Journal of Experimental Psychology: General
2017, Vol. 146, No. 2, 194–213

0096-3445/17/$12.00 http://dx.doi.org/10.1037/xge0000239

Repeated Evaluative Pairings and Evaluative Statements: How Effectively Do They Shift Implicit Attitudes?

Benedek Kurdi and Mahzarin R. Banaji
Harvard University

Six experiments, involving a total of 6,492 participants, were conducted to investigate the relative effectiveness of *repeated evaluative pairings* (REP; exposure to category members paired with pleasant or unpleasant images), *evaluative statements* (ES; verbally signaling upcoming pairings without actual exposure), and their combination (ES + REP) in shifting implicit social and nonsocial attitudes. Learning modality (REP, ES, and ES + REP) was varied between participants and implicit attitudes were assessed using an Implicit Association Test (IAT). Study 1 (N = 675) used fictitious social groups (NIFFs and LAAPs), Study 2 (N = 1,034) used novel social groups (humans with long vs. square faces), Study 3 (N = 1,072) used nonsocial stimuli (squares vs. rectangles), and Study 4 (N = 848) and Study 5 (N = 958) used known social groups (young vs. elderly; American vs. foreign). ES were more effective than REP and no less superior than ES + REP in producing implicit attitude change. Results were robust across social and nonsocial domains and for known and novel groups. Study 6 (N = 1,905) eliminated time on intervention, levels of construal, and expectancy effects as possible explanations for these findings. Associative theories of implicit evaluation posit that implicit attitudes should shift piecemeal over time; yet, in these experiments, one-shot language-based learning led to larger shifts in implicit attitude than exposure to stimulus pairings. Moreover, the redundancy observed in REP + ES suggests that attitude acquisition from repeated pairings and evaluative instructions may rely on shared mental representations.

Keywords: associative learning, evaluative conditioning, Implicit Association Test, implicit attitudes, propositional learning

Learning about conspecifics is among the most fundamental and momentous tasks performed by humans. To decide whether a person is friend or foe, and whether to approach or avoid them, access to the products of such learning in memory is essential. Such decisions, moreover, often depend upon immediate responses in complex and rapidly shifting social environments. Accordingly, the process of social learning requires maintenance of a delicate balance between holding on to an initial or even well-learned evaluation and adaptively changing that evaluation when novel, and potentially contradictory, information becomes available. Learning and updating are, therefore, critical to achieving social accuracy and even to survival.

Benedek Kurdi and Mahzarin R. Banaji, Department of Psychology, Harvard University.

Some of the data reported in this article have been presented at the Fourth Annual Symposium for Boston-Area Graduate Students in Psychology in April 2015, the Learning and Implicit Processes Lab, Ghent University, in August 2015, the 17th Annual Meeting of the Society for Personality and Social Psychology in January 2016, and the Fifth European Meeting on the Psychology of Attitudes in July 2016. All raw data files, analysis scripts, and stimuli used in this article are available for download from the Open Science Framework (https://osf.io/jyk8c/). We thank Susan Carey, Jan De Houwer, Sam Gershman, and Jason Mitchell for their insightful comments on previous drafts of this article, Mark Thornton for statistical advice, and Sarah Ryan for assistance with study coding.

Correspondence concerning this article should be addressed to Benedek Kurdi, Department of Psychology, Harvard University, Cambridge, MA 02138, E-mail: kurdi@g.harvard.edu

Learning in the social world, at its core, involves the formation of attitudes, defined as "a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor" (Eagly & Chaiken, 1993, p. 1). Fazio and colleagues (Fazio, Jackson, Dunton, & Williams, 1995; Fazio, Sanbonmatsu, Powell, & Kardes, 1986), Bargh and colleagues (Bargh, 1989; Bargh, Chaiken, Govender, & Pratto, 1992), and Greenwald and Banaji (1995) focused attention on the implicit nature of social evaluation. In this framework, implicit attitudes are relatively less consciously accessible, less controllable, and more automatic than their explicit counterparts (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009; Greenwald & Banaji, 1995; Kim, 2003). Moreover, they uniquely guide behavior in a wide range of situations (Devine, 1989; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Greenwald, Poehlman, Uhlmann, & Banaji, 2009).

However, how are implicit attitudes acquired and how do they shift in the face of new information? Social attitudes and modalities of acquisition and change can be categorized in myriad ways. A recent distinction that has emerged is that between associative (Gawronski & Bodenhausen, 2006; Greenwald & Banaji, 1995; Rydell & McConnell, 2006) and propositional (De Houwer, 2009, 2014; C. J. Mitchell, De Houwer, & Lovibond, 2009) processes of learning and change. The present research relies on the distinction between associative and propositional learning to ask a basic question that is yet to be investigated directly: Which one of these two learning modalities is more effective in shifting implicit attitudes? Answering this question can teach us about the nature of attitudes in their most fundamental form and has the potential to

lead to method development for changing attitudes via theoretically guided interventions.

The nature of the mental representations underlying evaluative learning has been contested (De Houwer, 2009, 2014; Gawronski & Bodenhausen, 2006, 2011), and many proposals have emerged from this debate. We are in agreement with De Houwer (2009) that conceptual confusion may be avoided by characterizing methods of learning in terms of actual operations rather than using more abstract labels. Far too often, abstract psychological constructs, including the most studied and central ones (such as *perception* or *memory*), are used to refer to a host of diverse phenomena and can map onto a range of different operationalizations, thus obscuring meaningful, and possibly unintended, differences across possible instantiations. To avoid fuzzy thinking about the learning mechanisms introduced here, we created novel labels that hew closely to the actual learning manipulations that were used in the experiments.

