed by Kawakami et al., we may have lacked the power to detect an effect of smaller effect size (i.e., for a medium effect: power = 0.62; for a small effect: power = 0.25). However, Bayesian analyses indicated that our findings already provide substantial evidence that subliminal AA training does not allow for changes in participants' implicit or explicit evaluations. Second, as suggested to us by one of the authors of Kawakami et al., the discrepancy between our results and the original results may relate to important differences in participants' baseline levels of racial prejudice due to cross-cultural differences. Our participant sample consisted of Belgian undergraduates, whereas participants in Kawakami et al.'s study were undergraduates from North America. Prejudice towards Black people may be more relevant (and more robust) for the latter population (e.g., due to a long history of conflict between these two racial groups; see Perlmutter, 1999). Because previous evidence suggests that AA training effects are observed only when the training is inconsistent with participants' racial bias (e.g., Phills et al., 2011), a smaller overall effect of AA training may be observed if a smaller subset of participants have this bias in our sample, which may explain why we did not find any subliminal AA training effects. Nevertheless, it should be noted that in our study, we did observe a strong preference for White people over Black people both on implicit and explicit measures of evaluation. To corroborate this, we examined the IAT scores of volunteers with Belgian or US nationality who completed a race IAT on the Project Implicit research website (https://implicit.harvard.edu) in the years 2002-2012 (Xu, Nosek, & Greenwald, 2014). We observed a significant preference for White people over Black people for both Belgian (M = 0.44, SD = 0.39), $t(71)$ = 9.46, p <.001, d = 1.13, and US participants (M = 0.37, SD = 0.43), $t(30343)$ = 152.19, p <.001, d = 0.86. Importantly, The IAT prejudice score was even slightly *lower* for US participants than for Belgian participants. Hence, the lack of substantial AA effects in this study could not be attributed to lower baseline levels of racial prejudice toward Black people in the present sample.

Alternatively, the effect that was observed by Kawakami et al. (2007) may have simply been a Type I error. In line with this idea, Bayesian analyses indicated that the data obtained by Kawakami et al. provided only anecdotal evidence in favor of the alternative hypothesis (i.e., that IAT scores differed significantly between approach Blacks training and control conditions). In contrast, our data provided substantial evidence for the null hypothesis. Another possible explanation is that in the experiment reported by Kawakami et al., some participants identified the stimulus-action contingencies and, though they failed to report this in the awareness questions, fuelled the effect. In line with this idea, our experiment provided evidence that (1) with the reported presentation times, participants are sometimes able to discriminate the race of the person presented in the photograph, and (2) participants may not always report this in the debriefing questions (e.g., the two participants with the best performance in the stimulus discriminability task did not report stimulus awareness). Moreover, we found that participants who performed the subliminal approach Blacks training (but not participants in the control group) sometimes identified the purpose of performing the AA training task. This may reduce participants' implicit prejudice if, for instance, participants adapt to the demands of the interpersonal context (Castelli & Tomelleri, 2008; Richeson & Ambady, 2003). In our study, however, this did not cause any significant differences in evaluations between participants in the approach Blacks training group and the control group. More specifically, we did not even find AA training effects for participants who were able to discriminate the race of the presented persons or reported seeing pictures of Black or White persons.

EXPERIMENT 2

Experiment 1 failed to provide support for the idea that training to approach respectively avoid subliminally presented photographs of Black and White people causes changes in the evaluations of these well-known social groups. In Experiments 2 and 3, we shifted our focus to the investigation of subliminal AA training effects for novel stimuli[3]. Subliminal AA training effects may be more robust for these stimuli because AA effects are more easily

[3] In both experiments the novel stimuli were introduced shortly before participants performed the AA training task to ensure that participants would not have to create a representation of the stimulus under conditions of subliminal presentation.

established for neutral as compared with initially valenced stimuli (e.g., Priester, Cacioppo, & Petty, 1996; Woud, Becker, Lange, & Rinck, 2013; Van Dessel, De Houwer, Gast, & Smith, 2015). This finding is related to the general idea that changing preferences is more difficult than establishing novel preferences (e.g., Gregg, Seibt, & Banaji, 2006). The stimuli we used in Experiment 2 were unfamiliar faces. Previous AA training studies have provided robust effects for these stimuli (Woud et al., 2008; 2013; Van Dessel, De Houwer, & Gast, 2015; but see: Vandenbosch & De Houwer, 2011) and evidence suggests that AA effects are larger for face stimuli compared to other pictorial stimuli (Laham, Kashima, Dix, & Wheeler, in press).

The purpose of Experiment 2 was twofold: (1) to establish whether changes in implicit and explicit evaluations of novel face stimuli can arise as the result of subliminal AA training, and (2) to compare effects of subliminal AA training with effects that result from instructions about stimulus-action contingencies. To this end, participants were informed about the contingencies between the four face stimuli and the AA actions in the subliminal AA training task before they performed the task. Unbeknownst to participants, the contingency information was incompatible with the AA training for half of the stimuli. Consequently, each face was assigned to one of the following four conditions: (1) approach training and approach instructions, (2) approach training and avoid instructions, (3) avoid training and approach instructions, or (4) avoid training and avoid instructions. Participants' evaluations of the faces were registered with explicit evaluative ratings and an implicit measure of evaluation (the evaluative priming task; Fazio, Sanbonmatsu, Powell, & Kardes, 1986).

This experimental set-up allowed us to test specific predictions derived from associative and propositional accounts of AA training effects. If AA training effects arise as the result of the gradual automatic formation of associations, then changes in evaluations should reflect the contingencies in the subliminal AA training task. These effects may be observed more strongly on implicit evaluations because implicit evaluations are considered to reflect the automatic activation of associations in memory (Gawronski & Bodenhausen, 2011). If, on the other hand, AA training effects critically depend on the acquisition of propositional information, then changes in evaluations should always be in line with the provided contingency information. Acquiring this information may influence both explicit and implicit stimulus evaluations (see De Houwer, 2014).

Method

Participants

Seventy-six native Dutch-speaking undergraduates (61 women) participated in exchange for a monetary reward of 5 euros. None of the participants had previously participated in Experiment 1.

Apparatus and Materials

Four photographs of female faces, selected from the set of materials used by Van Dessel, De Houwer, and Gast (2015), served as stimuli for the AA training task. For the evaluative priming task, seven positive words (the Dutch words for happy, pleasant, sweet, kind, friendly, sympathetic, and fun), and seven negative words (the Dutch words for unfriendly, irritating, hostile, bad, moody, unpleasant, and mean) were selected to serve as target stimuli and the four photographs of faces were used as primes.

The experiment was programmed in C-language and presented using the C-library Tscope package (Tscope 1.0.171.) on a Tori PC with a 19-inch monitor (80 Hz refresh rate), a keyboard and a joystick (Wingman Attack 2) attached to it.

Procedure

After participants had given informed consent, they were seated in front of a computer screen. Half of the participants read the following instructions (translated from Dutch):

In this experiment, you will see photographs of different faces. You will approach a specific number of these faces by pulling the joystick towards you. You will avoid other faces by pushing the joystick away from you.

The photographs of faces will be presented in such a way that you will not be able *to perceive them consciously. For this reason, we will present a word immediately following the presentation of a specific face. This word will indicate whether a face was presented that is to be approached or avoided.*

Please respond by moving the joystick forward when the word 'APPROACH' is presented and respond by moving the joystick backward when the word 'AVOID' is presented.

Before we start the task we will teach you which faces you will approach and which faces you will avoid.

These participants thus received instructions to approach by pulling a joystick towards them (an arm flexion movement) and to avoid by pushing the joystick away from them (an arm extension movement). The other half of the participants received identical instructions except that they were instructed to approach by moving themselves towards the screen with the joystick (an arm extension movement) and to avoid by moving themselves away from the screen with the joystick (an arm flexion movement).[4]

Subsequently, participants were shown the four faces that would be presented during the AA training task. Above two faces approach instructions were presented: "These are the faces you will have to approach" ('approach instruction faces') and avoid instructions were presented above the other two faces: "These are the faces you will have to avoid" ('avoid instruction faces'). Participants were asked to make sure that they would not forget which action belonged with each face.

The AA training task consisted of two blocks of 160 trials. During each training block, each of the four faces was presented on 40 occasions and was always presented with either the word 'approach' or the word 'avoid'. For each

[4] Some theories have argued that AA effects depend on whether the AA action consists of arm flexion or arm extension (e.g., Cacioppo, Priester, & Berntson, 1993), or on the specific instructions that are used to frame these actions as approach or avoidance (e.g., Laham et al., 2014). Because Kawakami et al. (2007) and Jones et al. (2013) used different action framing instructions (and incongruent mappings of movement to AA action label), we decided to include both action framing instructions in our experiment.

participant, one of the approach instruction and one of the avoid instruction faces was always paired with approach ('approach training faces') and the other faces were always paired with avoid ('avoid training faces'). To avoid biases, we randomized the assignment of faces to the AA training action and AA instruction action. Similar to Experiment 1, on each trial of the AA training task a moonscape mask was presented for 300 ms, followed by a photograph presented for 23 ms and a backward moonscape mask presented for 33 ms. Then the word approach or avoid was presented until participants responded correctly with the joystick by performing a vertical movement towards the screen or away from the screen. After 200 ms the next trial started.

For half the participants, the AA training task was immediately followed by an evaluative priming task. The other half of the participants first completed explicit ratings. In the evaluative priming task, participants categorized target words as either positive or negative using the E and I keys of a computer keyboard. The assignment of the response keys to either the positive or negative category was counterbalanced across participants and across task order and action framing conditions. Participants were instructed to perform this categorization task as quickly as possible, while making as few mistakes as possible. Participants were further told that they would see photographs of faces presented before the words and that they could look at these photographs, but that their task was simply to respond to the positive and negative words. In line with standard procedures at our lab (Spruyt, De Houwer, Hermans, & Eelen, 2007), a single trial consisted of a fixation cross presented for 500 ms, a blank screen for 500 ms, a prime for 200 ms, a post-prime pause for 50 ms and the target word in white font for 1500 ms or until the participant had given a response. Error feedback was presented on the screen (i.e., the Dutch word for 'wrong' presented in red font) for 250 ms if participants made an error. The inter-trial interval was set to vary randomly between 500 ms and 1500 ms. Participants completed 224 trials separated into two blocks, each containing 14 trials with each of the faces as prime and a positive or negative word as target, presented in random order.

The explicit ratings consisted of two questions for each of the faces. Participants indicated whether they liked the person in the photograph and whether they thought the person in the photograph was friendly on two eight-point Likert scales (0 = not liked at all/not friendly at all; 7 = liked a lot/very friendly). For each face, we collapsed these score ratings into one explicit rating score by averaging the respective ratings. The internal consistency of this score was good (mean Cronbach's $\alpha = 0.76$, $SD = 0.04$).

After the implicit and explicit evaluation tasks, participants completed questions assessing their memory for the instructed stimulus-action contingencies. Each of the faces was presented in a random order and participants were asked to indicate what action they had performed in response to this face according to the instructions. Participants answered by selecting an option on a dropdown menu with "Approach", "Avoid", and "I don't remember" as possible answers.

Subsequently, participants indicated whether they had ever seen any face during the joystick task by choosing from three options (i.e., "Yes, clearly", "Yes, but I couldn't identify which face", and "No, never"). Then each of the faces was presented again and participants reported on how many occasions they had seen this face during the subliminal training task by entering a number between 0 and 320.

Finally, participants performed a perceptibility task which consisted of 80 trials. Trials were identical to the subliminal AA training trials with the exception that (1) the words approach and avoid were replaced with strings of 'XXXX', and (2) participants indicated which of the four faces was presented on each trial by pushing one of four keys on the numeric keypad (1-4).

Results

Awareness and Perceptibility

On average, participants selected the correct instructed action for the faces 82% of the time (SD = 32%). The number of correct instructed action identifications was identical for faces with compatible instruction and training (e.g., approach instructions and approach training; M = 82%, *SD* = 34%) and faces with incompatible instruction and training (e.g., approach instructions and avoid training; M = 82%, *SD* = 34%).

In response to the question whether they had ever seen any face during the joystick task, one participant (1%) indicated ‘yes, clearly’, eleven participants (15%) indicated ‘yes, but I couldn’t identify which face’, and 64 participants (84%) indicated they had not seen any of the faces.

Participants’ overall detection accuracy in the perceptibility task was 26.17% (*SD* = 4.85%), which was significantly above the chance level of 25%, $t(75) = 2.10$, $p = .039$. Individual accuracy scores of 27 participants (36%) were above the 95% confidence interval (> 27.28%).

Implicit Evaluation

For the analysis of the evaluative priming task reaction time data, trials with an incorrect response were dropped (4.6%) as well as trials in which reaction times (RTs) were at least 2.5 standard deviations removed from an individual’s mean (2.9%)[5] (Spruyt, De Houwer, Hermans, & Eelen, 2007). RTs were subjected to a mixed analysis of variance (ANOVA) that contained 3 within-subject factors: Prime AA Instructions (approach, avoidance), Prime AA Training (approach, avoidance), Target Valence (positive, negative), and 2 between-subject factors: Order (evaluative priming task first, explicit rating task first) and Action Framing (approach by pulling, approach by moving towards the screen). We observed a main effect of Target Valence, $F(1,72) = 56.64$, $p < .001$, $\eta^2_p = 0.44$, indicating that participants were faster to detect a positive target ($M = 569$, $SD = 82$) than to detect a negative target ($M = 590$, $SD = 85$). We also observed a main effect of Action Framing, $F(1,72) = 6.14$, $p = .016$, $\eta^2_p = 0.08$, indicating that, overall, participants were faster if they had approached faces by pulling the joystick in the subliminal AA training task. More importantly, we observed an

---

[5] Including these trials did not result in any shift in significance for any of the reported effects.

interaction effect of Target Valence and Prime AA Instructions, $F(1,72) = 5.68$, $p = .020$, $\eta^2_p = 0.07$ (Table 2). Participants were faster when the valence of the target word was compatible with the valence of the AA instruction for the face prime (i.e., positive target and approach instruction face [*M* = 565, *SD* = 88], or negative target and avoid instruction face [*M* = 588, *SD* = 83]), than when they were incompatible (i.e., positive target and avoid instruction face [*M* = 572, *SD* = 77], or negative target and approach instruction face [*M* = 593, *SD* = 87]). In contrast, the interaction of Target Valence and Prime AA Training was not significant, $F(1,72) = 0.03$, $p = .87$, $\eta^2_p < 0.01$, nor were any other main or interaction effects, *F*s < 1.44., *p*s > .23.

Table 2.

Mean RTs in the evaluative priming task in Experiment 2 as a function of Prime AA Instructions, Prime AA Training, and Target Valence.

	Approach Instructions		Avoid Instructions	
	Approach training	Avoidance training	Approach Training	Avoidance training
Positive Target	565 (83)	565 (93)	571 (77)	572 (78)
Negative Target	593 (84)	593 (90)	586 (83)	589 (84)

Note. Standard deviations are in parentheses.

To further examine participants' implicit preference for approach training and approach instruction faces, we calculated two indices of AA effects. An index of the AA training effect was calculated by subtracting evaluative priming RTs for trials where target valence was compatible with the valence of the AA training for the face prime (i.e., trials with approach training face and positive target or avoid training face and negative target) from RTs for trials where target valence was incompatible with the valence of the AA training for the face prime (i.e., trials with avoid training face and positive target or approach training face and negative target). An index of the AA instruction effect was calculated in the same manner for trials where target valence and valence of the AA instructions for the face prime were compatible or incompatible. One-sample t-tests indicated a significant effect of AA instructions (*M* = 23, *SD* = 83), $t(75) = 2.40$, $p = .019$, $d = 0.28$, $BF_1 = 4.58$, but no effect of AA training (*M* = -2, *SD* = 94), $t(75) = -0.17$, $p =$

.87, d = 0.02, BF_0 = 6.49.

We performed a linear regression of participants' detection scores in the perceptibility task on the index of the AA training effect. We did not observe a significant positive intercept, b_1 = -23.27, $t(74)$ = -0.39, p = .70, indicating that participants whose detection performance was at chance level did not exhibit a significant AA training effect. The slope was also not significant, b_2 = 81.98, $t(74)$ = 0.37, p = .72, which indicates that the AA training effect was not a function of stimulus visibility.

Explicit evaluation

The explicit rating scores were subjected to a mixed ANOVA that included AA Instructions and AA Training as within-subjects factors, and Order and Action Framing as between-subjects factors. The analysis revealed a main effect of AA Instructions, $F(1,72)$ = 16.19, p < .001, η^2_p = 0.19 (Table 3). Participants preferred faces they were instructed to approach (M = 3.88, SD = 1.50) over faces they were instructed to avoid (M = 2.93, SD = 1.44). The main effect of Action Framing was also significant, indicating that participants who approached by pulling and avoided by pushing evaluated the faces more positively compared to participants where the framing was reversed, $F(1,72)$ = 7.18, p = .009, η^2_p = 0.09. No other effects reached significance, Fs < 1.94, ps >.16.

Table 3.

Mean explicit rating scores in Experiment 2 as a function of AA Instructions and AA Training.