The present research compares the separate and combined effects of two different forms of learning. The first learning procedure will be referred to as *repeated evaluative pairings* (*REP*), a learning modality commonly characterized as associative in nature. We define repeated evaluative pairings as "pairing [a] stimulus [the conditioned stimulus; CS] with other positive or negative stimuli [unconditioned stimuli; US]" (De Houwer, Thomas, & Baeyens, 2001). This learning modality, also known as evaluative conditioning (Baeyens & De Houwer, 1995; De Houwer et al., 2001; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010), is rooted in well-known associative learning procedures such as classical conditioning (Pavlov, 1927). Using this method, learning has been observed in numerous species, including goldfish, quail, hamsters, rats, dogs, cats, pigeons (Rescorla & Holland, 1982), and humans (Delgado, Olsson, & Phelps, 2006; C. J. Mitchell et al., 2009). In such paradigms, learning is assumed to occur via the transfer of valence between an intrinsically positive or negative stimulus and an initially neutral stimulus or action. Specifically, in evaluative conditioning, changes in the liking of a stimulus or its subjective positivity are thought to result from pairing that stimulus with other, positive or negative, stimuli (De Houwer et al., 2001).

A second learning procedure used in these experiments will be referred to as *evaluative statements* (*ES*), a learning modality commonly characterized as propositional in nature. Unlike nonhuman animals that lack the capacity for language or higher-order symbolic logic (Penn, Holyoak, & Povinelli, 2008), humans have the unique ability to learn about their environment solely through language.[1] For example, when one person says to another, "The Falaknuma Palace is the most beautiful hotel I have ever stayed in," the assumption is that the concept "Falaknuma Palace" is now linked to the positive trait "beautiful," and an attitude may form based on the recipient's cognitive response to the statement. In contrast with REP-based learning procedures, which apart from minor procedural variations are highly similar to each other, evaluative statements can be manifold. Classic work on attitudes used messages that varied in persuasive power to shift evaluations of varied attitude objects via high-level, relatively elaborative, reasoning processes (for reviews see Petty & Briñol, 2010; Wood, 2000). Other kinds of evaluative statements remain closer to REP-based learning in that they rely on mere signaling of upcoming repeated evaluative pairings. For instance, Olsson and Phelps (2004, 2007) demonstrated that human participants need not be exposed to actual shock for fear conditioning to occur; it is sufficient to simply inform participants of upcoming stimulus–shock pairings to obtain a learning effect of similar magnitude to presenting actual shock. The power of language alone, in the absence of direct experience, to create learning is intriguing and uniquely human.

Shifting Implicit Attitudes via Repeated Evaluative Pairings and Evaluative Statements

Olson and Fazio (2001) provided a clear demonstration of the effectiveness of *repeated evaluative pairings* (*REP*) in the domain of implicit attitude change. Olson and Fazio showed participants several hundred images, some of them in isolation and others in pairs. Critically, they embedded 40 unconditioned stimulus (US)–conditioned stimulus (CS) pairs among the stimuli. Two initially neutral Pokémon characters served as CS, and valenced words (e.g., "awesome" vs. "terrible") and images (e.g., of ice cream vs. a cockroach) served as US. At test, participants exhibited implicit preference in favor of the Pokémon character paired with positive US over the Pokémon character paired with negative US during the learning phase. The finding that REP influence implicit attitudes has been replicated numerous times (Gibson, 2008; Grumm, Nestler, & von Collani, 2009; C. J. Mitchell, Anderson, & Lovibond, 2003; Olson & Fazio, 2006; Prestwich, Perugini, Hurling, & Richetin, 2010). The robustness of this result may not be surprising given the abundance of evidence for associative learning and its pride of place in many theories of attitude change (Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006; Strack & Deutsch, 2004). Such findings are, however, also compatible with a propositional theory of implicit evaluation (De Houwer, 2014), which posits that the implicit attitude created as a result of exposure to stimulus pairings is mediated by propositional, rather than associative, mental representations. Although the stimuli may be presented in physiological or nonlinguistic form (e.g., shock or photographs), the assumption is that propositions about CS–US relations can also be formed on the basis of exposure to stimulus pairings. When receiving repeated CS–US pairings, humans can infer the underlying relationship and represent that information in propositional form.

A separate body of work has provided ample evidence that it is possible to shift implicit attitudes by using mere ES. For instance, Gregg, Seibt, and Banaji (2006) created strong novel implicit attitudes toward initially neutral fictitious social groups using vignettes that described one of the social groups as positive and the other social group as negative (see also Ranganath & Nosek, 2008) and, in another condition, by instructing participants to suppose that Group X has positive traits and Group Y has negative traits. In these studies, change in implicit attitudes was created on the basis of information that was highly diagnostic of the targets (see Cone & Ferguson, 2015). De Houwer (2006) showed the presence of such learning using an even more minimal manipulation that is

[1] This should not be taken to suggest that language cannot play a role in repeated evaluative pairing paradigms involving human participants. However, both traditionally and in our project, such paradigms have used quite uninformative verbal instructions with almost sole focus on exposure to stimulus pairings.

crucial to the present studies. He simply informed participants that nonwords (Experiment 1) or names from fictitious social groups (Experiment 2) would be paired with known positive or negative words in an upcoming evaluative conditioning task. Although participants were never exposed to any actual pairings between CS and US, novel implicit attitudes toward the initially neutral stimuli were found to have formed in line with this most minimal verbal instruction. Together, these studies demonstrate that ES are sufficient to shift implicit attitudes (see also Gast & De Houwer, 2013; Van Dessel, De Houwer, Gast, & Smith, 2015). Such a result may be surprising for at least two reasons. First, implicit attitudes are often recalcitrant: A large-scale comparative investigation showed the failure of many learning interventions to produce any change in implicit attitudes, although some did have impact (Lai et al., 2014). Second, the finding that one-shot language-based attitude induction modalities have the power to shift implicit attitudes contradicts strict theories of associative learning positing that implicit attitudes form as a result of exposure to stimulus associations in one's environment over extended periods of time. The result obtained by De Houwer (2006) that even minimal propositional statements can shift implicit evaluation is, however, compatible with single-process theories suggesting that REP and ES are two ways of creating the same underlying propositional representation.

Comparing REP and ES

It is established that, relative to a neutral baseline, implicit attitudes can shift either on the basis of REP or ES. However, do REP and ES give rise to implicit attitudes of comparable magnitude? And does the joint effect of REP and ES exceed their respective separate effects? The answer to these important questions is not known. Despite an impressive number of studies exploring the effects of REP, on the one hand, and the effects of ES, on the other hand, no systematic comparison of the two learning modalities has been undertaken, even though such a comparison would provide evidence about their relative power to influence attitudes, thus informing and constraining theories of implicit evaluation. The conspicuous lack of a direct comparison might have to do with the fact that the overwhelming majority of studies on implicit attitude acquisition and change have been conducted within the confines of two major theories (Greenwald, Pratkanis, Leippe, & Baumgardner, 1986), neither of which favor a comparative approach. As we discuss below, projects steeped in associative models (Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006; Strack & Deutsch, 2004) have mostly attempted to find a match between cognitive systems and learning modalities (ES change explicit attitudes vs. REP change implicit attitudes). On the other hand, work relying on propositional theories (De Houwer, 2009, 2014; Hughes, Barnes-Holmes, & De Houwer, 2011; C. J. Mitchell et al., 2009) has focused on demonstrating the—at least initially—surprising result that implicit attitude change via ES is possible. Neither theoretical perspective has had a stake in directly contrasting or combining REP and ES in the context of implicit evaluation even though both would gain from the results of such a test, whatever the outcome. Such a comparison as well as a test of the joint influence of the two learning modalities on implicit attitudes is the primary purpose of the present research.

In a sense, this project sought to do for the study of implicit social attitudes what Olsson and Phelps (2004, 2007) accomplished for fear conditioning, namely a direct comparison of REP and ES. Accordingly, in our studies, we induced the same implicit attitude toward the same attitude object within the same experiment with random assignment of participants to learning modality conditions, each of which involved either REP or ES. Some participants learned about an attitude object exclusively via exposure to target stimuli paired with valenced images, whereas other participants learned about the same attitude object using only verbal instructions and without any exposure to stimulus pairings.

Based on the extensive literature available we expected that both attitude induction modalities would create learning compared with baseline. The main focus here was to directly test the relative strength of the learning effects produced by each intervention. In terms of the comparison between REP and ES, the studies reported here had three possible outcomes: Both REP and ES could produce learning at comparable levels; REP might outperform ES; or ES might outperform REP. An additional condition involving joint presentation of both types of learning also allowed us to test whether the two together produce more robust attitude change than each one by itself. As discussed in more detail below, if the underlying representation is common to both interventions, then the two together need not be more robust than the single effect of the stronger of the two.

There are at least two lines of evidence that favor the superiority of REP. First, according to theories of associative learning and attitude change (Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006; Strack & Deutsch, 2004), unlike explicit attitudes that can be updated very rapidly, implicit attitudes shift gradually over time. If this is the case, the REP condition, which involves multiple stimulus pairings, should be more effective than ES, which relies on a verbal instruction to expect particular pairings, with no actual pairings presented. In fact, certain kinds of ES, such as instructing participants to adopt an egalitarian mindset, do not change implicit race attitudes at all (Lai et al., 2014). Second, when REP and ES are contradictory, REP usually outperform ES in their effect on implicit attitudes (but see Peters & Gawronski, 2011; Zanon, De Houwer, Gast, & Smith, 2014). For instance, Gregg, Seibt, and Banaji (2006, Experiment 3) found that informing participants that the REP to which they were exposed are invalid (an ES-based manipulation) does not have an impact on implicit evaluations. Similar results have been obtained using other paradigms involving supraliminally presented behavioral information versus subliminally presented valenced primes (Rydell & McConnell, 2006; Rydell, McConnell, Mackie, & Strain, 2006), affirmative versus negative sentences (DeCoster, Banner, Smith, & Semin, 2006), and REP versus ES containing causal information (Moran & Bar-Anan, 2013). These studies suggest that even in the absence of contradictory information, REP might be superior to ES.