	Approach Instructions		Avoid Instructions	
	Approach training	Avoidance training	Approach Training	Avoidance training
Explicit Rating score	3.91 (1.62)	3.86 (1.38)	2.87 (1.41)	2.99 (1.43)

Note. Standard deviations are in parentheses.

We calculated indices of the AA training and AA instruction effect by subtracting the explicit rating scores for avoid training faces from explicit rating scores for approach training faces and explicit rating scores for avoid instruction faces from explicit rating scores for approach instruction faces. These indices correlated significantly with the corresponding indices of AA effects on implicit evaluations, *r*s > 0.39, *p*s <.001. Similar to the results for implicit evaluations, we observed a significant effect of AA instructions (*M* = 1.90, *SD* = 4.06), *t*(74) = 4.09, $p < .001$, *d* = 0.47, BF_1 = 412.97, but no significant effect of AA training (*M* = -0.07, *SD* = 2.26), *t*(74) = -0.28, *p* = .78, *d* = 0.03, BF_0 = 7.00.

Discussion

In Experiment 2, we obtained no evidence that subliminal training to approach or avoid face stimuli causes changes in implicit or explicit evaluations of these stimuli. In contrast, we observed changes in both implicit and explicit evaluations as the result of instructions about the stimulus-action contingencies. This corroborates previous findings of AA instruction effects (Van Dessel, De Houwer, Gast, & Smith, 2015), and extends these findings by showing that AA instruction effects occur even if (subliminal) AA training is provided that is incompatible with the provided information. These findings are more consistent with a propositional account of AA training effects than with associative accounts which suggest that AA training effects result from the gradual and automatic formation of associations.

Some aspects of our procedure, however, may have impeded the detection of subliminal AA training effects. First, we may have lacked the power to detect a small effect (power = .53 to detect an effect-size of *d* = 0.20). However, Bayesian analyses indicated that our data provide substantial evidence that subliminal AA training does not cause changes in evaluations. In contrast, these analyses indicated strong evidence in favor of AA instruction effects. Second, the mere presence of AA instructions might have somehow interfered with the subliminal training effect. For instance, subliminal AA training may not work in the presence of opposite AA instructions because (1) these instructions allow the stimulus to acquire a specific valence (Van Dessel, De Houwer, Gast, & Smith, 2015), and (2) recently established (implicit) evaluations may not be easily undone (Gregg et al., 2006). Third, we used Kawakami et al. (2007)'s procedure for the subliminal presentation of the face stimuli. Though this procedure may allow participants to identify the race of the depicted person, it may be more difficult to register the identity of the face stimulus. Arguing against this explanation, however, research with event related brain potentials (ERPs) suggests that participants can process the identity of masked face stimuli that are presented for these short durations (e.g., Lee, Lim, Lee, & Choi, 2009). Moreover, performance in the detection task indicated that participants were able to discriminate the face stimuli better than chance level.

EXPERIMENT 3

In Experiment 3, we changed our set-up to create more optimal circumstances for subliminal AA training effects to occur. First, novel non-words were used as stimuli. Previous research suggests that AA effects on evaluations are weaker for associatively rich stimuli than for associatively-impoverished stimuli such as non-words (e.g., Priester et al., 1996). Second, we tried to maximize the possibility that participants would be able to register the identity of the stimuli during the subliminal training task by modeling the subliminal presentation procedure after studies that established subliminal priming effects on the basis of the identity of non-words (Holcomb & Grainger, 2006; Beyersmann, Castles, & Coltheart, 2009). Third, we tested a larger sample of participants (N = 96) to have sufficient statistical power (power = 0.80) to detect even a small effect size. Finally, we included a manipulation that might facilitate association formation under conditions of subliminal stimulus presentations (Custers & Aarts, 2011).

Custers and Aarts (2011) proposed that association formation strongly depends on attentional processes. When people experience specific contingencies in the environment this may allow for the formation of associations in memory if their attentional system is prepared to process these contingencies. Under these circumstances, associative learning effects may arise even in the absence of awareness of the contingencies or in the absence of stimulus awareness (see also Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006; Dijksterhuis & Aarts, 2010). They further argued that the existence of a contingency relationship in the environment may tune participants' attention to detect other such relationships. Thus, when participants experience specific contingencies before a learning phase with subliminal stimulus presentations, this may facilitate effects of associative learning. They tested this in three experiments and obtained evidence that the learning of predictive relations in a subliminal priming task was facilitated when participants engaged in predicting targets on the basis of cues in an earlier, unrelated task.

If attention can be tuned to process contingency relations and if this facilitates association formation, then participants who experienced specific contingencies in an earlier AA training task may show stronger effects of subliminal AA training. To test this, we designed an experiment where participants performed an AA training task with *supraliminal* stimulus presentations prior to performing the subliminal AA training task. For half the participants the supraliminal AA training task involved specific stimulus-action contingencies (attention-tuning condition), for the other participants the identity of the stimulus did not predict the action: 50% contingency (control condition).

Method

Participants

Ninety-six native Dutch-speaking undergraduates (78 women) participated in exchange for a monetary reward of 5 euros. None of the participants had previously participated in Experiments 1 or 2. Participants were randomly assigned to the attention-tuning condition or to the control condition.

Apparatus and Materials

Four non-words were selected as evaluation stimuli from previous studies that registered Dutch-speaking participants' evaluations of non-words (Zanon, De Houwer, & Gast, 2012). The non-words for the supraliminal AA training task were UDIBNON and SARICIK. For the subliminal AA training task the non-words were LOKANTA and FEVKANI. The attribute stimuli used in the IAT were four positive words (the Dutch words for happy, pleasant, fun, and nice) and four negative words (the Dutch words for mean, unpleasant, irritating, and bad).

Procedure

Upon entering, participants were seated at a desk and were informed that they would perform a series of tasks. They were asked to remember that some of these tasks involved reacting to two non-existing words, specifically LOKANTA and FEVKANI (the non-words for the subliminal training task). Participants were then instructed that in the first task they would respond to the words “approach” and “avoid” by making approach or avoidance movements with the joystick. The same action framing instructions were used as in Experiment 2.

The supraliminal AA training task consisted of 40 trials. The non-words SARICIK and UDIBNON were each presented on 20 occasions. For participants in the attention-tuning condition, one of the non-words was always presented with the word approach and the other non-word was always paired with the word avoid. Assignment of the non-words SARICIK and UDIBNON to the approach or avoidance action was counterbalanced across participants. For participants in the control condition each non-word was equally often followed by the words approach and avoid. Each trial started with the presentation of a white fixation cross presented in the center of the screen for 500 ms, followed by a forward mask of # symbols presented for 500 ms. Then, one of the non-words was displayed for 500 ms after which the word approach or avoid was presented until participants responded correctly with the joystick by performing a vertical movement towards the screen or away from the screen. The inter-trial interval was 200 ms.

Subsequently, participants were informed that they would now perform another joystick task. Trials in the subliminal training task were identical to the supraliminal training task trials with the following exceptions. First, there were 2 blocks of 50 trials and in each of the blocks the non-word FEVKANI and the non-word LOKANTA were each presented on 25 occasions. Second, for participants in both conditions one of the non-words was always presented with the word approach and the other non-word was always presented with the word avoid. Assignment of the non-words FEVKANI and LOKANTA to the approach or avoidance action was counterbalanced across participants and across conditions. Third, the presentation time of the non-words was limited to 40 ms.

In the IAT, participants categorized attribute words as 'positive' or 'negative' and target words FEVKANI and LOKANTA as 'Fevkani' or 'Lokanta' by using the E and I keys of a computer keyboard. To avoid that stimuli were classified only on the basis of simple perceptual features, each target stimulus was presented in four different fonts (lower case Arial Black, upper case Arial Black, lower case Fixedsys, and upper case Fixedsys), resulting in 8 different target stimuli. All other procedural details of the IAT were identical to Experiment 1.

Participants then completed liking ratings and thermometer ratings of self-reported warmth feelings towards each of the four non-words on two 8-point Likert scales (0 = not liked/warm at all; 7 = completely liked/warm). For each non-word, we collapsed these ratings into one explicit rating score by averaging the respective ratings. The internal consistency of this measure was good (mean Cronbach's $\alpha = 0.76$, $SD = 0.04$).

Next, participants answered questions about the supraliminal AA training task: They indicated (1) whether they had noticed any regularities in the presentations of the words approach or avoid and the non-words UDIBNON and SARICIK, and (2) what action word had been presented most often with these non-words by choosing from four options (i.e., 'approach', 'avoid', 'both an equal number of times' or 'I don't remember'). For the subliminal training task, participants reported whether they had ever seen a stimulus appear before the words approach or avoid (yes/no). Then, participants were informed that the non-words FEVKANI and LOKANTA had been presented during this task and they indicated whether they had ever seen these words and what action word they thought had been presented most often with these non-words.

Finally, participants performed a d' perceptibility task which consisted of 50 trials. Trials were identical to trials in the subliminal AA training task with the exception that (1) the words approach and avoid were replaced with strings of 'XXXX', and (2) participants indicated whether FEVKANI or LOKANTA had been presented by pushing key 1 or 2 on the numeric keypad.

Results

Awareness and Perceptibility

First, we investigated participants' awareness of the contingencies in the supraliminal AA training task. As expected, participants in the attention-tuning condition indicated more often that they had noticed regularities in the presentation of action words and non-words (52% of the time) than participants in the control condition (25% of the time), $t(94) = 2.81$, $p = .006$. Participants in the attention-tuning condition selected the correct action for the non-words (M = 44%, SD = 48%) more often than the incorrect action (M = 5%, SD = 21%), $t(47) = 4.79$, $p < .001$. Participants indicated that they did not know the correct action for the nonwords 43% of the time (SD = 48%) and that they had performed both actions an equal number of times 8% of the time (SD = 24%).

For the subliminal AA training task, 20 participants (21%) indicated that they had noticed that a stimulus was sometimes presented before the approach/avoid word and16 participants (17%) indicated that the non-words FEVKANI and LOKANTA were presented. Participants did not select the correct action for the non-words (M = 6%, SD = 23%) significantly more often than the incorrect action (M = 5%, SD = 21%), $t(95) = 0.32$, $p = .75$.

Participants' overall accuracy in the perceptibility task was 65% (SD = 11%). The signal detection sensitivity measure d' was computed with hits defined as LOKANTA responses on trials where LOKANTA was presented, and false alarms as LOKANTA responses on trials where FEVKANI was presented. The d' score indicated that participants' detection performance was above chance level (M = 0.86, SD = 0.71), $t(95) = 11.81$, $p < .001$.

Supraliminal AA training

We examined whether supraliminal AA training caused changes in explicit rating scores for contingency aware and contingency unaware non-words. Non-words were classified as contingency aware if participants had indicated the correct action for the non-word (44%) and as contingency unaware if they had indicated the incorrect action or if they indicated that they didn't know the correct action or that both actions had been performed an equal number of times (56%). Analyses were restricted to the data of participants in the attention-tuning condition because only these participants experienced specific stimulus-action contingencies. In line with Van Dessel, De Houwer, and Gast (2015), analyses were performed with item-based linear mixed effects models as implemented in R package lme-4 (Bates, Maechler, Bolker, & Walker, 2014). We tested for an effect of supraliminal AA training by fitting a model that included Participant as a random factor and the fixed factors of Action (approach, avoid), Non-word (Saricik, Udibnon), and Contingency Awareness (contingency aware, contingency unaware). We observed a main effect of Action, $\chi^2(1) = 3.84$, $p = .050$, which was qualified by a marginally significant interaction effect of Action and Contingency Awareness, $\chi^2(1) = 2.88$, $p = .090$. A preference for approached non-words ($M = 3.48$, $SD = 1.45$) over avoided non-words ($M = 2.50$, $SD = 1.11$) was observed for contingency aware non-words, $\chi^2(1) = 5.80$, $p = .016$, $d = 0.67$, $BF_1 = 5.81$, but not for contingency unaware non-words, $\chi^2(1) = 0.03$, $p = .87$, $d = 0.07$, $BF_0 = 3.01$.

Subliminal AA Training

IAT D4 scores were calculated such that higher scores indicate a preference for the approached non-word over the avoided non-word. Importantly, we did not observe a significant preference for the approached non-word ($M = -0.03$, $SD = 0.51$), $t(95) = -0.52$, $p = .60$, $BF_0 = 9.14$. An ANOVA on IAT score with Non-word (Lokanta, Fevkani), Condition (attention-tuning, control), and Action Framing (approach by pulling, approach by moving towards the screen) as between-subjects factors only revealed a main effect of Non-word, $F(1,88) = 15.68$, $p < .001$, $BF_1 = 209.41$, indicating that participants preferred LOKANTA over FEVKANI. The main effect of Condition was not significant, $F(1,88) = 0.41$, $p = .52$, $BF_0 = 3.95$. We did not observe a significant preference for the approached non-word for participants in the attention-tuning or control condition, ts < 0.07, ps > .95, BF_0s > 4.44. No other main or interaction effects were significant, Fs < 0.98, ps >

.32. We performed a linear regression of d' scores on IAT scores. The intercept, b_1 = -0.06, $t(94)$ = -0.73, p = .47, and slope were not significant, b_2 = 0.04, $t(94)$ = 0.51, p = .61.

An explicit preference score for the approached non-word was calculated by subtracting participants' explicit rating score for the avoided non-word, from the explicit rating score for the approached non-word (M = -0.39, SD = 4.16). This score correlated significantly with the IAT score, $r(94)$ = 0.59, p < .001. Importantly, we did not observe a significant preference for the approached non-word (M = 6.94, SD = 2.45) over the avoided non-word (M = 7.32, SD = 2.43), $t(95)$ = -0.91, p =.37, BF_0 = 11.38. An ANOVA on the explicit rating score revealed only the main effect of Non-Word, $F(1,88)$ = 28.00, p < .001, BF_1 = 11261.91, but no main effect of Condition, $F(1,88)$ = 0.18, p = .67, BF_0 = 4.39, or any interaction effect, Fs < 1.35, ps > .24. We did not observe a preference for the approached non-word in the attention-tuning or control condition, ts < 0.01, ps > .99, BF_0s > 6.03.

Discussion

Despite more optimal conditions for subliminal perception and an increased power for detecting small effects we did not obtain evidence for the conclusion that subliminal AA training caused changes in evaluations of novel non-words. We also did not find more robust subliminal AA training effects when participants experienced specific stimulus-action contingencies in an earlier supraliminal AA training task. This might indicate that the attention-tuning manipulation did not enhance participants attention to contingency relations (but note that participants in the attention-tuning condition did indicate awareness of the contingencies in the supraliminal AA training task) or that tuning participants attention to process these relations does not facilitate associative learning (but see Custers & Aarts, 2011). Alternatively, the findings by Custers and Aarts (2011) may simply not extend to AA training effects because awareness of the specific contingencies is crucial for AA training effects rather than attention to contingency relations in general. In line with this idea, we observed effects of *supraliminal* AA training only when participants indicated awareness of the stimulus-action contingencies, corroborating previous findings (Van Dessel, De Houwer, & Gast, 2015).

GENERAL DISCUSSION

In three experiments with different stimuli and procedures, we tested whether subliminal AA training changes stimulus evaluations. Participants repeatedly performed AA movements in response to approach or avoidance cues that were preceded by subliminally presented stimuli: faces of Black or White people (Experiment 1), unfamiliar faces (Experiment 2), or unfamiliar non-words (Experiment 3). In contrast to previous findings (Kawakami et al., 2007; Jones et al., 2013), we did not observe changes in implicit or explicit evaluations of these stimuli that could be attributed to the subliminal AA training. We did, however, observe AA effects when participants were provided with correct or incorrect information about the stimulus-action contingencies in the subliminal AA training task and when participants performed *supraliminal* AA training and were aware of the stimulus-action contingencies.

How can these results be reconciled with previous findings?

Kawakami et al. (2007) and Jones et al. (2013) reported effects of subliminal AA training in two and three experiments, respectively; we report three experiments that do not show these effects. To explain these discrepant results two options can be considered. On the one hand, it is possible that AA training cannot cause changes in stimulus evaluations when the stimuli are presented subliminally during the training phase. Under that assumption, we might look at the original findings and suggest that the reported effects are simply Type I error, that is, a false rejection of the null hypothesis. In line with this idea, Bayesian analyses indicated that these original results provided only anecdotal evidence for subliminal AA training effects while our results consistently provided more substantial evidence favoring the idea that subliminal

AA training does not allow for changes in evaluations. A Bayes Factor meta-analysis including the data of the successful subliminal AA training studies by Kawakami et al. and Jones et al., as well as our studies indicates that the available evidence favors the null hypothesis that subliminal AA training does not influence implicit evaluations, $BF_0 = 3.64$, or explicit evaluations, $BF_0 = 6.51$ (Table 1). However, some caution is warranted when interpreting these results, most importantly because the subliminal training studies all differed substantially in their methodology. Alternatively, it is possible that the effects reported by Kawakami et al. and Jones et al. occurred because stimuli were on some occasions presented above the perceptual threshold. By using d' measures of stimulus perceptibility, we found evidence that, with the subliminal AA training procedure used by Kawakami et al., some participants' are able to detect the presented stimuli. Although we did not observe that this caused changes in evaluations in our experiments, this may have contributed to the original reports of subliminal AA training. One could object that in three of the experiments reported by Kawakami et al. and Jones et al. effects were observed even though participants did not report awareness of the stimuli when probed. As previously contended, however, the accuracy of these reports may be limited because (1) awareness was assessed only after the conditioning and evaluation phase, and (2) participants do not always report having been aware of a stimulus despite having some subjective experience of it (Cleeremans, 2001).