On the other hand, single-process theories of implicit evaluation (De Houwer, 2009, 2014; Hughes et al., 2011; C. J. Mitchell et al., 2009) have advanced the idea that REP and ES should produce learning effects of similar magnitude. If, as posited by these theories, all human learning is mediated by the same kinds of propositional representations, neither learning modality should be inferior or superior to the other. In fact, when Gast and De Houwer (2013) induced novel implicit attitudes toward the same nonwords

using ES (Experiment 2a) and REP (Experiment 2b), the magnitude of implicit attitudes was found to be highly similar across the two experiments. However, the comparison between REP and ES remains post hoc and incidental because in this project participants were not randomly assigned to conditions and time on intervention was not fixed across studies. Nonetheless, it is possible that the effects of both attitude induction modalities might be comparable under more controlled conditions as well.

Finally, although unlikely given the research we have reviewed, it is conceivable that under the conditions created by the current experiments, ES will be superior to REP in producing change in implicit attitudes. As of now, no evidence suggesting this outcome exists and no theories would make such a prediction. Thus, the studies we conduct offer the first test that allows for the possibility of such a result to emerge. Unlike previous studies conducted under the dual-process view (e.g., Moran & Bar-Anan, 2013), our design made it possible for ES to exert their effect without the influence of countervailing information, and unlike previous studies conducted under the propositional view (Gast & De Houwer, 2013), we ensured, to the extent possible, equivalent exposure to REP and ES.

In addition to the question of relative effectiveness, we sought to explore whether attitude acquisition effects from REP and from ES are additive or redundant. In line with single-process theories of implicit evaluation (De Houwer, 2014; C. J. Mitchell et al., 2009), the effects of REP and ES on implicit attitudes should be comparable and redundant because both are mediated by the same propositional representations (e.g., "X is good and Y is bad"). Under this theory, one would expect the combination of both attitude induction modalities to produce a learning effect on par with their separable effects. On the other hand, dual-process theories of implicit evaluation might offer two different predictions depending on whether they are strictly associationist or allow for interactions between different learning modalities. Under strict associative learning theories (Rydell et al., 2006; Strack & Deutsch, 2004), ES should produce at most a negligible learning effect on implicit attitudes, which renders the question of additivity versus redundancy moot. On the other hand, some dual-process theories (Gawronski & Bodenhausen, 2006) allow for interactions between the explicit and implicit system. Under such theories, the effects of both attitude induction modalities should be additive, with REP creating implicit attitudes by strengthening links between conceptual nodes in long-term memory and ES affecting implicit attitudes via a separate pathway mediated by explicit attitudes.

Overview of the Present Studies

Each of Studies 1–5 reported below consisted of a learning phase and a test phase. Based on the considerations outlined above, the learning phase included four between-participants conditions: (a) a *control condition* in which we measured participants' implicit preferences at baseline; (b) a *REP condition* in which stimuli representing one social or nonsocial category were paired with positive images and stimuli representing another social or nonsocial category were paired with negative images over a sequence of trials; (c) an *ES condition* in which participants were informed that they would be exposed to stimulus pairings but in reality, the verbal instruction about stimulus pairings served as the sole attitude induction experience; and (d) a *combined (ES + REP) condition* in which verbal instructions were followed by actual stimulus pairings. After the learning phase, participants completed an Implicit Association Test measuring their implicit attitudes toward the two targets included in the learning phase and, finally, explicit attitudes toward the targets were measured.

Intrinsically valenced photographs (Study 4) and valenced line drawings (all other studies) served as US that would produce the necessary evaluative changes in the REP condition. To examine the generalizability and potential moderators of our findings, attitude objects—represented by visual CS—were varied across studies. Study 1 used names from two initially neutral fictitious social groups; Study 2 used more ecologically valid social stimuli (drawings of long and square faces); Study 3 used nonsocial targets (squares and rectangles); Study 4 used faces from pre-existing social groups, young and elderly people; and Study 5 tested another pre-existing social contrast (US vs. foreign) using photographs and drawings as CS. Finally, in Studies 6A–6C we used novel social stimuli (fictitious names and individuals with long and square faces) to eliminate possible confounds that may have accounted for some of the results observed in the first five studies, including time spent on intervention, levels of construal, and expectancy effects.

In these studies, some of the categories we used, such as fictitious groups and geometric shapes, might be expected to be relatively evaluatively neutral. However, as several previous studies have reported (Bargh et al., 1992; Glaser & Banaji, 1999), implicit measures can detect surprising preferences in one or another direction even when the stimuli seem to be evaluatively equal. For instance, as we show below, most participants exhibit a baseline implicit preference in favor of squares over rectangles (perhaps because of their greater symmetry) but the opposite preference for faces, rectangular over square (perhaps because they are thinner). Therefore, the conditions of greatest interest will be those where attitude objects are paired with attributes that are evaluatively opposite of the baseline attitude.

Study 1

Previous research has established that implicit attitudes can be shifted either using REP of stimuli (Gibson, 2008; Grumm et al., 2009; Olson & Fazio, 2001, 2006; Prestwich et al., 2010) or using ES about an attitude object (De Houwer, 2006; Gast & De Houwer, 2013; Gregg et al., 2006; Van Dessel et al., 2015). In Study 1, we sought to directly compare these two learning modalities in terms of their effectiveness in shifting implicit attitudes. To our knowledge, this is the first study that directly contrasts the effectiveness of the two attitude induction modalities within the same experiment, using random assignment of participants to conditions, and fixing the time spent on each learning experience. The relative contribution of each learning modality would be difficult to gauge without such a test. Moreover, a third joint condition was included to measure the combined effects of the two learning modalities, that is, to probe whether ES followed by REP create more robust learning of attitudes than either REP or ES in isolation. In Study 1, we used two attitudinally neutral fictitious social groups, the Laapians and the Niffians (Gregg, 2000), as target stimuli.

Method

Participants and design. We recruited 707 volunteers from the Project Implicit educational website (https://implicit.harvard.edu/implicit/) to participate in the study. Participants who did not complete the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), which was the main dependent measure of the study, could not be included in subsequent analyses ($N = 22$). In line with the recommendations of Greenwald, Nosek, and Banaji (2003), participants whose response latencies were below 300 ms on more than 10% of all IAT trials, suggesting inattentive responding, were also excluded from all analyses ($N = 10$). This resulted in a final sample size of $N = 675$ participants ($N = 446$ female, mean age = 39.24, $SD = 15.17$ years). For online research, involving no payment, these are excellent retention rates. Because obtaining no differences across learning modality conditions was a potentially interesting finding, we used a large sample to make sure that we had sufficient power to detect differences provided that they existed in the population. In fact, post hoc power calculated on the basis of the effect size that we obtained was 1 to within machine precision. Participants were randomly assigned to one of four learning modalities (control vs. REP vs. ES vs. combined) and one of two congruency conditions (learning congruent vs. incongruent with overall, not individually based, prevailing attitudes), with both factors varied between participants. In the context of the present study, we refer to the Laapian–good condition as the congruent condition because in the control condition 98 participants showed an implicit preference in favor of Laapians over Niffians, and only 72 participants exhibited the opposite preference. This distribution does not differ from chance, $p = .054$ by binomial test, and congruency did not modulate the main effect of learning modality; therefore, the analyses reported below collapse across this factor. Assignment to conditions was fully random, that is, the number of participants assigned to each condition was not fixed. One hundred seventy participants were assigned to the control condition, 154 to the REP condition, 159 to the ES condition, and 192 to the combined condition.

Materials. Cartoon drawings of intrinsically positive objects, including a beach, a flower, a heart, an ice cream cone, and a sun, and intrinsically negative objects, including an insect, a frowny face, a fleeing man, a snake, and an assassin, served as US.[2] Names from two fictitious social groups, the Laapians and the Niffians (Gregg, 2000), served as CS. The names were all pronounceable nonsense words consisting of three syllables that followed a certain phonological pattern. Laapians' names ended with the syllable *lap* (e.g., Deebolap or Maasolap), whereas Niffians' names ended with the syllable *nif* (e.g., Ibbonif or Yossanif). These stimuli had been pretested and were found to be attitudinally neutral on both explicit and implicit measures (Gregg, 2000).

Procedure and Measures

Overview of the procedure. The procedure consisted of a learning phase, followed by an implicit and an explicit measure of attitudes. In the learning phase, each participant was randomly assigned to one of four learning modalities (control vs. REP vs. ES vs. combined) in a between-participants design. We varied learning modality conditions between participants for two reasons. First, repeated administrations of the IAT (Greenwald et al., 1998) are known to decrease the magnitude of the IAT effect (Nosek, Banaji, & Greenwald, 2002). Second, taking a preintervention IAT selectively reduces the effects of REP on implicit attitudes (Lai et al., 2014), which would have given an unfair advantage to the ES condition in the present studies. Moreover, each participant was also assigned to one of two congruency conditions, that is, they learned either that Laapians are good and Niffians are bad (congruent condition) or that Niffians are good and Laapians are bad (incongruent condition). To make learning modalities fully comparable to each other, time spent on acquiring the attitude was fixed to 160 s across all learning conditions. Initial instructions, stimulus pairs, and evaluative statements appeared on the screen automatically for a fixed amount of time. Immediately after the learning episode, participants completed an IAT designed to measure attitudes toward the two target groups, followed by two feeling thermometer items tapping participants' explicit attitudes toward the target groups. Finally, participants were thanked and debriefed.

Learning Modality Conditions

REP condition. In this condition, the learning phase consisted of exposure to stimulus pairings. No verbal statements were used to explain the relationship between stimuli and no verbal labels were used to refer to the target groups or the US. Participants received general instructions explaining the nature of the learning task, followed by a presentation of the full set of CS and US.[3] Participants were then exposed to 37 trials in a standard visual–visual evaluative conditioning paradigm (Levey & Martin, 1975), that is, both CS and US were presented in the visual modality. On each trial, one CS (a Laapian or Niffian name) and one US (a line drawing of an intrinsically positive or negative object) were presented simultaneously next to each other in the center of the screen for 2,500 ms, followed by an intertrial interval of 1,000 ms, consisting of a blank screen. Within participants, target group and valence were perfectly matched with each other, that is, a certain target group (Laapians or Niffians) was paired only with positive US and the other target group was paired only with negative US. The pairings of specific CS to specific US were randomized for each participant.

ES condition. In this condition, learning occurred by presenting a verbal instruction; participants did not experience any actual stimulus pairings. Upon entering the study, participants received some general explanations describing the nature of the learning task. Crucially, participants were informed that they would see parings of names with pictures such that one target group (e.g., Laapians) would always be paired with pictures of pleasant things and the other target group (e.g., Niffians) would always be paired with pictures of unpleasant things, followed by a presentation of the full set of CS and US. In fact, participants were not exposed to stimulus pairings.