On the other hand, it is possible that subliminal AA training effects can be reliably observed, but we failed to do so. The detection of these effects, then, may require certain methodological idiosyncrasies that we failed to incorporate in our experiments. First, for some reason, our procedures might have interfered with the perception of the subliminal stimuli (Bargh & Morsella, 2008). It is unlikely that this was a problem in our experiments. Not only did we model our subliminal presentation procedures after Kawakami et al.'s study and other studies that report robust subliminal priming effects, also did the d' perceptibility scores indicate that some participants were even able to consciously identify the identity of the presented stimuli. Note, however, that d' perceptibility scores may overestimate awareness to the stimuli (Vermeiren & Cleeremans, 2012; but see Amihai, 2012). For instance, directing participants' attention to the subliminal stimuli during the d' perceptibility task may lead to higher visibility of these stimuli. Second, AA training may cause changes in evaluations only when the training involves a sufficiently large number of training trials (e.g., because associative learning is a slow and gradual process; Rydell & McConnell, 2006). Again we see no reason why this would be a problem in our studies. We matched the number of training trials to that in Kawakami et al. (Experiment 1) or to the number of training trials recommended by Woud, Becker, and Rinck (2011) for AA training with novel stimuli. Third, subliminal AA training effects may be limited to well-known stimuli such as spiders and well-known social groups. For instance, visual attention may be more strongly directed toward a well-known stimulus because the stimulus' related evaluation is particularly accessible (Roskos-Ewoldsen and Fazio, 1992; Young & Fazio, 2013). However, in both our experiments with novel stimuli, we presented the stimuli in the instructions directly preceding the subliminal training task to ensure that a representation of the stimulus would recently have been made accessible. Moreover, participants were asked to remember these stimuli (or stimulus-action contingencies), which may even facilitate selective visual attention (Soto, Humphreys, & Heinke, 2006; Soto, Heinke, Humphreys, & Blanco, 2005) and thus learning (e.g., Mackintosh, 1975).

In sum, although we realized different variations of an AA training task that are representative of AA training tasks in the literature, we did not replicate or establish effects on subliminally presented stimuli in any of these variations. Of course such a set of results can never exclude the possibility that an effect on subliminally presented stimuli might occur under a different set of conditions. It does, however, indicate that these effects, if they exist, depend on boundary conditions that are not yet identified. Further research is necessary to determine if and under what circumstances subliminal AA training effects can be reliably established. In the absence of this research, one should be cautious in drawing strong theoretical conclusions on the basis of previous reports of subliminal AA training effects. This also means that evidence for subliminal AA training is currently not reliable enough to be treated as conclusive evidence in the debate between association formation versus propositional models of associative learning in general and AA training specifically.

Contingency awareness and AA training

The current results fit nicely with the idea that awareness of the stimulus-action contingencies mediates AA training effects (Van Dessel, De Houwer, & Gast, 2015). If contingency awareness is necessary to obtain AA training effects, then subliminal AA training effects should not occur. Of course, the reversed inference does not hold: The fact that we did not observe AA training effects in the absence of awareness of the stimuli does not prove that AA training cannot cause changes in evaluations in the absence of contingency awareness. However, some aspects of our data do support a causal role of contingency information in AA training effects. First, in Experiment 2, we obtained evidence that aware contingency knowledge can lead to AA effects without actual practice of the AA contingencies. When participants were informed about the stimulus-action contingencies that would be presented in the subliminal AA training task, they exhibited changes in implicit and explicit evaluations in line with this information (also see Van Dessel, De Houwer, Gast, & Smith, 2015). This effect occurred even when the contingency information contrasted with the contingencies that were actually presented in the training task. Second, the results of Experiment 3 provide evidence that contingency awareness is also a necessary condition for AA training effects. When participants were provided with supraliminal AA training, a preference for approached non-word stimuli was observed only when participants were able to consciously report the relation between non-word and action. This corroborates previous findings that AA training influences evaluations (of novel faces) only in the presence of contingency awareness (Van Dessel, De Houwer, & Gast).

Theoretical implications

The current results have important implications for accounts of AA training effects. Most importantly, they pose a challenge to theories that assume that the repeated paring of stimuli and AA actions allows for the automatic formation of associations that underlie (implicit) evaluations (Woud et al., 2013; Phills et al., 2011). For our findings to be reconciled with these accounts it seems necessary to assume that specific boundary conditions determine when these automatic effects may arise. Current association formation theories, however, remain unclear on the necessary environmental conditions (Shanks, 2007). In Experiment 3, we explored one possible boundary condition of unconscious associative learning (i.e., whether attention is tuned to process contingency relations; see Custers & Aarts, 2011). However, we failed to find evidence that our attention-tuning manipulation moderated subliminal training effects. In the absence of a clear description of boundary conditions for association formation, the current results may be more easily explained by an alternative, propositional account of AA training effects. Because a propositional account suggests that the acquisition of information about stimulus-action contingencies directly causes changes in implicit and explicit stimulus evaluations, this fits well with the observation that contingency awareness plays an important role in establishing AA training effects.

Our findings may also advance the debate about the validity of dual-process and single-process propositional models of evaluative learning. Dual-process models assume two routes of evaluative learning: learning on the basis of the automatic formation of associations and learning on the basis of conscious propositional reasoning (Smith & DeCoster, 2000; Strack & Deutsch, 2004). In contrast, single-process propositional theories postulate that all learning necessarily involves the conscious acquisition of propositional information (e.g., De Houwer, 2009; Mitchell, De Houwer, & Lovibond, 2009). Proponents of these theories argue that unambiguous evidence in favor of unconscious associative learning is scarce and the existence of an automatic association formation mechanism can therefore be questioned (see Mitchell et al., 2009, for a review). In the evaluative conditioning (EC) literature, for example, investigations which used more fine-grained methodologies often failed to find evidence for subliminal EC and for EC without contingency awareness (Pleyers, Corneille, Luminet, & Yzerbyt, 2007; Gast, De Houwer, & De Schryver, 2012; but see Hütter, Sweldens, Stahl, Unkelbach, & Klauer, 2012; see Sweldens, Corneille, & Yzerbyt, 2014, for a recent review).

In the area of AA training effects, Kawakami et al. (2007) and Jones et al. (2013) report evidence for subliminal effects. As we noted in the introduction, these findings have the potential to provide strong support for the hypothesis that preferences can be learned unconsciously, and hence support the idea that an automatic association formation mechanism exists. This is not only relevant for understanding AA training effects, but also for the more general debate about associative, propositional, and dual-process models. Considering the relevance of such findings the reports of unconscious AA training effects, however, have not been scrutinized sufficiently, unlike reports of unaware learning in other research areas. In the current paper we both evaluate the previous evidence for subliminal AA training effects and report three additional studies that test it. Across our analyses we consistently failed to find reliable evidence for the existence of subliminal AA training effects. We nevertheless hope that future studies will further try to establish the empirical validity of findings that, if valid, may provide strong evidence for the existence of an association formation mechanism.

Our results also have important implications for theories of evaluation. Evaluation is often explained in terms of dual-process models in which implicit and explicit evaluation depend on different systems or processes (Gawronski & Bodenhausen, 2006, 2011, 2014; Rydell & McConnell, 2006; Rydell, McConnell, Mackie, & Strain, 2006; Rydell, McConnell, Strain, Claypool, & Hugenberg, 2007). Implicit evaluations are typically thought to emerge from associative mental processes that operate through the spreading activation of associations in memory. Because associations are assumed to be formed gradually and automatically, implicit evaluations may not be easily changed (Gregg et al., 2006). Explicit evaluations are typically assumed to be governed by propositional processes that operate on the basis of propositional reasoning, and therefore may more strongly and immediately reflect new information. In line with this idea, Rydell and McConnell (2006) showed that verbal information about a novel stimulus caused changes in the explicit, but not implicit evaluation of this stimulus, whereas the repeated pairing of the same stimulus with valenced primes, presented subliminally, changed only implicit evaluations. In Experiment 2, we provided evidence that directly contrasts these results. Both implicit and explicit evaluations changed as the result of verbal information about the contingencies between a stimulus and AA action, but not as the result of the repeated pairing of the stimulus with AA actions. This accords with recent findings that both implicit evaluations and explicit evaluations can be changed even by a single piece of verbal information (Cone & Ferguson, 2015; De Houwer, 2006; Gast & De Houwer, 2012; Peters & Gawronski, 2011; Van Dessel et al., 2015; Zanon et al., 2014) and fits more easily with theories which assume that propositional information influences both implicit and explicit evaluations (De Houwer, 2014). Again, further research is required to distinguish between the various accounts. A first step in this endeavour can be to try and replicate any studies reporting evidence that clearly favors the idea that implicit and explicit evaluative change relies on separate processes.

Concluding Remarks

On the basis of the available evidence, we believe that there is insufficient evidence to conclude that AA training causes changes in evaluations in the absence of awareness. If subliminal AA training effects do exist, they must be subjected to very stringent boundary conditions. In contrast, evidence seems to be accumulating that conscious propositional knowledge about stimulus-action contingencies plays an important role in AA training effects. Nevertheless, given the important theoretical implications of subliminal AA training effects, we hope that future studies will further explore these effects and attempt to replicate studies reporting these effects.

REFERENCES

Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *Handbook of social psychology*. Worcester, Mass: Clark University Press.

Amihai I. (2012). Problems in using d' measures to assess subjective awareness. *Iperception, 3*, 783–785. doi: 10.1068/i0563ic

Bargh, J. A., & Morsella, E. (2008*).* The unconscious mind*. Perspectives on Psychological Science,* 3, 73–79. doi:10.1111/j.1745-6916.2008.00064.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-6. http://CRAN.R-project.org/package=lme4

Beyersmann, E., Castles, A., & Coltheart, M. (2011). Early morphological decomposition during visual word recognition: Evidence from masked transposed-letter priming. *Psychonomic Bulletin and Review*, *18*, 937-942. doi: 10.1037/0022-3514.49.4.937

Cacioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary determinants of attitudes. II: Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology, 65*, 5-17. doi: 10.1037//0022-3514.65.1.5

Castelli, L., & Tomelleri, S. (2008). Contextual effects on prejudiced attitudes: When the presence of others leads to more egalitarian responses. *Journal of Experimental Social Psychology, 44,* 679-686. doi: 10.1016/j.jesp.2007.04.006

Cleeremans, A. (2001). Conscious and unconscious processes in cognition. In N. Smelser & P. B. Baltes (Eds.), *International encyclopedia of social and behavioral sciences,* Vol. 4 (pp. 2584–2589). London: Elsevier.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cone, J., & Ferguson, M. J. (2015). He Did What? The Role of Diagnosticity in Revising Implicit Evaluations. *Journal of Personality and Social Psychology, 108*, 37-57. doi: 10.1037/pspa0000014

Cornelis, I., & Van Hiel, A. (2006). The impact of cognitive styles on authoritarianism based conservatism and racism. *Basic and Applied Social Psychology, 28*, 37-50. doi: 10.1207/s15324834basp2801_4

Custers, R., & Aarts, H. (2011). Learning of predictive relations between events depends on attention, not on awareness. *Consciousness and Cognition, 20*, 368–378. doi: 10.1080/10640266.2012.715515

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

De Houwer, J. (2014). A Propositional Model of Implicit Evaluation. *Social and Personality Psychology Compass, 8*, 342-353. doi: 10.1111/spc3.12111

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences, 10*, 204–211. doi: 10.1016/j.tics.2006.03.007

Dijksterhuis, A., & Aarts, H. (2010). Goals, attention, and (un)consciousness. *Annual Review of Psychology, 61*, 467–490. doi: 10.1146/annurev.psych.093008.100445

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229–238. doi: 10.1037//0022-3514.50.2.229

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

Gast, A., De Houwer, J., & De Schryver, M. (2012). Evaluative conditioning can be modulated by memory of the CS-US Pairings at the time of testing. *Learning and Motivation, 43*, 116-126. doi: 10.1016/j.lmot.2012.06.001

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*, 692–731. http://dx.doi.org/10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2011). The associative–propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology, 44*, 59–127. http://dx.doi.org/10.1016/B978-0-12-385522-0.00002-0

Gawronski, B., & Bodenhausen, G. V. (2014). The associative–propositional evaluation model: Operating principles and operating conditions of evaluation. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 188–203). New York, NY: Guilford Press.

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*, 1464–1480. doi: 10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197–216. doi: 10.1037/0022-3514.85.2.197

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,* 1-20. doi: 10.1037/0022-3514.90.1.1

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

Holcomb, P.J., & Grainger, J. (2006). On the time-course of visual word recognition: an ERP investigation using masked repetition priming. *Journal of Cognitive Neuroscience, 18*, 1631–1643. doi: 10.1162/jocn.2006.18.10.1631

Hütter, M., Sweldens, S., Stahl, C., Unkelbach, C., & Klauer, K. C. (2012). Dissociating contingency awareness and conditioned attitudes: Evidence of contingency-unaware evaluative conditioning. *Journal of Experimental Psychology: General, 141*, 539–557. doi: 10.1037/a0026477

Jeffreys, H. (1961). Theory of Probability. Oxford: Oxford University Press.

Jones, C. R., Vilensky, M. R., Vasey, M. W., & Fazio, R. H. (2013). Approach behavior can mitigate predominately univalent negative attitudes: Evidence regarding insects and spiders. *Emotion, 13*, 989-996. doi: 10.1037/a0033164

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder: Improving implicit racial evaluations and interracial interactions through approach behaviors. *Journal of Personality and Social Psychology, 92*, 957-971. doi: 10.1037/0022-3514.92.6.957

Kawakami, K., Steele, J. R., Cifa, C., Phills, C. E., & Dovidio, J. F. (2008). Approaching math increases math = me and math = pleasant. *Journal of Experimental Social Psychology, 44*, 818–825. doi:10.1016/j.jesp.2007.07.009

Laham, S. M., Kashima, Y., Dix, J., Wheeler, M., & Levis, B. (2014). Elaborated contextual framing is necessary for action-based attitude acquisition. *Cognition & Emotion, 28*,1119-1126. doi: 10.1080/02699931.2013.867833

Laham, S. M., Kashima, Y., Dix, J., & Wheeler, M. (in press).A meta-analysis of the facilitation of arm flexion and extension movements as a function of stimulus valence. *Cognition & Emotion.* doi:10.1080/02699931.2014.968096

Lee T. H., Lim S. L., Lee K. Y, Choi J. S. (2009). Facilitation of visual processing by masked presentation of a conditioned facial stimulus. *Neuroreport, 20,* 750–754. doi: 10.1097/wnr.0b013e32832a5cfa

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82,* 276–98. doi: 10.1037/h0076778

Macmillan, N. A., & Creelman, C. D. (2005). Detection Theory: A User's Guide. Mahwah, NJ: Lawrence Erlbaum.

McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination and racism* (pp. 91- 126). New York: Academic.

Merikle, P., Smilek, D., & Eastwood, J. (2001). Perception without awareness: Perspectives from cognitive psychology. *Cognition, 79*, 115–134. doi: 10.1016/S0010-0277(00)00126-8

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

Perlmutter, P. (1999). Legacy of hate: A short history of ethnic, religious, and racial prejudice in America. New York: M.E. Sharp.