Combined condition. Like in the ES condition, participants were informed that one target group would always be paired with pleasant things, whereas the other target group would always be paired with unpleasant things. Like in the REP condition and

[2] The stimuli for Study 1 and all other studies are available for download from the Open Science Framework (OSF; https://osf.io/jyk8c/).

[3] The full text of the instructions is available for download from OSF (https://osf.io/jyk8c/).

unlike in the ES condition, participants were exposed to CS–US pairings. To keep time spent on learning consistent across conditions, participants in the combined condition were exposed to 20, rather than 37, conditioning trials.[4]

Control condition. The structure of the control condition was identical to the combined condition; however, the target groups were never mentioned and participants were not exposed to any presentations of the CS. Rather, some participants were informed that they would see positive images paired with positive images and negative images paired with negative images, whereas other participants were informed that they would see positive and negative images paired with each other. Like in the combined condition, participants were subsequently exposed to pairings; however, instead of CS–US pairings, they saw 20 US–US pairings. Thus, no task-relevant learning could occur.

Implicit attitudes. Following the learning phase of the study, participants completed an IAT (Greenwald et al., 1998) as the measure of implicit attitudes toward the two target groups. The IAT was chosen as an implicit measure because it tends to produce large effect sizes both in general and specifically in the context of attitude induction via evaluative conditioning (Hofmann et al., 2010; Nosek et al., 2007). The IAT consisted of five blocks with 20, 20, 40, 20, and 40 trials each. In Block 1, participants sorted positively and negatively valenced words[5] and in Block 2, they sorted Laapian and Niffian names used in the learning phase. In Block 3, the first critical block, both sorting tasks were combined. For participants in the control condition, the initial assignment of valence to target group was randomized. For participants who underwent prior learning, the first critical block of the IAT was always in line with the prior attitude induction. For instance, if a participant learned that Laapians are good and Niffians are bad, they sorted Laapians with positive words and Niffians with negative words on the first critical block. This was necessary to make sure that the IAT remained purely a test of previously acquired attitudes rather than forcing participants to override prior learning by presenting counterattitudinal statements. In Block 4, participants practiced the new assignment of target groups to response keys and finally, in Block 5, they completed a combined task with the opposite assignment of target group to valence.

Explicit attitudes. Following the IAT, participants were asked to respond to two feeling thermometer items, each of which measured explicit attitudes toward one of the target groups. On the first item participants were instructed to indicate how warmly or coldly they felt toward Laapians. On the second item they were instructed to indicate how warmly or coldly they felt toward Niffians. Responses were provided on 10-point Likert scales, anchored by "extremely warmly" on the left-hand side and "extremely coldly" on the right-hand side.

In line with the research reviewed earlier, the primary interest of the present work is in the effects of REP and ES on implicit attitudes. We collected matching data on explicit attitudes in the hope that the results may provide some insight into dissociations between more and less automatic evaluations. Given their status in this research, explicit measures were always administered following the IAT, and are best viewed as manipulation checks. As such, data from these measures will be reported briefly for each experiment, with the focus on whether they are evaluatively in the same direction as the IAT score. Given the attitudinally neutral targets used in the present study, no dissociation between implicit and explicit attitudes is expected.

Results[6]

Manipulation check. As in many such experiments, explicit and implicit attitudes were significantly correlated with each other, Pearson's $r = .36$ [.30; .43], $t(664) = 10.04$, $p < .001$, underscoring the soundness of the manipulation and experimental procedure.[7]

Implicit attitudes. Participants' implicit attitudes toward the two target groups were calculated using the scoring algorithm recommended by Greenwald, Nosek, and Banaji (2003), with positive D scores indicating an implicit preference in line with the prior learning episode. Descriptive statistics (means and *SD*s) by condition are reported in Table 1 but will not be discussed in detail given that the coefficients in the regression analysis below express condition means and mean differences across conditions.

To assess the relative strength of the attitude induction modalities, D scores were submitted to a linear regression, with learning modality (control vs. REP vs. ES vs. combined) as the sole predictor.[8] Because we used dummy coding with the control condition as the reference category, the intercept corresponds to the mean of the control condition and the slope parameters correspond to learning effects compared with control. The intercept did not significantly differ from zero, $b = 0.00$ [−0.07; 0.08], $t(671) = 0.11$, $p = .912$, indicating a neutral baseline in the control condition. More important, however, REP created a statistically significant learning effect compared with baseline, $b = 0.43$ [0.32; 0.54], $t(671) = 7.73$, $p < .001$. ES produced an even more sizable learning effect, $b = 0.50$ [0.39; 0.61], $t(671) = 9.09$, $p < .001$, and the strongest learning effect was observed in the combined condition, $b = 0.55$ [0.44; 0.65], $t(671) = 10.46$, $p < .001$.

To directly investigate the relative effectiveness of the REP, ES, and combined conditions relative to each other, we drew 10,000 bootstrap samples for each mean difference and calculated 95, 99, and 99.9% bootstrap confidence intervals (CIs) around the esti-

[4] In future studies, time can be allowed to vary with number of trials held constant, but there is no reason to assume that one trade-off is superior to the other. Because time was the crucial dimension to compare ES to REP, it was used by extension for this third condition as well. Moreover, as described in more detail in the context of Study 6A, increasing the number of evaluative conditioning trials from 20 to 37 does not create a stronger implicit attitude.