Peters, K. R., & Gawronski, B. (2011). Are we puppets on a string? Comparing the impact of contingency and validity on implicit and explicit evaluations. *Personality and Social Psychology Bulletin, 37*, 557–569. doi: 10.1177/0146167211400423

Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in Western Europe. *European Journal of Social Psychology, 25*, 57-75. doi: 10.1002/ejsp.2420250106

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. doi: 10.1037/a0022159

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,* 130–144. doi: 10.1037/0278-7393.33.1.130

Priester, J. R., Cacioppo, J. T., & Petty, R. E. (1996). The influence of motor processes on attitudes toward novel versus familiar semantic stimuli. *Personality and Social Psychology Bulletin, 22*, 442–447. doi:10.1177/0146167296225002

Richeson, J.A., & Ambady, N. (2003). Effects of situational power on automatic racial prejudice. *Journal of Experimental Social Psychology, 39,* 177-183. doi: 10.1027/1614-0001/a000142

Roskos-Ewoldsen, D. R., & Fazio, R. H. (1992). On the orienting value of attitudes: Attitude accessibility as a determinant of an object's attraction of visual attention. *Journal of Personality and Social Psychology, 63*, 198–211. doi: 10.1037/0022-3514.63.2.198

Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin and Review, 18*, 682–689. doi: 10.3758/s13423-011-0088-7

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237. doi: 10.3758/PBR.16.2.225

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*, 995–1008. doi: 10.1037/0022-3514.91.6.995

Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of two minds: Forming and changing valence-inconsistent implicit and explicit attitudes. *Psychological Science, 17*, 954–958. doi: 10.1111/j.1467-9280.2006.01811.x

Rydell, R., McConnell, A. R., Strain, L. M., Claypool, H. M., & Hugenberg, K. (2007). Implicit and explicit attitudes respond differently to increasing amounts of counterattitudinal information. *European Journal of Social Psychology, 37*, 867–878. doi: 10.1002/ejsp.393

Shanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *Quarterly Journal* of *Experimental Psychology, 60*, 291–309. doi: 10.1080/17470210601000581

Smith, E. R.. & DeCoster, J. (2000). Dual process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4,* 108-131. doi: 10.1207/S15327957PSPR0402_01

Snodgrass, M., Bernat, E., & Shevrin, H. (2004). Unconscious perception: A model-based approach to method and evidence. *Perception and Psychophysics, 66*, 846–867. doi: 10.3758/BF03194978

Soto, D., Heinke, D, Humphreys, G. W., & Blanco, MJ. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human perception and Performance, 31*, 248–261. doi: 10.1037/0096-1523.31.2.248

Soto, D., Humphreys, G. W., & Heinke, D. (2006). Working memory can guide pop-out search. *Vision Research, 46*, 1010–1018. doi: 10.1016/j.visres.2005.09.008

Spruyt, A., De Houwer, J., Hermans, D., & Eelen, P. (2007). Affective priming of non-affective semantic categorization responses. *Experimental Psychology, 54*, 44–53. doi: 10.1027/1618-3169.54.1.44

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*, 220-247. doi: 10.1207/s15327957pspr0803_1

Sweldens, S., Corneille, O. & Yzerbyt, V. (2014). The Role of Awareness in Attitude Formation Through Evaluative Conditioning. *Personality and Social Psychology Review, 18*, 187-209. doi: http://dx.doi.org/10.1177/1088868314527832

Vandenbosch, K., & De Houwer, J. (2011). Failures to induce implicit evaluations by means of approach-avoid training. *Cognition and Emotion, 25,* 1311-1330. doi: 10.1080/02699931.2011.596819

Van Dessel, P., De Houwer, J., & Gast, A. (2015). Approach –Avoidance Training effects are Moderated by Awareness of Stimulus-Action Contingencies. Manuscript submitted for publication.

Van Dessel, P., De Houwer, J., Gast, A., & Smith, C. T. (2015). Instruction-Based Approach–Avoidance Effects: Changing Stimulus Evaluation via the Mere Instruction to Approach or Avoid Stimuli. *Experimental Psychology, 62*, 161-169. doi: 10.1027/1618- 3169/a000282

Van Hiel, A. & Mervielde, I. (2005). Authoritarianism and social dominance orientation: Relationships with various forms of racism. *Journal of Applied Social Psychology 35*, 2323-2344. doi: 10.1111/j.1559-1816.2005.tb02105.x.

Vermeiren A., & Cleeremans A. (2012) The validity of d' measures. *PloS ONE, 7*, e31595. doi: 10.1371/journal. pone.0031595.g003

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 100*, 426–432. doi:10.1037/a0022790

Woud, M. L., Becker, E. S., & Rinck, M. (2008). Implicit evaluation bias induced by approach and avoidance. *Cognition and Emotion, 22*, 1187-1197. doi: 10.1080/02699930801906843

Woud, M. L., Becker, E. S., & Rinck, M. (2011). Induction of implicit evaluation biases by approach-avoidance training: A commentary on Vandenbosch and De Houwer. *Cognition and Emotion, 25*, 1331-1338. doi: 10.1080/02699931.2011.628300

Woud, M. L., Becker, E. S., Lange, W. G., & Rinck, M. (2013). To like or not to like: The effect of approach-avoidance training on neutral, angry and smiling face stimuli. *Psychological Reports, 113*, 199-216. doi: 10.2466/21.07.PR0.113x10z1

Woud, M. L., Maas, J., Becker, E.S., & Rinck, M. (2013). Make the manikin move: Sumbolic approach-avoidance responses affect implicit and explicit face evaluations. *Journal of Cognitive Psychology, 25*, 738-744. doi: 10.1080/20445911.2013.817413

Xu, K., Nosek, B. A., & Greenwald, A. G. (2014). Race IAT 2002-2013 [dataset]. Open Science Framework, Available at: https://osf.io/52qxl/.

Young, A. I., & Fazio, R. H. (2013). Attitude accessibility as a determinant of object construal and evaluation. *Journal of Experimental Social Psychology, 49*, 404-418. doi: 10.1016/j.jesp.2012.12.009

Zanon, R., De Houwer, J., Gast, A., Smith, C. T. (2014). When does relational information influence evaluative conditioning? *Quarterly Journal of Experimental Psychology, 67*, 2105-2122. doi: 10.1080/17470218.2014.907324

CHAPTER 6

GENERAL DISCUSSION

In the current dissertation, we examined the effects of approach and avoidance (AA) training and instructions on implicit evaluation. Implicit evaluation can be defined as the automatic impact of stimuli on evaluative responses (De Houwer, Gawronski, & Barnes-Holmes, 2013) and is most often attributed to the activation of associations in memory (for a review, see Hughes, Barnes-Holmes, & De Houwer, 2011). Our research was inspired by the idea that propositions, rather than associations, guide implicit evaluation, as proposed by the single-process propositional model of evaluation (De Houwer, 2014). Based on this idea, we examined whether propositions mediate the effects of AA training on implicit evaluations. Unlike to what is typically assumed on the basis of association formation models of AA training (Woud, Maas, Becker, & Rinck, 2013; Phills, Kawakami, Tabi, Nadolny, & Inzlicht, 2011), propositional models imply that the effects of AA training involve the conscious acquisition of propositional knowledge (De Houwer, 2009; Mitchell, De Houwer, & Lovibond, 2009). The focus of my dissertational research was twofold. First, we examined whether the acquisition of stimulus-action contingency knowledge via instructions can cause changes in implicit evaluations. Second, we examined whether effects of AA training on implicit evaluations depend on conscious awareness of the stimulus-action contingencies. The results presented in each chapter are outlined in detail in the following section.

CRITICAL OVERVIEW OF THE STUDIES AND FINDINGS

Research line 1: Effects of AA instructions on implicit evaluation

In **Chapter 2**, we investigated whether typical AA effects occur also when participants do not perform the AA actions but are merely instructed about the contingencies between stimuli and AA actions. Five experiments were conducted to answer this question. In each experiment, participants received instructions to approach one stimulus and avoid another stimulus. They were asked to remember this information well as they would first need to complete a different task. Immediately following these instructions, participants' implicit evaluations of the two stimuli were registered. Results showed that AA instructions can influence implicit evaluations. Participants exhibited more positive implicit evaluations of the stimuli they were instructed to approach than of the stimuli they were instructed to avoid. However, these effects were critically dependent on the type of stimulus that was used. Though AA instructions clearly influenced evaluations of novel stimuli (i.e., fictitious social groups and novel non-words), no AA instruction effects were observed on evaluations of well-known stimuli (i.e. well-known social groups). This pattern of results was obtained regardless of whether implicit evaluations were registered with an Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) or using an evaluative priming task (Fazio, Sanbonmatsu, Powell, & Kardes, 1986). Similarly, we observed effects of AA instructions on explicit evaluations of fictitious social groups but not on explicit evaluations of well-known social groups (i.e., self-reported ratings of liking and warmth).

The present results show that changes in implicit (and explicit) stimulus evaluations can result not only from extended AA training but also from mere instructions about the relations between stimuli and future AA movements. These findings support the idea that propositional processes play an important role in the learning of implicit evaluations (De Houwer, 2009, 2014; Mitchell et al., 2009). Even AA training effects, which are often considered prototypical examples of effects that result from automatic processing in an associative system, may (at least partly) result from processes that involve the acquisition of propositional information.

There are, however, several reasons why it would be premature to conclude that AA training effects are mediated by propositional processes. First, we examined effects of AA instructions rather than effects of AA training. Though instructions about stimulus-action contingencies are often included in the procedures of AA training studies (e.g., Kawakami, Phills, Steele, & Dovidio, 2007), this is not always the case (e.g., Woud et al., 2013). Also, in studies that do include AA instructions, effects may result from several distinct processes. For instance, changes in liking may arise partly as the result of propositional processes that involve the acquisition of propositional contingency information via instructions and partly as a result of the (gradual) acquisition of associations through actual AA training. Thus, the present research might merely demonstrate a second pathway to evaluative change in addition to evaluative change via actual AA behaviors. Such an explanation would resonate with dual-process models that suggest that propositional processes and associative processes can produce (distinct) changes in implicit evaluation (e.g., the Associative-Propositional Evaluation [APE] model; Gawronski & Bodenhausen, 2006, 2011, 2014).

Second, even though our statistical tests had sufficient power to detect even a small effect, we did not observe AA instruction effects on evaluations of well-known stimuli. In contrast, AA training has been known to produce such effects (Kawakami et al., 2007). This observation is consistent with a core assumption of many evaluation theories that the learning of implicit evaluations reflects the gradual buildup of strength of an associative representation as the result of repeated pairings (Strack & Deutsch, 2004; Rydell & McConnell, 2006; Gawronski & Bodenhausen, 2006). Because implicit evaluations of well-known stimuli should reflect strong associations, these evaluations may change only through repeated experiences (e.g., actual AA training). Dual process-models like the APE model can explain why we do find effects on implicit evaluations of novel stimuli. Implicit evaluations of novel stimuli may change also when externally provided propositions influence participants' explicit evaluation of the stimulus and contribute to the proactive construction of *new* associations (Whitfield & Jordan, 2009). This explanation, however, does not fit well with the observation that changes in implicit evaluations were not fully mediated by changes in explicit evaluations (Experiments 4 and 5). We explored this issue further in the next chapter.

In **Chapter 3**, we examined mediation of the effect of AA instructions on implicit evaluations by changes in explicit evaluations. The results described in Chapter 2 provided preliminary evidence that the impact of AA instructions on implicit evaluations is not fully mediated by changes in explicit evaluation. In the studies reported in Chapter 3, participants were instructed to approach or avoid the names of members of two fictitious social groups. Before presenting these AA instructions, we provided participants with information about the evaluative traits of these social groups. We observed that trait instructions cancelled the effects of AA instructions on explicit evaluations but not on implicit evaluations. Participants exhibited an implicit preference for the to-be-approached social group over the to-be-avoided social group even though information was available that was more diagnostic of the valence of the groups. This pattern of results was observed also when AA instructions were supplemented with actual AA training. Statistical mediation analyses further corroborated that both AA instructions and AA training can have a direct influence on implicit evaluation (i.e., unmediated by changes in explicit evaluation).

These findings reveal important similarities between instruction-based and training-based AA effects. Changes in implicit evaluations can occur even when the AA instructions (and AA training) provide information that is not considered highly diagnostic of the evaluative properties of the stimulus and therefore do not influence explicit evaluations. This contrasts with the assumption of the APE model that instructions can influence implicit evaluations only if the provided information is considered valid and causes deliberate changes in explicit evaluations (Gawronski & Bodenhausen, 2006). Our results suggest that AA instruction effects are not entirely the result of controlled, non-automatic processes that involve the intentional use of the provided information. For instance, effects cannot be exclusively due to the fact that participants infer that the experimenter wants them to evaluate the stimulus in a certain way and they comply with this demand (De Houwer, Beckers, & Moors, 2007). Rather, the acquisition of propositional information about stimulus-action contingencies may allow for an *automatic* effect on implicit evaluation. More specifically, AA instructions cause changes in implicit evaluation that seem to occur in the absence of participants' intention that the changes take place. The processes underlying this effect thus seem to possess at least one feature of automaticity (Moors & De Houwer, 2006). The current results fit well with the idea that propositional information can automatically influence implicit evaluation (De Houwer, 2014).

However, it would again be premature to make strong conclusions about the causal role of propositional processes in AA effects on the basis of the available evidence. Though the current findings reveal important similarities between the effects of AA training and those of AA instructions we cannot conclude that both types of effects are due to the same mental processes (e.g., the formation and activation of propositions). For instance, AA training might still provide an additional source for changes in implicit evaluations due to the acquisition and strengthening of stimulus-action associations (Gawronski & Bodenhausen, 2006). In line with this idea, we found that AA training adds to the effect of AA instructions on implicit evaluations. Participants exhibited a stronger AA effect on implicit evaluation when AA instructions were supplemented with AA training. This added effect was observed only for participants who received trait instructions. This observation fits with the idea that the gradual formation of associations is necessary for changing evaluations of stimuli that previously acquired a specific valence (Rydell & McConnell, 2006). It also fits well with the evidence reported in Chapter 2 that AA instructions produce smaller effects on evaluations of valenced, well-known, stimuli than on evaluations of entirely novel stimuli. The current results extend these findings by showing that added experience in approaching or avoiding the stimuli causes stronger effects on evaluations of valenced stimuli. However, they also show that AA instructions *can* influence evaluations of valenced stimuli under certain conditions, indicating that propositional processes might also play a role in changing evaluations of stimuli that previously acquired a specific valence. Note also that the observation of an added effect of actual training does not imply that AA training effects necessarily also reflect the operation of association-formation mechanisms. For instance, receiving additional training might facilitate the acquisition or activation of propositional information which could influence implicit evaluation. Future studies are required to examine the (nature of the) mental processes that underlie this added effect of AA training.

In the current studies, all participants received AA instructions (either supplemented with AA training or not). It is therefore difficult to make conclusions about the processes that underlie AA training effects that do not include such instructions. For instance, propositional processes might only play a role in AA effects when participants are informed about the stimulus-action contingencies. A fundamental question that remains is whether AA training effects will depend on propositional contingency information even in the absence of AA instructions. In the next research line we provide an answer to this question.

Research line 2: The role of contingency awareness in AA training effects

The aim of **Chapter 4** was to test whether the effect of AA training on implicit evaluations depends on participants' awareness of the stimulus-action contingencies. If AA training effects are the result of propositional processes then these effects should hinge on participants' ability to acquire conscious propositional contingency information, which should (at least partially) be reflected in contingency awareness (Mitchell et al., 2009). In the present studies, we used a variant of the procedure introduced by Woud, Becker, and Rinck (2008) to examine AA training effects on implicit evaluations of novel face stimuli. In this procedure, participants repeatedly approach or avoid novel faces on the basis of a certain stimulus feature (i.e., the color of the frame surrounding the face pictures) that is unrelated to the target feature of the stimulus (i.e., the identity of the face pictures). Subsequently, implicit evaluations of the faces are registered with an evaluative priming task. The previous studies that used this procedure produced mixed results. Whereas Woud and colleagues reported finding a significant effect of AA training on implicit evaluations of novel faces, Vandenbosch and De Houwer (2011) failed to find any evidence for AA training effects in five experiments and failed to reproduce the effect reported by Woud et al. when reanalyzing their data. For the current studies, we adapted the procedure in a number of ways such that the acquisition of information about stimulus-action contingencies would be facilitated. For instance, we used a smaller number of evaluative stimuli (i.e., pictures of 8 instead of 12 neutral faces) and allowed participants to experience a 100% contingency between a certain face and AA action. Importantly, we also made adaptations to address the importance of contingency awareness in AA training effects. First, we tried to capture participants' awareness of the experienced face-action contingencies by measuring participants' memory about the relation between faces and actions. Second, we provided half of the participants with instructions that specified the stimulus-action contingencies.

Results of Experiment 1 indicated that contingency awareness moderated AA training effects on implicit (and explicit) evaluations of novel faces. Participants exhibited a preference for approached faces over avoided faces only if these faces had been correctly identified as an approached or avoided face. In contrast, our manipulation of contingency awareness via contingency instructions failed to produce any evidence that contingency awareness influenced AA training effects even though the manipulation did influence measured contingency awareness. Experiment 2 involved the same procedure with the exception that none of the participants received any contingency information. We observed that AA training caused significant changes in implicit and explicit evaluations, and that these effects were moderated by participants' awareness of the stimulus-action contingencies.

These results support the idea that AA training influences the liking of a stimulus only after participants acquire conscious propositional knowledge about the relation between AA action and stimulus. They accord with our previous findings that the acquisition of propositional contingency information (via instructions) causes changes in implicit (and explicit) stimulus evaluations and provide strong evidence that propositional processes play an important role in AA training effects.

The observation that contingency awareness moderates AA training effects does not fit well with the idea that an automatic association formation mechanism underlies these effects. However, we cannot rule out the possibility that associative processes do play some role in AA training effects. On the one hand, it is possible that the observed changes in implicit evaluations actually resulted from the (gradual) formation and activation of associations. There are several reasons why changes in liking that are due to association formation might have correlated with contingency awareness. For instance, participants could have trusted their feelings toward the stimuli to answer the contingency awareness questions, relying on affect-as-information. In such cases, the AA training effect on evaluations may be stronger for stimuli classified as contingency aware than for stimuli classified as contingency unaware (see Hütter, Sweldens, Stahl, Unkelbach, & Klauer, 2012 for a similar argument in the context of evaluative conditioning). Alternatively, both the formation of associations during action performance and contingency awareness may critically depend on the same boundary conditions (e.g., participants' attention to the identity of the stimuli). Supporting the idea that contingency knowledge does not mediate AA training effects, our experimental manipulation of contingency awareness (i.e., whether instructions about stimulus-action contingencies were also presented) did not significantly influence stimulus evaluations. However, given our previous findings, we are confident that AA instructions (which in essence also specify stimulus-action contingencies) can influence implicit and explicit evaluations of novel stimuli. In fact, we did obtain AA instruction effects with a similar procedure and similar stimuli in another study (see Chapter 5, Experiment 2). Failure to find this effect in Experiment 1 may simply be due to a lack of power to detect such an effect in the between-subjects analyses.

On the other hand, our procedures may have interfered with the observation of AA effects that result from automatic association formation processes. For instance, because participants were aware of most of the stimulus-action contingencies in both our experiments, statistical power was more strongly reduced for analyses that tested AA effects on evaluations of contingency unaware faces than for analyses that tested AA effects on evaluations of contingency aware faces. This might explain why we only observed significant AA effects for contingency aware faces. However, two types of evidence go against this explanation. First, Bayesian analyses indicated that our data provided strong evidence that AA training influences evaluations of contingency aware faces and already provided substantial evidence that AA training does not influence evaluations of contingency indiscriminate faces. Second, Vandenbosch and De Houwer (2011) used a similar procedure with the exception that they used a larger number of stimuli and less-than-perfect contingencies, which is likely to reduce contingency awareness. A high-powered analysis including the data of 9 experiments with this procedure did not provide any evidence for effects of AA training. Of course, we can never fully exclude the possibility that other procedural details impeded the detection of effects due to association formation in our studies (and in the studies by Vandenbosch & De Houwer). Though it is currently unclear what these procedural details might be, some evidence supports this idea and indicates that AA training can, under certain conditions, produce effects on implicit evaluations in the absence of contingency awareness (i.e., with subliminal stimulus presentations such as in the studies of Kawakami et al., 2007). We explored this issue further in the next chapter.

In **Chapter 5**, we tested whether AA training with subliminal stimulus presentations can cause changes in implicit evaluations. Experiment 1 was an exact replication of the experiment by Kawakami et al. (2007, Experiment 2) who found subliminal AA training effects on implicit evaluations of Black and White social groups. We were unable to replicate this result. In contrast, Bayesian analyses of our data provided substantial evidence that training to approach subliminally presented photographs of Black people and avoid subliminally presented photographs of White people does not produce changes in participants' implicit (or explicit) evaluations of these groups.

Experiment 2 examined effects of subliminal AA training on evaluations of novel face stimuli. In this experiment, participants were informed about the contingencies between four face stimuli and the AA actions in the subliminal AA training task before they performed the task. Unbeknownst to participants, contingency instructions specified the correct contingencies for only two of the four face stimuli. Results showed that the contingency instructions influenced implicit and explicit evaluations of the stimuli but subliminal AA training did not.

Finally, Experiment 3 investigated effects of subliminal and supraliminal AA training on evaluations of non-existing words. This study included a manipulation designed to facilitate participants' attention to contingencies in the subliminal AA training task. As Custers and Aarts (2011) argued, such a manipulation might facilitate the formation of associations in memory under conditions of subliminal stimulus presentations. Participants performed a supraliminal AA training task in which they repeatedly approached or avoided certain non-words. For half the participants the supraliminal AA training task involved specific stimulus-action contingencies (attention-tuning condition), for the other participants the identity of the stimulus did not predict the action (control condition). After performing the supraliminal AA training task, participants performed an AA training task with subliminal presentations of other non-word stimuli. Even though most participants in the attention-tuning condition were aware of the contingencies in the supraliminal AA training task and exhibited typical effects of the supraliminal AA training, we observed no effects of subliminal AA training on implicit or explicit evaluation in this condition or in the control condition. We also found that the effects of the supraliminal AA training critically depended on contingency awareness.

These findings provide further evidence that propositional processes play an important role in AA training effects. AA training might produce changes in evaluations only when participants are aware of the stimuli and of the stimulus-action contingencies. A meta-analysis including all the available evidence on subliminal AA training currently favors the hypothesis that subliminal AA training does not influence implicit or explicit evaluations. This contrasts with the idea that the repeated paring of stimuli and AA actions allows for the automatic formation of associations that underlie (implicit) evaluations (Woud et al., 2013; Phills et al., 2011). Of course, it is still possible that subliminal AA training effects can be reliably observed, but that we merely failed to do so in the current studies. For instance, subliminal AA training effects may depend on specific boundary conditions such as whether participants have had ample previous experience with the stimuli (and therefore arise only with well-known stimuli). Currently, however, we believe there is insufficient evidence to conclude that subliminal AA training can cause changes in evaluations. It will be of particular importance to demonstrate the existence of subliminal AA training effects in studies that include sensitive checks of perceptual stimulus discriminability and contingency awareness.

To sum up, the experiments reported in the four empirical chapters provided strong evidence that the acquisition of propositional contingency information plays a crucial role in AA effects. We observed that (1) the acquisition of stimulus-action contingency information produces (direct) changes in implicit evaluations and (2) that AA training effects on implicit evaluations are moderated by awareness of the stimulus-action contingencies.

THEORETICAL IMPLICATIONS

Now that we have provided a brief overview and discussion of the different findings reported in the current thesis, in this section we consider the various implications of our findings for current and future theories about the mental processes underlying AA effects and implicit evaluation. We first address how our findings fit with current (associative) accounts of AA effects. Next, we introduce alternative, propositional, accounts of AA effects. Finally, we discuss implications for current theories of stimulus evaluation.

Associative accounts of AA effects

As we discussed in the General Introduction, a number of different explanations for AA training effects have been put forward that all incorporate the assumption that association formation mechanisms underlie these effects. The current results strongly constrain these associative accounts.

First, the finding that typical AA effects can be observed even in the absence of actual AA actions does not fit well with the idea that an association formation mechanism underlies AA effects. It is most often assumed that associations only form gradually as the result of many different experiences and therefore require a large number of stimulus-action pairings (Kawakami et al., 2007; Woud et al., 2008). We see two ways in which AA instruction effects can be reconciled with current associative accounts. On the one hand, AA instruction effects might differ from AA training effects in that they do not result from the same mental processes. For instance, whereas actual AA training might produce changes in implicit evaluations that are the result of automatic association formation processes, AA instruction effects might depend on other, more controlled, processes. One way to test this idea is by comparing the (automaticity) features of instruction-based and training-based AA effects. In Chapter 3, we reported evidence that both AA instructions and AA training can produce unintentional changes in implicit evaluations. Future research is needed to establish whether instruction-based and training-based effects are similar also regarding other features, for example regarding uncontrollability (see Gawronski, Balas, & Creighton, 2014).

On the other hand, both AA instruction and AA training effects might depend on association formation processes. Some theorists have argued that strong associations can form even as the result of a single pairing (e.g., Fazio, 2007; Field, 2006). Though instructions do not involve any pairing of AA action and stimulus, they do involve other pairings. For instance, in AA instructions a certain stimulus (e.g., the word UDIBNON) is paired with valenced action words (e.g., the word ‘approach’). This one-time pairing might be sufficient to create an associative link between the representations of the stimulus and a valenced representation. Automatic activation of this association might produce the observed changes in implicit evaluation (Woud et al., 2008). Note that participants can also engage in mental rehearsal of the instructions. This mental rehearsal could allow for multiple (covert) pairings of stimulus and action words (through inner speech). Alternatively, AA instruction effects might occur when participants engage in mental simulation of AA actions (see Glenberg & Kaschak, 2002, for an overview of evidence that people engage in the mental simulation of actions that are described in instructions). Current associative accounts of AA effects can accommodate AA instruction effects if they assume that mental simulation allows for the formation of associations between representations of the self and stimulus representations (Phills et al., 2011) or between stimulus representations and motivational representations (Cacioppo, Priester, & Berntson, 1993). Note, however, that in our AA instruction studies participants were not informed how they would perform the approach and avoidance behavior which makes it is less likely that effects resulted from mentally simulated action.

Second, current associative accounts cannot easily explain why AA training effects only seem to arise when participants are aware of stimulus-action contingencies. Because association formation is often considered an automatic process, AA training effects are typically thought not to depend on contingency awareness (Kawakami et al., 2007). As we discussed in the previous section, the current results can be reconciled with an association formation mechanism if one assumes that specific boundary conditions determine when association formation occurs and these conditions also determine contingency awareness. Current theories, however, remain unclear on the necessary environmental conditions that determine (automatic) association formation (Shanks, 2007) and AA training effects (Vandenbosch & De Houwer, 2011).

Propositional accounts of AA effects

In general, we believe that the current results can be more easily explained by alternative, propositional, accounts of AA effects. In line with propositional models of EC (De Houwer, 2009; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010; Mitchell et al., 2009), AA training might influence the liking of a stimulus only after participants acquire conscious propositional knowledge about the relation between action and stimulus. There a number of possible pathways via which the acquisition of this information might lead to changes in implicit (and explicit) stimulus evaluations.

First, when participants acquire propositional contingency information, they might generate novel propositions by making specific inferences. For instance, participants who learn that they approach a stimulus may infer that they like this stimulus, and participants who learn that they avoid a stimulus may infer that they do not like this stimulus. These inferences could arise because of the knowledge that, typically, positive things are approached and negative things are avoided. People may have learned this rule through previous experiences during which they approached liked stimuli and avoided disliked stimuli. Although this knowledge does not logically imply the reverse (i.e., that approached things are good and avoided things are bad), people are known to be prone to affirm the consequent (i.e., conclude that A is true on the basis of the fact that A implies B and B is present). Note that people make inferences not only in a controlled manner but also when they have no intention to make them or when they are unaware of the inference itself (Critcher & Risen, 2014; Uleman, Saribay, & Gonzalez, 2008; see Kruglanski & Gigerenzer, 2011, for an overview of evidence of automatic inferences). The inferences underlying AA effects might also occur under certain conditions of automaticity (e.g., because the application of the underlying rule is well-practiced, Kruglanski & Gigerenzer). Thus, inferential processes might allow for the acquisition of novel propositional information in either a more automatic or a more deliberate manner.

Once participants infer that the approached stimulus is good and the avoided stimulus is bad, they may use this newly acquired propositional information for evaluation. The process by which this new information influences evaluation might also be automatic to a greater or lesser degree. For instance, more deliberate changes in liking may arise when participants regard the AA information as a valid source of information or when they believe that the experimenter wants them to evaluate the stimulus in a certain way and they want to comply with this demand (De Houwer et al., 2007). The observation that AA effects on implicit evaluations can occur also in the absence of changes in explicit evaluations suggests that the changes in liking are not necessarily deliberate. Rather, the newly acquired propositional information might influence implicit evaluation in a more automatic manner (De Houwer, 2014). For instance, presentation of the approached stimulus during implicit evaluation tasks may allow for the automatic (e.g., fast and unintentional) activation of the proposition that this stimulus is positive. This might facilitate certain spontaneous evaluative responses (e.g., fast categorization responses in IAT blocks where the approached stimulus is categorized with the same key as liked stimuli).

A second possible propositional explanation for AA effects centers on the idea that the acquisition of propositional contingency information triggers the activation of other propositional information through similarity-based retrieval principles rather than through inferential processes. For instance, presentation of an approached stimulus during an evaluation task may allow for the immediate activation of the contingency information (e.g. that stimulus A was approached). In turn, this may trigger the activation of other information as a function of the similarity between the information and previously acquired information (Hintzman, 1986; Nosofsky & Palmeri, 1997). For instance, participants may activate information or memories that relate to the concept 'approach' (e.g., that they consider this concept positive) which might influence participants' immediate response to the stimulus (e.g., evoke a positive feeling or facilitate responding 'positive' in an implicit evaluation task).

The described propositional accounts can easily explain the current results.

First, these accounts assume that changes in liking can occur only when participants obtain stimulus-action contingency information. Because the acquisition of contingency information should (at least partly) be reflected in contingency awareness, AA effects might critically depend on contingency awareness. Note that the relation between measured contingency awareness and the (previous) acquisition of contingency information is not necessarily perfect. Thus, these accounts do not exclude the possibility that participants might exhibit AA effects even when they are unaware of the contingency information (a) at the time of contingency measurement, (b) at the time of evaluation measurement, or (c) at the time of acquisition of the propositional contingency information (but see Mitchell et al., 2009). Each of these questions can be addressed in future studies and could provide further information on the automatic or controlled nature of these effects. Second, propositional accounts assume that acquired propositional information will influence evaluation independent of whether this information is acquired on the basis of experience or on the basis of instructions (De Houwer, 2009; Mitchell et al.). They therefore predict the existence of AA instruction effects a priori and can explain why AA instruction effects share important similarities with AA training effects.

Overall, we believe that the current findings fit better with a propositional explanation of AA effects than with current associative explanations. Because our results indicate that propositional processes are involved in AA effects, a propositional account that can explain these results is to be preferred over an account that additionally postulates the existence of an entirely different second mechanism (i.e., association formation) for reasons of parsimony.

Mental process accounts of implicit evaluation

The current findings also have implications for current and future theories of stimulus evaluation. First, our results are difficult to reconcile with single-process associative models of stimulus evaluation (e.g., Fazio, 2007; Olson & Fazio, 2009). These models typically assume that co-occurrences (e.g., stimulus-action pairings) are automatically encoded in the form of mental associations that mediate implicit evaluation. This assumption strongly contrasts with our findings that instructions about pairings influenced implicit evaluation, indicating that the formation of implicit evaluations does not necessitate (repeated) pairings. It also contrasts with the observation that effects that do involve actual pairings depend on contingency awareness rather than on experienced stimulus-action contingencies.

Second, most dual-process models of evaluation also assume that associations underlie implicit evaluation and that these associations form gradually as the result of many pairings (e.g., Smith & DeCoster, 2000; Rydell & McConnell, 2006). These models therefore also have great difficulty in dealing with the current results. However, contemporary dual-process models in which association formation processes can interact with propositional learning processes can better explain our findings. For instance, the APE model developed by Gawronski and Bodenhausen (2006), assumes that the formation of associations can be mediated by the (propositional) encoding of the pairings. As a result, this model would predict that AA instructions can influence both implicit and explicit evaluations. However, the APE model also assumes that the acquisition of propositional information influences implicit evaluations only indirectly (i.e., through a mediating influence on explicit evaluation). Participants need to consciously consider the evaluative stimulus as positive or negative before a change in implicit evaluations may occur (Gawronski & LeBel, 2008). This contrasts with the observation that AA instructions can influence implicit evaluations even in the absence of changes in explicit evaluations. To accommodate the current findings, dual-process models would need to make additional assumptions (e.g., that propositional information can have a direct influence on implicit evaluation).

Finally, purely propositional models of evaluation (e.g., De Houwer, 2009; Mitchell et al., 2009; De Houwer, 2014) postulate that implicit and explicit evaluation depends entirely on propositional processes. Because our findings imply that propositional processes play an important role in AA effects, these models fit best with the current results. Both AA training and AA instructions may allow for the acquisition of propositional contingency information. As we described in the previous section, there are a number of possible pathways via which this might influence implicit (and explicit) evaluation in a purely propositional manner.

LIMITATIONS AND FUTURE DIRECTIONS

In this final section, we first describe important limitations of our work and possible future directions for research on the effects of AA instructions and AA training. We then provide recommendations for future research that aims to shed light on the mental processes underlying AA effects and implicit evaluation.

A first issue that necessitates further examination is to what extent AA effects depend on the nature of the stimuli. In Chapter 2 we found that the nature of the stimulus can be an important boundary condition of AA instruction effects. AA instructions influenced evaluations of novel but not well-known social groups. This contrasts with findings that AA training can influence evaluations of well-known groups (Kawakami et al., 2007). In Chapter 3 we observed that AA training adds to the effect of AA instructions on implicit evaluations of stimuli that previously acquired a specific valence (through trait instructions) but not of entirely novel stimuli. This study also found that, overall, AA instruction and AA training effects were strongly reduced for the valenced stimuli, a finding that accords with other studies showing that AA effects depend on the valence of the stimuli (Centerbar & Clore, 2006; Jones, Vilensky, Vasey, & Fazio, 2013; Woud, Becker, Lange, & Rinck, 2013). For instance, Woud et al. found that AA training influenced evaluations of faces with neutral but not angry or smiling expressions. Further studies are needed to examine the stimulus qualities that determine the effectiveness of AA effects and the extent to which the impact of stimulus qualities is similar for instruction-based and training-based effects. This issue could be addressed by directly comparing effects of AA instructions and AA training on evaluations of novel and well-known stimuli (e.g., prejudiced social groups) or on evaluations of stimuli that have a positive, negative or neutral valence. Such studies might not only extend knowledge about the moderators of AA instruction and AA training effects but might also provide further insight into the mental processes that underlie these effects. For instance, if we find that AA training but not AA instructions can cause changes in implicit evaluations of well-known stimuli this would support the idea that distinct processes underlie these effects.

More generally, we hope that future research will further address the similarities and differences of AA instruction and AA training effects. Such studies can be important also for practical reasons. For instance, if AA instructions and AA training produce similar effects, then using instructions would be a more cost-effective way to change a person’s evaluations than providing ample AA training. If we do find that there are important dissociations between instruction-based and training-based AA effects (e.g., for changing evaluations of well-known stimuli), follow-up studies could investigate whether observed dissociations can be reduced under certain circumstances. This research might also provide important information about the nature of the processes underlying AA effects. For instance, one could compare effects of AA training and effects of AA instructions that are presented repeatedly to examine whether effects depend on repeated pairings (of action and stimulus or action word and stimulus within an instruction) rather than on actual performance of the actions. It might also be interesting to investigate whether dissociations are reduced when participants receive (hypnotic) suggestions that they experienced a large number of stimulus-action pairings. Under these circumstances, participants acquire propositional information that is also acquired with AA training procedures but not with typical AA instruction procedures. This allows us to examine whether effects depend on the acquisition of this propositional information.