[5] Positive items included *love*, *peace*, *joy*, *happy*, *sweet*, *glory*, and *success*. Negative items included *hate*, *war*, *devil*, *bomb*, *bitter*, *agony*, and *failure*.

[6] Raw data and analysis scripts for all studies are available for download from OSF (https://osf.io/jyk8c/).

[7] The mean levels of explicit attitudes for each experimental condition are available on OSF (https://osf.io/jyk8c/).

[8] In lieu of the analyses reported here, that is, linear regressions combined with bootstrapping for additional contrasts, it would have been possible to conduct two-way ANOVAs, with REP (present vs. absent) and ES (present vs. absent) as the two predictors. We decided against this possibility given that the combined condition was not a perfect combination of the REP and ES conditions, among other reasons because time on intervention was fixed across the three learning conditions and thus the combined condition did not include the same number of stimulus pairings as the REP condition. However, the raw data files published on OSF enable other investigators to pursue this or any other alternative data analysis approach.

Table 1
IAT D Score Means and SDs (in Parentheses) by Congruency and Learning Condition (Control; REP = Repeated Evaluative pairings; ES = Evaluative statements; Combined) for Studies 1–5

	Congruent				Incongruent			
	Control	REP	ES	Combined	Control	REP	ES	Combined
Study 1	—	—	—	—	.00 (.47)	.44 (.56)	.51 (.47)	.56 (.51)
Study 2	.48 (.46)	.54 (.43)	.57 (.42)	.67 (.39)	.09 (.52)	.17 (.56)	.41 (.50)	.46 (.44)
Study 3	.49 (.40)	.50 (.45)	.65 (.38)	.69 (.38)	.03 (.52)	.54 (.45)	.55 (.47)	.55 (.49)
Study 4	.61 (.35)	.66 (.38)	.71 (.35)	.72 (.35)	−.61 (.32)	−.35 (.46)	−.19 (.46)	−.13 (.50)
Study 5	.75 (.35)	.78 (.43)	.89 (.32)	.88 (.33)	−.63 (.39)	−.57 (.39)	−.29 (.46)	−.31 (.57)

Note. Positive D scores indicate implicit attitudes in line with the prior learning. IAT = Implicit Association Test.

mate of the difference. If the 95% CI includes zero, we conclude that the two condition means are not significantly different. If the 95, 99, or 99.9% CI does not include zero, we conclude that the condition means are significantly different from each other with $p < .05$, $p < .01$, or $p < .001$, respectively. For the REP versus ES contrast, we found a difference of $b_{diff} = 0.07$ [−0.04; 0.18] in favor of ES; however, it failed to reach statistical significance, $p > .05$. Whereas the combined condition added significant value to the REP condition, $b_{diff} = 0.12$ [0.01; 0.23], $p < .05$, the difference between the ES and combined conditions was not significant, $b_{diff} = 0.05$ [−0.05; 0.15], $p > .05$.

In terms of absolute strength, the implicit attitudes found in each of the three learning modality conditions were on par with some of the strongest implicit attitudes found in American adult populations toward real groups such as the implicit preference for young over elderly people or for Whites over Blacks (Nosek, Greenwald, & Banaji, 2005). In terms of their relative strength, REP and ES produced comparable learning effects; however, the combined condition added value only to the REP condition and not to the ES condition.

Discussion

Three learning modalities and a control condition were created with the aim of testing the relative effectiveness of two different methods of implicit attitude change involving two fictitious social groups. All three learning modalities proved to be effective even though one of them, ES, relied on a most minimal instruction. We found strong implicit attitudes on par with some of the most robust social group attitudes observed in previous work. This study reveals the ease with which implicit attitudes can be created, both in time taken to produce an effect (less than 3 min) and the minimal instruction that is needed to produce them, at least in the ES condition. More important, Study 1 replicates prior research showing that exposing participants to ES simply stating that the attitude object *will be paired* with positive or negative stimuli can result in strong implicit attitudes toward the target (De Houwer, 2006; Gast & De Houwer, 2013; Van Dessel et al., 2015).

That ES were not inferior to REP in shifting attitudes is a new result and surprising for a number of reasons. First, the REP condition was arguably more attentionally intense than the ES condition. Whereas in the former condition participants were presented with a new stimulus pair every 3.5 s, in the latter condition participants were merely asked to keep reading some instructions that remained stationary on the screen. Second, informing participants that they will view particular pairings and then never presenting the pairings should not have given rise to any strong attitudinal effects, considering that this information is not in any way diagnostic of the targets (Cone & Ferguson, 2015). Moreover, based on these results it seems that learning via REP and learning via ES are, at least to some extent, redundant. REP did not offer any incremental value beyond ES, suggesting that the learning created by the combined intervention is subsumed under the learning created by the ES intervention alone. For these reasons, it is particularly striking that ES allowed the formation of any attitude at all, let alone as robust as the one observed. However, this first result may reflect possible artifacts of the experimental situation. Moreover, it is at least theoretically possible that despite the exceedingly low p value, the result is a Type 1 error.