Another important limitation of the current studies is that implicit evaluations were registered only with the IAT or with an evaluative priming task. Though these are the two most often used implicit evaluation measures, it may be important for future research to extend its scope to other implicit measures that we have not considered in the present dissertation (e.g., the Affective

Misattribution Procedure, Payne, Cheng, Govorun & Stewart, 2005; the Extrinsic Affective Simon Test, De Houwer, 2003; the Go/No-Go Association Test, Nosek & Banaji, 2001). Note that in all our studies the AA effects on evaluations as registered with the IAT, evaluative priming task, or explicit evaluation measures, revealed similar patterns of results (e.g., AA instructions influence evaluations of novel but not well-known stimuli, AA training effects occur only in the presence of contingency awareness,…). The only exception is that, in the presence of trait instructions, we observed AA instruction and AA training effects on implicit evaluations as registered with the IAT but not on explicit ratings of liking. Examining whether such a dissociation can be observed also when other measures of implicit and explicit evaluation are used might further inform us about the automaticity of the processes that underlie AA effects and about the extent to which different evaluation measures are sensitive to these processes.

A related issue is that the current studies only used one specific question to capture participants' awareness of the experienced stimulus-action contingencies. In the evaluative conditioning literature a number of different approaches have been used for the measurement of contingency awareness (see Sweldens, Corneille, & Yzerbyt, 2014, for an overview). This is important because no single measure can accurately capture contingency awareness. For instance, one problem with the measure that we used is that it does not satisfy the immediacy criterion (i.e., that the test follows the learning episode quickly to avoid forgetting or interference, Lovibond & Shanks, 2002). It may be useful in future studies to include a measurement of contingency awareness at the time of learning (i.e., during performance of the AA training task). However, to address whether contingency awareness mediates AA effects the best approach would be to use experimental procedures rather than correlational procedures (Dedonder, Corneille, Bertinchamps, & Yzerbyt, 2014; Gawronski & Walther, 2012). The subliminal AA training procedure provides one way to do this because the experimental set-up can prevent contingency awareness but this procedure has important limitations (see Sweldens et al.). We hope that future studies also use other experimental procedures to further examine the issue of contingency awareness (e.g., by relying on parafoveal presentations of stimuli, Dedonder et al.).

We now turn our attention to implications and limitations of our research for possible practical applications. One important issue is that all the experiments reported in this dissertation restricted investigation to the effects of AA procedures on implicit (and explicit) evaluations. Recently, there have been many reports that AA training can also influence other (dysfunctional) behaviors such as alcohol consumption (Wiers, Eberl, Rinck, Becker, & Lindenmeyer, 2011), unhealthy food consumption (Schumacher, Kemps & Tiggeman, 2016), cigarette consumption (Wittekind, Feist, Schneider, Moritz, & Fritzsche, 2015), or fear responses (Dibbets & Fonteyne, 2015; Amir, Kuckertz, & Najmi, 2013; Huijding, Muris, Lester, Field, & Joosse, 2011). These findings are often considered to be mediated by training-induced changes in implicit evaluations (Jones et al., 2013; Wiers et al.). In line with this idea, some studies have provided evidence that the induction of changes in implicit evaluations directly influences some of the dysfunctional behaviors under investigation (e.g., Houben, Havermans, & Wiers, 2010). However, other research has shown that changes in implicit evaluations often do not mediate AA training effects on dysfunctional behavior (e.g., Dibbets & Fonteyne; Lester et al., 2015; Eberl et al., 2014). It thus seems important to further investigate to what extent the effects of AA training on dysfunctional behaviors depend on changes in (implicit) evaluation. If these effects are mediated by changes in evaluation, then, based on the current findings, it can be argued that the acquisition of contingency information may be an important factor also for these studies. Comparing the effects of AA instructions and AA training on these behaviors could help provide an answer to this question.

An alternative explanation for AA training effects on certain dysfunctional behaviors is that AA training modifies participants' automatic tendencies to approach or avoid specific stimuli which, in turn, influences the behaviors (see Sharbanee et al., 2014). In line with this idea, research has shown that (1) people exhibit an automatically activated approach-or avoidance tendency for certain well-known stimuli (i.e., they are faster to make either approach or avoidance movements in response to these stimuli; see Laham, Kashima, Dix, & Wheeler, 2015, and Phaf, Mohr, Rotteveel, & Wicherts, 2014, for recent overviews), and (2) this bias can be modified via AA training (Eberl et al., 2014; Dibbets & Fonteyne, 2015; Schumacher et al., 2016). However, it is unclear precisely *how* this “cognitive bias modification” may cause the observed changes in behavior. Moreover, some evidence suggests that automatic AA tendencies do not logically relate to the behaviors under investigation (Spruyt et al., 2013; Snelleman, Schoenmakers, de Mheen, 2015). For instance, Spruyt and colleagues observed that the likelihood of relapse in alcohol-dependent patients increased as the relative tendency to avoid alcohol increased. In future studies, it might be interesting to investigate the role of propositional processes also in these effects. One could test a number of different possible propositional accounts of these effects. For instance, when participants repeatedly perform AA actions in response to a specific stimulus they might infer that they will now perform these actions more easily. In the context of alcohol addiction, an alcoholic patient who repeatedly performs avoidance movements to alcoholic stimuli could, for instance, make the inference that they will better be able to avoid alcohol in the future. This self-efficacy belief might help a person to refrain from alcohol abuse such that avoid alcohol training can reduce relapse rates in alcoholic patients (Wiers et al., 2011).

In general, however, the studies we reported in this dissertation provide only limited information about possible practical implications. Though we observed changes in implicit and explicit evaluations as the result of AA training and AA instructions, it is unclear whether these procedures can produce ‘meaningful’ effects. Future studies are needed to investigate whether (both) these procedures can install evaluations that influence other (real-world) behaviors and to what extent these effects are robust or fragile and short-lived.

Finally, we want to end this dissertation by providing recommendations for further investigations about the mental processes that underlie AA effects and evaluation. First, we believe that, in order to advance research, it is crucial to define theoretical models that make clear, testable predictions. For instance, to gain more information about the processes underlying AA effects it might be useful to describe a propositional account that specifies the inferences that might underlie these effects. Novel predictions can then be derived from this account (e.g., that AA effects should be reduced if the invalidity of these inferences is highlighted) and testing these predictions in future studies may further our knowledge on the processes underlying AA effects. We also recommend this approach for testing cognitive theories of evaluation and classes of models such as dual-process and propositional models. Because these models often have a high degree of flexibility, distinguishing between these models on the basis of a single set of data is difficult, if not impossible. Proponents of these models can always make post-hoc adaptations to explain obtained results. To advance research it seems important that theories clearly specify the boundary conditions that determine when these processes will take place. For instance, we hope that dual-process theories will further specify the conditions that determine when associations are formed and to what extent this association formation process occurs automatically.

The second point we want to raise is that to promote scientific progress it seems crucial to clearly separate explanations that are situated at different levels of analysis (see De Houwer, 2011). For instance, AA training effects can be explained at the functional level of analysis (i.e., by explaining effects in terms of the causal impact of elements in the environment) and at the cognitive or mental process level (i.e., by explaining effects in terms of mental constructs such as the formation of propositions). Because AA training procedures involve pairings (or 'associations') of actions and stimuli these effects were traditionally defined at the cognitive level, as training procedures that install associations. Effects were assumed to rely on automatic association-formation processes and most studies were performed under this (rarely specified) assumption. However, from a functional perspective, AA training effects can be defined as changes in behavior that are due to the pairing of actions and stimuli and can in principle be due to a number of different processes. Accepting the claim that AA training involves association-formation might hamper the opportunity for novel theoretical expansion. A similar argumentation can be made for the case of implicit evaluation. If implicit evaluations are defined as changes in associations in memory (as is often the case, see Hughes et al., 2011), this rules out the possibility that changes in implicit evaluations can be due to different mechanisms. As we have shown in the current studies, important new insights can be obtained by questioning these assumptions and investigating the propositional processes underlying AA training effects and implicit evaluation.

OVERALL CONCLUSION

In the current dissertation our aim was to learn more about the conditions under which AA training causes changes in implicit evaluations and about the mental processes that underlie these effects. We obtained clear evidence that propositional processes play an important role in AA effects. This conclusion raises questions about the prevailing view that these evaluative learning effects are due to automatic association formation mechanisms. We believe that questioning the assumption that automatic effects of (evaluative) learning depend on association formation processes and examining the role of propositional processes in these effects may lead to important new discoveries and can have important practical implications. As such, the present thesis sets the stage for these new developments.

REFERENCES

Amir, N., Kuckertz, J. M., & Najmi, S. (2013). The effect of modifying automatic action tendencies on overt avoidance behaviors. *Emotion, 13,* 478-484. doi: 10.1037/a0030443

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*, 133–142. doi: 10.1016/0005-7967

Cacioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary determinants of attitudes II: Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology*, *65*, 5–17. doi: 10.1037/0022-3514.65.1.5

Centerbar, D. & Clore, G.L. (2006). Do approach-avoidance actions create attitudes? *Psychological Science, 17*, 22-29. doi: 10.1111/j.1467-9280.2005.01660.x

Critcher, C. R., & Risen, J. L. (2014). If he can do it, so can they: Incidental exposure to counterstereotypically-successful exemplars prompts automatic inferences. *Journal of Personality and Social Psychology, 106*, 359-379

Custers, R., & Aarts, H. (2011). Learning of predictive relations between events depends on attention, not on awareness. *Consciousness and Cognition, 20*, 368–378. doi: 10.1080/10640266.2012.715515

Dedonder, J., Corneille, O., Bertinchamps, D., & Yzerbyt, V. (2014). Overcoming Correlational Pitfalls: Experimental Evidence Suggests That Evaluative Conditioning Occurs for Explicit But Not Implicit Encoding of CS–US Pairings. *Social Psychological and Personality Science, 5*, 250-257. doi: 10.1177/1948550613490969

De Houwer, J. (2003). The Extrinsic Affective Simon Task. *Experimental Psychology, 50*, 77-85. doi: 10.1027//1618-3169.50.2.77

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

De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science, 6,* 202–209.

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

De Houwer, J., Beckers, T., & Moors, A. (2007). Novel attitudes can be faked on the Implicit Association Test. *Journal of Experimental Social Psychology, 43*, 972-978. doi: 10.1016/j.jesp.2006.10.007

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology*, *24*, 252–287. doi: 10.1080/10463283.2014.892320

Dibbets, P., Fonteyne, R. (2015). High spider fearfuls can overcome their fear in a virtual approach-avoidance conflict task. *Journal of Depression & Anxiety, 4*, 182, 1-7.

Eberl C., Wiers R. W., Pawelczack, S., Rinck, M., Becker, E. S. , Lindenmeyer J. (2014). Implementation of approach bias re-training in alcoholism-how many sessions are needed? *Alcoholism: Clinical and experimental research, 38,* 587-594. doi: 10.1111/acer.12281.

Fazio, R. H. (2007). Attitudes as Object–Evaluation Associations of Varying Strength. *Social Cognition, 25,* 603−637. doi: 10.1521/soco.2007.25.5.603

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*, 229–238. doi: 10.1037/0022-3514.50.2.229.

Field, A. P. (2006). Is conditioning a useful framework for understanding the development and treatment of phobias? *Clinical Psychology Review*, *26*, 857-875. doi: 10.1016/j.cpr.2005.05.010

Gawronski, B., Balas, R., & Creighton, L. (2014). Can the formation of conditioned attitudes be intentionally controlled? *Personality & Social Psychology Bulletin*, *40*, 419–432. doi: 10.1177/0146167213513907

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692–731. doi: 10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model. Theory, evidence, and open questions. *Advances in Experimental Social Psychology*, *44*, 59-127. doi: 10.1016/B978-0-12-385522-0.00002-0

Gawronski, B., & Bodenhausen, G. V. (2014). Implicit and explicit evaluation: A brief review of the associative-propositional evaluation model. *Social and Personality Psychology Compass*, *8*, 448–462. doi: 10.1111/spc3.v8.8

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*, 1355-1361. doi: 10.1016/j.jesp.2008.04.005

Gawronski, B., & Walther, E. (2012). What do memory data tell us about the role of contingency awareness in evaluative conditioning? *Journal of Experimental Social Psychology, 48*, 617-623. doi: 10.1016/j.jesp.2012.01.002

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*, 558–565. doi: 10.3758/BF03196313

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*, 1464–80. doi: 10.1037/0022-3514.74.6.1464

Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review, 93*, 411-428.

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

Houben, K., Havermans, R., & Wiers, R. W. (2010). Learning to dislike alcohol: Conditioning negative implicit attitudes towards alcohol and its effect on drinking behavior. *Psychopharmacology, 211*, 79-86. doi: 10.1007/s00213-010-1872-1

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). the Dominance of Associative Theorizing in Implicit Attitude Research: Propositional and Behavioral Alternatives. *Psychological Record*, *61*, 465–496.

Huijding, J., Muris, P., Lester, K. J., Field, A. P., & Joosse, G. (2011). Training children to approach or avoid novel animals: Effects on self-reported attitudes and fear beliefs and information-seeking behaviors. *Behaviour Research and Therapy, 49*, 606-613. doi: 10.1016/j.brat.2011.06.005

Hütter, M., Sweldens, S., Stahl, C., Unkelbach, C., & Klauer, K. C. (2012). Dissociating contingency awareness and conditioned attitudes: Evidence of contingency-unaware evaluative conditioning. *Journal of Experimental Psychology: General, 141*, 539–557. doi: 10.1037/a0026477

Jones, C. R., Vilensky, M. R., Vasey, M. W., & Fazio, R. H. (2013). Approach behavior can mitigate predominately univalent negative attitudes: Evidence regarding insects and spiders. *Emotion, 13*, 989-996. doi: 10.1037/a0033164

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder:Improving implicit racial attitudes and interracial interactions through approach behaviors. *Journal of Personality and Social Psychology*, *92*, 957–971. doi: 10.1037/0022-3514.92.6.957

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberative judgements are based on common principles. *Psychological Review, 118*, 97–109.

Laham, S. M., Kashima, Y., Dix, J., & Wheeler, M. (2015). A meta-analysis of the facilitation of arm flexion and extension movements as a function of stimulus valence. *Cognition and Emotion, 29*, 1069-1090. doi: 10.1080/02699931.2014.968096

Lester, K. J., Lisk, S. C., Mikita, N., Mitchell, S., Huijding, J., Rinck, M. & Field, A. P. (2015). The effects of verbal information and approach-avoidance training on children's fear-related responses. *Journal of Behavior Therapy and Experimental Psychiatry, 48*, 40-49.

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in pavlovian conditioning: Empirical evidence and theoretical implications. J*ournal of Experimental Psychology: Animal Behavior Processes, 28*, 3-26.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *The Behavioral and Brain Sciences*, *32*, 183–198. doi: 10.1017/s0140525x09000855

Moors, A., & De Houwer, J. (2006). Automaticity: A Theoretical and Conceptual Analysis. *Psychological Bulletin*, *132*, 297–326. doi: 10.1037/0033-2909.132.2.297

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19*, 625-666. doi: 10.1521/soco.19.6.625.20886

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266-300.

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*, 277-293. doi: 10.1037/0022-3514.89.3.277

Phaf, R. H., Mohr, S. E., Rotteveel, M., & Wicherts, J. M. (2014). Approach, avoidance, and affect: a meta-analysis of approach-avoidance tendencies in manual reaction time tasks. *Frontiers in Psychology, 5*, 1–16. doi:10.3389/fpsyg.2014.00378

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. doi: 10.1037/a0022159

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*, 995–1008. doi: 10.1037/0022-3514.91.6.995

Schumacher, S., Kemps, E., & Tiggeman, M. (2016). Bias modification training can alter approach bias and chocolate consumption. *Appetite, 96*, 219-224. doi: 10.1016/j.appet.2015.09.014

Shanks, D. R. (2007). Associationism and cognition: human contingency learning at 25. *Quarterly Journal of Experimental Psychology*, *60*, 291–309. doi: 10.1080/17470210601000581

Sharbanee J. M., Hu, L., Stritzke, W. G. K., Wiers, R. W., Rinck, M., MacLeod, C. (2014) The Effect of Approach/Avoidance Training on Alcohol Consumption Is Mediated by change in Alcohol Action Tendency. *PLoS ONE 9*(1): e85855. doi:10.1371/journal.pone.0085855

Smith, E. R., & DeCoster, J. (2000). Dual process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4,* 108-131. doi: 10.1207/S15327957PSPR0402_01

Snelleman, M., Schoenmakers, T. M., de Mheen, D. V. (2015). Attentional Bias and Approach/Avoidance Tendencies Do Not Predict Relapse or Time to Relapse in Alcohol Dependency. *Alcoholism-Clinical and Experimental Research, 9*, 1734-1739. doi: 10.1111/acer.12817

Spruyt, A., De Houwer, J., Tibboel, H., Verschuere, B., Crombez, G., Verbanck, P., Hanak, C., Brevers, D., & Noël, X. (2013). On the predictive validity of automatically activated approach/avoidance tendencies in abstaining alcohol-dependent patients. *Drug and Alcohol Dependence, 127*, 81-86. doi: 10.1016/j.drugalcdep.2012.06.019

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*, 220-247. doi: 10.1207/s15327957pspr0803_1

Sweldens, S., Corneille, O. & Yzerbyt, V. (2014). The Role of Awareness in Attitude Formation Through Evaluative Conditioning. *Personality and Social Psychology Review, 18*, 187-209. doi: http://dx.doi.org/10.1177/1088868314527832

Uleman, J. S., Saribay, S. A., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology, 59*, 329-360. doi: 10.1146/annurev.psych.59.103006.093707

Vandenbosch, K., & De Houwer, J. (2011). Failures to induce implicit evaluations by means of approach–avoid training. *Cognition & Emotion*, *25*, 1311–1330. doi: 10.1080/02699931.2011.596819

Whitfield, M., & Jordan, C. H. (2009). Mutual influence of implicit and explicit attitudes. *Journal of Experimental Social Psychology*, *45*, 748–759. doi: 10.1016/j.jesp.2009.04.006

Wiers, R.W., Eberl, C., Rinck, M., Becker, E. & Lindenmeyer, J. (2011). Re-training automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science, 22,* 490-497. doi: 10.1177/0956797611400615

Wittekind, C. E., Feist, A., Schneider, B. C., Moritz, S., & Fritzsche, A. (2015). The approach-avoidance task as an online intervention in cigarette smoking: A pilot study. *Journal of Behavior Therapy and Experimental Psychiatry*, *46*, 115–120. doi: 10.1016/j.jbtep.2014.08.006

Woud, M. L., Becker, E. S., & Rinck, M. (2008). Implicit evaluation bias induced by approach and avoidance. *Cognition & Emotion*, *22*, 1187–1197. doi: 10.1080/02699930801906843

Woud, M. L., Becker, E. S., Lange, W. G., & Rinck, M. (2013). To like or not to like: The effect of approach-avoidance training on neutral, angry and smiling face stimuli. Psychological Reports, 113, 199-216. doi: 10.2466/21.07.PR0.113x10z1

Woud, M. L., Maas, J., Becker, E. S., & Rinck, M. (2013). Make the manikin move: Symbolic approach–avoidance responses affect implicit and explicit face evaluations. *Journal of Cognitive Psychology*, *25*, 738–744. doi: 10.1080/20445911.2013.817413

NEDERLANDSTALIGE SAMENVATTING

Gedrag wordt in sterke mate bepaald door preferenties (Allport, 1935). Zo zal iemand bijvoorbeeld sneller een stimulus benaderen die men positief evalueert (e.g., een lekker koekje) en een stimulus vermijden die men negatief evalueert (e.g., een gevaarlijke spin). De studie van evaluatie neemt dan ook een belangrijke rol in binnen psychologisch onderzoek. Evaluatie kan worden gedefinieerd als de invloed van stimuli op evaluatieve responsen (De Houwer, Gawronski, & Barnes-Holmes, 2013). Onder bepaalde omstandigheden zullen stimuli op spontane wijze een evaluatieve respons uitlokken (Zajonc, 1980). Dergelijke evaluaties noemen we impliciete evaluaties (De Houwer, 2009a). Ze worden gemeten met behulp van zogenaamde impliciete maten van evaluatie zoals de Impliciete Associatie Test (IAT; Greenwald, McGhee, & Schwartz, 1998). Evaluaties die op een meer opzettelijke en gecontroleerde manier ontstaan (e.g., iemands antwoord op de vraag in welke mate men een bepaalde stimulus aangenaam vindt) noemen we expliciete evaluaties.

Onderzoek toont aan dat impliciete evaluaties een belangrijke determinant zijn van bepaalde (spontane) gedragingen (zie Gawronski & Payne, 2010 voor een overzicht). Recent werden dan ook specifieke methoden ontwikkeld om impliciete evaluaties te vormen en te veranderen. Deze methoden bestonden hoofdzakelijk uit het herhaaldelijk samen aanbieden van een stimulus met een andere, positieve of negatieve, stimulus (evaluatieve conditionering: Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010) of met positieve of negatieve acties (benaderings- en vermijdingstraining: Kawakami, Phills, Steele, & Dovidio, 2007). Deze methodes zijn gebaseerd op het idee dat impliciete evaluaties veroorzaakt worden door de automatische activatie van associaties tussen representaties in het geheugen. Traditionele theorieën van evaluatie veronderstellen dat dergelijke associaties automatisch worden gevormd wanneer twee gebeurtenissen (herhaaldelijk) samen voorkomen (Strack & Deutsch, 2004; Rydell & McConnell, 2006; Gawronski & Bodenhausen, 2006).

Recent echter werden alternatieve evaluatietheorieën voorgesteld die veronderstellen dat impliciete evaluaties het gevolg zijn van propositionele processen zoals de vorming en activatie van proposities (De Houwer, 2009b, 2014; Mitchell, De Houwer, & Lovibond, 2009). Proposities verschillen van associaties omdat ze relationele informatie kunnen bevatten (e.g., 'gebeurtenis A WORDT VEROORZAAKT DOOR gebeurtenis B'). Recente studies bieden evidentie voor het idee dat propositionele processen een rol spelen bij impliciete evaluatie. Ten eerste tonen recente studies aan dat impliciete evaluaties kunnen veranderen wanneer proefpersonen via instructies propositionele informatie verkrijgen over de evaluatieve stimuli (Castelli, Zogmaister, Smith, & Arcuri, 2003; Gregg, Seibt, & Banaji, 2006). Dit gaat rechtstreeks in tegen de veronderstelling van traditionele evaluatietheorieën dat impliciete evaluaties enkel kunnen veranderen wanneer associaties geleidelijk aan worden gevormd aan de hand van herhaaldelijke paringen (Rydell & McConnell, 2006; Smith & DeCoster, 2000). Ten tweede is er evidentie dat herhaaldelijke paringen niet automatisch leiden tot een verandering in impliciete evaluaties. Hoewel verscheidene studies vonden dat impliciete evaluaties van stimuli positiever worden na herhaaldelijke benadering van de stimuli en negatiever na herhaaldelijke vermijding van stimuli is dit niet steeds het geval (Becker, Jostmann, Wiers, & Holland, 2015; Vandenbosch & De Houwer, 2011). Dit gaat in tegen het idee dat het uitvoeren van deze responsen automatisch leidt tot veranderingen in de associaties die impliciete evaluaties onderliggen (Phills, Kawakami, Tabi, Nadolny, & Inzlicht, 2011). Effecten van benaderings- en vermijdingstraining lijken afhankelijk te zijn van omstandigheden die tot op heden nog onbekend zijn (Vandenbosch & De Houwer).

HUIDIGE PHD THESIS

Het doel van deze thesis was om meer inzicht te krijgen in de omstandigheden waaronder benaderings- en vermijdingstraining leidt tot veranderingen in impliciete evaluaties en de mentale processen die deze effecten onderliggen. Specifiek gingen we na in welke mate deze effecten afhankelijk zijn van propositionele processen die gebaseerd zijn op de bewuste verwerving of acquisitie van propositionele informatie. Enerzijds gingen we na of het verschaffen van propositionele informatie via instructies over het samen voorkomen van stimuli en benaderings- of vermijdingsacties kan leiden tot veranderingen in impliciete evaluaties. Anderzijds onderzochten we of de effecten van benaderings- en vermijdingstraining op impliciete evaluaties afhankelijk zijn van de mate waarin men zich bewust is van de contingenties tussen de stimuli en de uitgevoerde acties.

Onderzoekslijn 1: Effecten van benaderings- en vermijdingsinstructies op impliciete evaluatie

In **Hoofdstuk 2** onderzochten we of veranderingen in impliciete evaluaties ook kunnen worden geobserveerd wanneer proefpersonen geen benaderings- of vermijdingsacties uitvoeren maar enkel instructies krijgen over de contingenties tussen stimuli en deze acties. We voerden 5 experimenten uit om deze vraag te beantwoorden. In elk experiment kregen proefpersonen instructies om één stimulus te benaderen en een andere stimulus te vermijden. Ze werden gevraagd om deze informatie goed te onthouden aangezien ze eerst een andere taak zouden moeten uitvoeren. Meteen na deze instructies werden de impliciete evaluaties van de stimuli gemeten. We observeerden dat benaderings- en vermijdingsinstructies impliciete evaluaties kunnen beïnvloeden. Proefpersonen vertoonden meer positieve evaluaties van de stimuli die ze volgens de instructies zouden benaderen dan van de stimuli die ze zouden vermijden. Deze effecten bleken echter afhankelijk van de aard van de stimulus die werd gebruikt. Hoewel benaderings- en vermijdingsinstructies een sterke invloed hadden op evaluaties van nieuwe, onbekende stimuli (i.e., fictieve sociale groepen of niet-bestaande woorden), vonden we geen effecten op evaluaties van bekende stimuli (i.e., vertrouwde sociale groepen). We vonden eveneens dat benaderings- en vermijdingsinstructies een invloed hadden op expliciete evaluaties van fictieve sociale groepen maar niet van vertrouwde sociale groepen.

Gebaseerd op deze resultaten concludeerden we dat veranderingen in impliciete evaluaties niet noodzakelijk het gevolg zijn van uitgebreide benaderings- en vermijdingstraining maar ook kunnen voorkomen als gevolg van instructies over de relatie tussen stimuli en toekomstige benaderings- en vermijdingsacties. Deze resultaten bieden evidentie voor propositionele evaluatietheorieën die veronderstellen dat propositionele processen een belangrijke rol spelen in het leren van impliciete evaluaties (De Houwer, 2014). Echter, recente duale-proces modellen van evaluatie zoals het associatief-propositionele evaluatie model (APE model; Gawronski & Bodenhausen, 2006) kunnen deze resultaten eveneens verklaren. Volgens dit model zouden impliciete evaluaties van onbekende stimuli kunnen worden veranderd als gevolg van de acquisitie van propositionele informatie indien er eerst een verandering is in expliciete evaluaties. Deze verandering zou er namelijk voor zorgen dat er nieuwe associaties worden gevormd die een verandering in impliciete evaluaties voor onbekende stimuli kan veroorzaken (Whitfield & Jordan, 2009). Echter, deze verklaring is niet in overeenstemming met de observatie dat veranderingen in impliciete evaluaties als gevolg van benaderings- en vermijdingsinstructies niet volledig gemedieerd waren door veranderingen in expliciete evaluaties (in Experiment 4 en 5). We onderzochten dit verder in het volgende hoofdstuk.

In **Hoofdstuk 3** onderzochten we de mediatie van het effect van benaderings- en vermijdingsinstructies op impliciete evaluaties via veranderingen in expliciete evaluaties. In deze twee experimenten kregen proefpersonen instructies om de namen van leden van fictieve sociale groepen te benaderen of te vermijden. Voorafgaand aan deze instructies werd informatie verschaft over de karaktertrekken van de leden van deze sociale groepen (e.g., ‘Niffieten zijn vredelievend, beschaafd en behulpzaam en Luupieten zijn agressief, woest en kwaadaardig’). We observeerden dat het verschaffen van deze informatie de effecten van benaderings- en vermijdingsinstructies teniet deed voor expliciete maar niet voor impliciete evaluaties. Proefpersonen vertoonden een grotere impliciete voorkeur voor de groep die ze zouden benaderen dan voor de groep die ze zouden vermijden ondanks het feit dat ze informatie ter beschikking hadden die meer diagnostisch was over de valentie van de stimuli. Dit resultatenpatroon werd eveneens gevonden wanneer de benaderings- en vermijdingsinstructies werden aangevuld met benaderings- en vermijdingstraining. Statistische mediatie analyses toonden verder aan dat zowel benaderings- en vermijdingsinstructies als benaderings- en vermijdingstraining een directe invloed hadden op impliciete evaluaties (i.e., niet gemedieerd door veranderingen in expliciete evaluaties).

Deze resultaten tonen aan dat er belangrijke gelijkenissen zijn tussen benaderings- en vermijdingseffecten die gebaseerd zijn op instructies en op training. Beide procedures lijken een invloed te hebben op impliciete evaluaties zelfs indien proefpersonen niet de intentie hebben om de instructies of training hun (expliciete) evaluaties te laten beïnvloeden. De processen die onderliggen aan instructie-effecten lijken dus, net als de effecten van training, een automatische invloed te hebben op impliciete evaluaties. Deze resultaten zijn in overeenstemming met het idee dat de acquisitie en activatie van proposities impliciete evaluaties op automatische wijze kan beïnvloeden (De Houwer, 2014). Bijvoorbeeld, benaderings- en vermijdingsinstructies zouden ertoe kunnen leiden dat men de inferentie maakt dat de te-benaderen stimulus positief is en de te-vermijden stimulus negatief. De automatische (e.g., niet-intentionele) activatie van deze nieuwe propositionele informatie zou iemands prestaties in impliciete evaluatietaken kunnen beïnvloeden.

Echter, de huidige bevindingen bewijzen niet dat de effecten van benaderings- en vermijdingstraining (altijd) het resultaat zijn van propositionele processen. In de huidige studies kregen proefpersonen steeds informatie over de contingenties tussen acties en stimuli. Een belangrijke vraag is of de effecten van benaderings- en vermijdingstraining op impliciete evaluatie ook afhankelijk zijn van propositionele processen wanneer men geen informatie krijgt over deze contingenties via instructies. Op deze vraag zochten we een antwoord in de volgende hoofdstukken.

Onderzoekslijn 2: De rol van contingentie bewustzijn in de effecten van benaderings- en vermijdingstraining

Het doel van **Hoofdstuk 4** was om na te gaan of de effecten van benaderings- en vermijdingstraining afhankelijk zijn van de mate waarin proefpersonen zich bewust zijn van de contingenties tussen stimuli en benaderings- en vermijdingsacties. Indien deze effecten het resultaat zijn van propositionele processen dan zouden ze afhankelijk moeten zijn van de mate waarin proefpersonen de mogelijkheid hebben om propositionele contingentie informatie te verkrijgen. Dit zou (ten minste gedeeltelijk) gereflecteerd moeten zijn in contingentie bewustzijn. In deze studies gebruikten we een variant van de procedure die geïntroduceerd werd door Woud, Becker, en Rinck (2008) om effecten van benaderings-en vermijdingstraining op impliciete evaluaties van onbekende gezichten na te gaan. Hierbij benaderen of vermijden proefpersonen herhaaldelijk bepaalde foto's van gezichten op basis van een kenmerk van de stimulus (e.g., de kleur van het kader dat de foto's omringt) dat niet gerelateerd is aan de identiteit van de gezichten. Hierna worden impliciete evaluaties van de gezichten gemeten. Voorgaande studies die deze procedure gebruikten vonden gemengde resultaten. Hoewel Woud en collega's een significant effect rapporteerden van de benaderings- en vermijdingstraining op impliciete evaluaties van de gezichten, vonden Vandenbosch en De Houwer (2011) geen enkele evidentie voor een dergelijk effect in vijf experimenten en konden ze de effecten van Woud en collega's niet reproduceren in een her-analyse van hun data. Voor de huidige studies hebben we de procedure aangepast zodat de acquisitie van informatie over stimulus-actie contingenties makkelijker zou worden. We gebruikten bijvoorbeeld een kleiner aantal evaluatieve stimuli (i.e., foto's van 8 in plaats van 12 gezichten). Tevens maakten we aanpassingen zodat het belang van contingentie bewustzijn kon worden nagegaan. Ten eerste, probeerden we bewustzijn van de ervaren contingenties na te gaan door te meten in welke mate proefpersonen de relatie tussen gezichten en acties konden rapporteren. Ten tweede, kregen de helft van de proefpersonen expliciete informatie over de contingenties en gingen we het effect hiervan na op evaluaties.

De resultaten van Experiment 1 toonden aan dat contingentie bewustzijn de effecten van benaderings- en vermijdingstraining op impliciete (en expliciete evaluaties) van gezichten modereert. Proefpersonen vertoonden enkel een voorkeur voor benaderde gezichten over vermeden gezichten indien ze voor deze gezichten correct konden aangeven of ze deze benaderd of vermeden hadden. Onze manipulatie van contingentie bewustzijn via instructies bood geen evidentie dat contingentie bewustzijn een invloed had op benaderings- en vermijdingsinstructies. Echter, onze voorgaande studies toonden wel een dergelijk effect van instructies aan. Het niet vinden van dit effect in Experiment 1 is dan ook waarschijnlijk te wijten aan beperkte statistische power. In Experiment 2 werd dezelfde procedure gebruikt als in Experiment 1 met uitzondering dat geen enkele van de proefpersonen informatie kreeg over de contingenties. Resultaten toonden aan dat benaderings- en vermijdingstraining significante verschillen veroorzaakte in impliciete en expliciete evaluaties, en dat deze effecten gemodereerd werden door contingentie bewustzijn.

Deze resultaten bieden evidentie voor het idee dat benaderings- en vermijdingstraining enkel leidt tot veranderingen in evaluaties indien proefpersonen propositionele kennis hebben over de relatie tussen actie en stimulus. De bevindingen zijn in lijn met onze voorgaande bevindingen dat de acquisitie van contingentie informatie via instructies veranderingen in impliciete (en expliciete) evaluaties kan veroorzaken. Ze bieden sterke evidentie dat propositionele processen een belangrijke rol spelen in benaderings- en vermijdingseffecten. De resultaten zijn niet in overeenstemming met het idee dat de automatische vorming van associaties effecten van benaderings- en vermijdingstraining onderligt (Phills et al., 2011; Woud et al., 2008). Men kan deze resultaten wel in overeenstemming brengen met associatieve verklaringen voor benaderings- en vermijdingstraining effecten indien men ervan uitgaat dat bepaalde procedurele details van deze studie ervoor zorgden dat een automatisch effect als gevolg van associatieve processen werd verhinderd. Hoewel het onduidelijk is wat deze procedurele details zouden kunnen zijn, is er enige evidentie voor dit idee. Namelijk, sommige studies geven aan dat benaderings- en vermijdingstraining, onder bepaalde condities, kan leiden tot effecten op impliciete evaluaties in de afwezigheid van contingentie bewustzijn (i.e., bij subliminale presentatie van de getrainde stimuli: Kawakami et al., 2007; Jones, Vilensky, Vasey, & Fazio, 2013). We onderzochten dit verder in het volgende hoofdstuk.

In **hoofdstuk 5** onderzochten we of benaderings- en vermijdingstraining kan leiden tot veranderingen in impliciete evaluaties van stimuli die sub-liminaal (i.e., onder de perceptuele bewustzijnsdrempel) worden aangeboden. Experiment 1 was een exacte replicatie van het experiment van Kawakami en collega's (2007, Experiment 2) dat evidentie bood dat subliminale benaderings- en vermijdingstraining een invloed kan hebben op impliciete evaluaties van raciale groepen. We konden dit resultaat niet repliceren. In tegenstelling, Bayesiaanse analyses van onze data boden substantiële evidentie dat training om subliminaal gepresenteerde prenten van personen met zwarte huidskleur te benaderen en subliminaal gepresenteerde prenten van blanke personen te vermijden niet leidt tot veranderingen in evaluaties van deze groepen. Experiment 2 onderzocht de effecten van subliminale benaderings- en vermijdingstraining op evaluaties van onbekende gezichten. In dit experiment werden proefpersonen eerst geïnformeerd over de contingenties tussen vier gezichtsstimuli en benaderings- en vermijdingsacties in de subliminale trainingstaak alvorens ze deze taak uitvoerden. Echter, de contingentie-instructies boden slechts correcte informatie voor twee van de vier stimuli. De resultaten boden sterke evidentie dat de contingentie-instructies impliciete en expliciete evaluaties van de stimuli beïnvloeden maar niet de subliminale benaderings- en vermijdingstraining. Experiment 3 onderzocht de effecten van *sub*liminale en *supra*liminale benaderings- en vermijdingstraining op evaluaties van niet-bestaande woorden. Resultaten toonden aan dat supraliminale benaderings- en vermijdingstraining een invloed had op stimulus evaluaties maar dat deze effecten afhankelijk waren van contingentie bewustzijn. Daarentegen observeerden we geen effecten van subliminale benaderings- en vermijdingstraining.

CONCLUSIE

De huidige thesis had tot doel om onze kennis uit te breiden over de processen die onderliggen aan benaderings- en vermijdingseffecten op impliciete evaluaties. Onze resultaten bieden sterke evidentie dat propositionele processen zoals de acquisitie van propositionele contingentie informatie een belangrijke rol spelen in deze effecten. We vonden dat (1) de acquisitie van informatie over de contingenties tussen stimuli en benaderings- en vermijdingsacties via instructies (directe) veranderingen in impliciete evaluaties kan veroorzaken en (2) effecten van benaderings- en vermijdingstraining op impliciete evaluaties gemodereerd worden door bewustzijn van stimulus- actie contingenties. Deze resultaten zijn niet in overeenstemming met het idee dat deze effecten veroorzaakt worden door de automatische vorming en activatie van associaties in het brein. We hopen dat toekomstig onderzoek verder de rol van propositionele processen in (automatische effecten op) impliciete evaluatie zal onderzoeken.

REFERENTIES

Allport, G. (1935). Attitudes. In Murchison, C. (Ed.). *A Handbook of Social Psychology* (pp. 789-843). Worcester, MA: Clark University Press.

Becker, D., Jostmann, N. B., Wiers, R. W., & Holland, R. W. (2015). Approach avoidance training in the eating domain: Testing the effectiveness across three single session studies. *Appetite*, *85*, 58–65. doi: 10.1016/j.appet.2014.11.017.

Castelli, L., Zogmaister, C., Smith, E. R., & Arcuri, L. (2004). On the automatic evaluation of social exemplars. *Journal of Personality and Social Psychology*, *86*, 373–387. doi: 10.1037/0022-3514.86.3.373

De Houwer, J. (2014). A Propositional Model of Implicit Evaluation. *Social and Personality Psychology Compass, 8*, 342-353. doi: 10.1111/spc3.12111

De Houwer, J. (2009a). How do people evaluate objects? A brief review. *Social and Personality Psychology Compass, 3*, 36-48. doi: 10.1111/j.1751-9004.2008.00162.x

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

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology*, *24*, 252–287. doi: 10.1080/10463283.2014.892320

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*, 692–731. http://dx.doi.org/10.1037/0033-2909.132.5.692

Gawronski, B., & Payne, B. K. (2010.) *Handbook of implicit social cognition.* New York, NY: Guilford Press.

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*, 1464–80. doi: 10.1037/0022-3514.74.6.1464

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,* 1-20. doi: 10.1037/0022-3514.90.1.1

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

Jones, C. R., Vilensky, M. R., Vasey, M. W., & Fazio, R. H. (2013). Approach behavior can mitigate predominately univalent negative attitudes: Evidence regarding insects and spiders. *Emotion, 13*, 989-996. doi: 10.1037/a0033164

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder:Improving implicit racial attitudes and interracial interactions through approach behaviors. *Journal of Personality and Social Psychology*, *92*, 957–971. doi: 10.1037/0022-3514.92.6.957

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

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. doi: 10.1037/a0022159

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*, 995- 1008. doi: 10.1037/0022-3514.91.6.995

Smith, E. R., & DeCoster, J. (2000). Dual-Process Models in Social and Cognitive Psychology: Conceptual Integration and Links to Underlying Memory Systems. *Personality and Social Psychology Review*, *4*, 108–131. doi: 10.1207/S15327957PSPR0402_01

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*, 220-247. doi: 10.1207/s15327957pspr0803_1

Vandenbosch, K., & De Houwer, J. (2011). Failures to induce implicit evaluations by means of approach–avoid training. *Cognition & Emotion*, *25*, 1311–1330. doi: 10.1080/02699931.2011.596819

Whitfield, M., & Jordan, C. H. (2009). Mutual influence of implicit and explicit attitudes. *Journal of Experimental Social Psychology*, *45*, 748–759. doi: 10.1016/j.jesp.2009.04.006

Woud, M. L., Becker, E. S., & Rinck, M. (2008). Implicit evaluation bias induced by approach and avoidance. *Cognition & Emotion*, *22*, 1187–1197. doi: 10.1080/02699930801906843

Zajonc, R. B. (1980). Feeling and thinking. *American Psychologist*, *35*, 151–175. doi: 10.1037/0003-066X.35.2.151

DATA STORAGE FACT SHEETS

% Data Storage Fact Sheet

% Name/identifier study: Chapter 2: Instruction-Based Approach-Avoidance Effects

% Author: Pieter Van Dessel, Jan De Houwer, Anne Gast, Colin Tucker Smith

% Date: 12 March 2015

1. Contact details
==

1a. Main researcher

- name: Pieter Van Dessel
- address: Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium
- e-mail: Pieter.VanDessel@UGent.be

1b. Responsible Staff Member (ZAP)

- name: Jan De Houwer
- address: Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium
- e-mail: Jan.DeHouwer@UGent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies
==

* Reference of the publication in which the datasets are reported:

Van Dessel, P., De Houwer, J., Gast, A., & Smith, C.T. (2015). Instruction-Based Approach-Avoidance Effects: Changing Stimulus Evaluation via the Mere Instruction to Approach or Avoid Stimuli. *Experimental Psychology, 62*, 161–169. doi: 10.1027/1618-3169/a000282

* Which datasets in that publication does this sheet apply to?:

All.

3. Information about the files that have been stored
==

3a. Raw data
----

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO

If NO, please justify:

* On which platform are the raw data stored?
  - [X] researcher PC
  - [X] research group file server
  - [ ] other (specify): ...

* Who has direct access to the raw data (i.e., without intervention of another person)?
  - [X] main researcher
  - [X] responsible ZAP
  - [X] all members of the research group
  - [ ] all members of UGent
  - [ ] other (specify): ...

3b. Other files
----

* Which other files have been stored?

  - [X] file(s) describing the transition from raw data to reported results. Specify: ...
  - [X] file(s) containing processed data. Specify: ...
  - [X] file(s) containing analyses. Specify: ...
  - [ ] files(s) containing information about informed consent
  - [ ] a file specifying legal and ethical provisions
  - [X] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
  - [ ] other files. Specify: ...

* On which platform are these other files stored?

  - [X] individual PC
  - [X] research group file server
  - [ ] other: ...

* Who has direct access to these other files (i.e., without intervention of another person)?

  - [X] main researcher
  - [X] responsible ZAP
  - [X] all members of the research group
  - [ ] all members of UGent
  - [ ] other (specify): ...

4. Reproduction
===

* Have the results been reproduced independently?: [X] YES (Experiments 4+5) / [ ] NO

* If yes, by whom (add if multiple):

  - name: Colin Tucker Smith
  - address: 945 Center Drive, Gainesville, FL 32611-2250, VSA
  - affiliation: Department of Psychology, University of Florida
  - e-mail: colinsmith@ufl.edu

v0.2

% Data Storage Fact Sheet

% Name/identifier study: Chapter 3: Instructing Implicit Processes: When Instructions to Approach or Avoid Influence Implicit but not Explicit Evaluation

% Author: Pieter Van Dessel, Jan De Houwer, Anne Gast, Colin Tucker Smith, Maarten De Schryver

% Date: 1 December 2015

1. Contact details
===

1a. Main researcher

- name: Pieter Van Dessel
- address: Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium
- e-mail: Pieter.VanDessel@UGent.be

1b. Responsible Staff Member (ZAP)

- name: Jan De Houwer
- address: Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium
- e-mail: Jan.DeHouwer@UGent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies
===

* Reference of the publication in which the datasets are reported:

Van Dessel, P., De Houwer, J., Gast, A., Smith, C. T., & De Schryver, M. (2016). Instructing Implicit Processes: When Instructions to Approach or Avoid Influence Implicit but not Explicit Evaluation. *Journal of Experimental Social Psychology, 63*, 1-9. doi: 10.1016/j.jesp.2015.11.002

* Which datasets in that publication does this sheet apply to?:

All.

3. Information about the files that have been stored
===

3a. Raw data
----

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO
If NO, please justify:

* On which platform are the raw data stored?
  - [X] researcher PC
  - [X] research group file server
  - [ ] other (specify): ...

* Who has direct access to the raw data (i.e., without intervention of another person)?

  - [X] main researcher
  - [X] responsible ZAP
  - [X] all members of the research group
  - [ ] all members of UGent
  - [ ] other (specify): ...

3b. Other files
----

* Which other files have been stored?

  - [X] file(s) describing the transition from raw data to reported results. Specify: ...
  - [X] file(s) containing processed data. Specify: ...
  - [X] file(s) containing analyses. Specify: ...
  - [ ] files(s) containing information about informed consent
  - [ ] a file specifying legal and ethical provisions
  - [X] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
  - [ ] other files. Specify: ...

* On which platform are these other files stored?

  - [X] individual PC
  - [X] research group file server
  - [ ] other: ...

* Who has direct access to these other files (i.e., without intervention of another person)?

  - [X] main researcher
  - [X] responsible ZAP
  - [X] all members of the research group
  - [ ] all members of UGent
  - [ ] other (specify): ...

4. Reproduction
===

* Have the results been reproduced independently?: [X] YES / [ ] NO

* If yes, by whom (add if multiple):

   - name: Maarten De Schryver
   - address: Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium
   - affiliation: Department of Experimentel-Clinical and Health psychology, Ghent University
   - e-mail: Maarten.DeSchryver@UGent.be

v0.2

% Data Storage Fact Sheet

% Name/identifier study: Chapter 4: Approach-Avoidance Training effects are

Moderated by Awareness of Stimulus-Action Contingencies

% Author: Pieter Van Dessel, Jan De Houwer, Anne Gast

% Date: 1 February 2016

1. Contact details
==

1a. Main researcher

- name: Pieter Van Dessel
- address: Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium
- e-mail: Pieter.VanDessel@UGent.be

1b. Responsible Staff Member (ZAP)

- name: Jan De Houwer
- address: Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium
- e-mail: Jan.DeHouwer@UGent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies
==

* Reference of the publication in which the datasets are reported:

Van Dessel, P., De Houwer, J., & Gast, A. (2016). Approach-Avoidance Training effects are Moderated by Awareness of Stimulus-Action Contingencies. *Personality and Social Psychology Bulletin, 42*, 81-93. doi: 10.1177/0146167215615335

* Which datasets in that publication does this sheet apply to?:

All.

3. Information about the files that have been stored
==

3a. Raw data
----

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO
If NO, please justify:

* On which platform are the raw data stored?

  - [X] researcher PC
  - [X] research group file server
  - [ ] other (specify): ...

* Who has direct access to the raw data (i.e., without intervention of another person)?

  - [X] main researcher

- [X] responsible ZAP
- [X] all members of the research group
- [ ] all members of UGent
- [ ] other (specify): ...

3b. Other files
----

* Which other files have been stored?

- [X] file(s) describing the transition from raw data to reported results. Specify: ...
- [X] file(s) containing processed data. Specify: ...
- [X] file(s) containing analyses. Specify: ...
- [ ] files(s) containing information about informed consent
- [ ] a file specifying legal and ethical provisions
- [X] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
- [ ] other files. Specify: ...

* On which platform are these other files stored?

- [X] individual PC
- [X] research group file server
- [ ] other: ...

* Who has direct access to these other files (i.e., without intervention of another person)?

- [X] main researcher
- [X] responsible ZAP
- [X] all members of the research group
- [ ] all members of UGent
- [ ] other (specify): ...

4. Reproduction
==

* Have the results been reproduced independently?: [X] YES / [ ] NO

* If yes, by whom (add if multiple):

- name: Jan Lammertyn
- address: Ghent University, Henri Dunantlaan 1, B-9000 Ghent, Belgium
- affiliation: Ghent University
- e-mail: Jan.Lammertyn@UGent.be

v0.2

% Data Storage Fact Sheet

% Name/identifier study: Chapter 5: Failures to Change Stimulus Evaluations by means of Subliminal Approach and Avoidance Training

% Author: Pieter Van Dessel, Jan De Houwer, Arne Roets, Anne Gast

% Date: 1 February 2016

1. Contact details
==

1a. Main researcher

- name: Pieter Van Dessel
- address: Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium
- e-mail: Pieter.VanDessel@UGent.be

1b. Responsible Staff Member (ZAP)

- name: Jan De Houwer
- address: Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium
- e-mail: Jan.DeHouwer@UGent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies
==

* Reference of the publication in which the datasets are reported:

Van Dessel, P., De Houwer, J., Roets, A., & Gast, A. (2016). Failures to Change Stimulus Evaluations by means of Subliminal Approach and Avoidance Training. *Journal of Personality and Social Psychology, 110*, e1-e15. doi: 10.1037/pspa0000039

* Which datasets in that publication does this sheet apply to?:

All.

3. Information about the files that have been stored
==

3a. Raw data
----

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO
If NO, please justify:

* On which platform are the raw data stored?

  - [X] researcher PC
  - [X] research group file server
  - [ ] other (specify): ...

* Who has direct access to the raw data (i.e., without intervention of another person)?

  - [X] main researcher

- [X] responsible ZAP
- [X] all members of the research group
- [ ] all members of UGent
- [ ] other (specify): ...

3b. Other files
----

* Which other files have been stored?

- [X] file(s) describing the transition from raw data to reported results. Specify: ...
- [X] file(s) containing processed data. Specify: ...
- [X] file(s) containing analyses. Specify: ...
- [ ] files(s) containing information about informed consent
- [ ] a file specifying legal and ethical provisions
- [X] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
- [ ] other files. Specify: ...

* On which platform are these other files stored?

- [X] individual PC
- [X] research group file server
- [ ] other: ...

* Who has direct access to these other files (i.e., without intervention of another person)?

- [X] main researcher
- [X] responsible ZAP
- [X] all members of the research group
- [ ] all members of UGent
- [ ] other (specify): ...

4. Reproduction
===

* Have the results been reproduced independently?: [] YES / [X] NO

* If yes, by whom (add if multiple):

- name:
- address:
- affiliation:
- e-mail:

v0.2