Study 2

In Study 1 we found that REP and ES both created learning effects. The learning effect produced by ES was on par with the learning effect produced by REP, and the combined condition only added value to REP and not to ES. The stimuli used in Study 1 offered the undeniable advantage of a neutral baseline, which is notoriously difficult to achieve in the context of implicit social attitudes. By virtue of this attitudinal neutrality, we had complete control over participants' learning, without any contaminating effects of prior exposure to the stimuli. However, in experimental psychology there is often a conflict between internal and external validity (Banaji & Crowder, 1989, 1991). In Study 2 we sought to tilt the balance toward the latter and used more ecologically valid social target stimuli—drawings of individuals with long and square faces. Based on pretesting we expected that participants would have a fairly strong implicit preference in favor Longfaces at baseline. Seemingly neutral stimuli are often found to be nonneutral using implicit measures of attitudes (Bargh et al., 1992; Fazio et al., 1986; Glaser & Banaji, 1999; Gregg et al., 2006). Thus, the incongruent condition (i.e., the condition in which participants' attitudes are allowed to move away from, rather than closer to, the prevailing attitude) presents a cleaner test case for our hypotheses, considering that in the congruent learning condition ceiling effects might occur. In fact, to avoid this problem, many classic attitude induction studies (Cacioppo, Petty, & Morris, 1983; Petty & Cacioppo, 1979; Petty, Wells, & Brock, 1976) and studies on attitude induction via EC and REP (De Houwer, 2006; Lai et al., 2014; Van Dessel et al., 2015) contain only incongruent learning conditions. Accordingly, our analyses focus on the incon-

gruent condition. Findings from the congruent condition are reported in a footnote and discussed in more detail only in the context of results collapsing across experiments. Moreover, if REP is to be a purely nonlinguistic learning modality, Study 1 was problematic in that group members were presented as words. In Study 2, all stimuli—including the good–bad UC attributes and the long-faced and square-faced individuals serving as CS—were presented nonlinguistically, in pictures.

Method

Participants. We recruited 1162 volunteers from the Project Implicit educational website to participate in the study. Of these, 119 participants were excluded from all analyses for failing to complete the IAT and 9 participants were excluded because they had response latencies below 300 ms on more than 10% of all IAT trials. This resulted in a final sample size of $N = 1034$ participants ($N = 584$ female, mean age = 33.48, $SD = 14.47$ years). This sample size, combined with the effect size obtained, yielded post hoc power of >.99 in the incongruent condition and .86 in the congruent condition. Three hundred fourteen participants were assigned to the control condition, 226 to the REP condition, 250 to the ES condition, and 244 to the combined condition. Moreover, each participant was also assigned to one of two congruency conditions, that is, they learned either that Longfaces are good and Squarefaces are bad (congruent condition, $N = 506$) or that Squarefaces are good and Longfaces are bad (incongruent condition, $N = 528$). We refer to the Longfaces–good condition as the congruent condition because in the control condition 210 participants showed an implicit preference in favor of Longfaces over Squarefaces, and only 104 participants exhibited the opposite preference, $p < .001$, by binomial test.

Materials and procedure. The learning procedure (including the US) and the implicit and explicit attitude measures paralleled the ones used in Study 1. However, unlike in Study 1, drawings of human faces served as CS. Faces were chosen as target stimuli because of their singular ability to represent a social being. Previous research has shown that human faces carry a wealth of social information that gives rise to instantaneous trait inferences (Ballew & Todorov, 2007; Olivola & Todorov, 2010; Oosterhof & Todorov, 2008; Todorov, Mandisodza, Goren, & Hall, 2005; Todorov, Said, Engell, & Oosterhof, 2008; Willis & Todorov, 2006). Neonates preferentially attend to faces (C. C. Goren, Sarty, & Wu, 1975), and the psychological dispositions inferred from faces by young children are remarkably similar to those inferred by adults (Cogsdill & Banaji, 2015; Cogsdill, Todorov, Spelke, & Banaji, 2014). However, face-to-trait inferences are dynamic and malleable (Hehman, Flake, & Freeman, 2015), suggesting that faces can be the targets of new learning.

In the present study we used two sets of faces that differed along a salient perceptual dimension (length-to-width ratio) for the purposes of attitude induction (cf. Hill, Lewicki, Czyzewska, & Schuller, 1990; Lewicki, 1986). Each face was generated by morphing photographs of two young White men facing the camera, cropped at the neck and displaying a neutral facial expression. In the following step, the faces were subjected to a sketch effect to create the appearance of line drawings. Crucially, in the last step of preparation, we manipulated the length-to-width ratio of the faces, yielding a set of long faces (length-to-width ratio = 2:1) and a set of square faces (length-to-width ratio = 4:3).

Results

Manipulation check. Explicit and implicit attitudes were found to be moderately correlated with each other, $r = .25$ [.20; .31], $t(1015) = 8.46$, $p < .001$, underscoring the soundness of our design and manipulation.

Implicit attitudes. Descriptive statistics by condition are reported in Table 1. To assess the relative strength of the attitude induction modalities, participants' D scores from the incongruent condition[9] were submitted to a linear regression. In the incongruent condition, the intercept significantly differed from zero, $b = 0.09$ [0.01; 0.16], $t(524) = 2.20$, $p = .028$, indicating a weak baseline preference for Squarefaces. REP did not produce a significant learning effect compared with baseline, $b = 0.08$ [−0.04; 0.20], $t(524) = 1.35$, $p = .176$. In contrast, the ES condition, $b = 0.32$ [0.20; 0.44], $t(524) = 5.21$, $p < .001$, and the combined condition, $b = 0.37$ [0.25; 0.49], $t(524) = 6.18$, $p < .001$, produced robust learning effects. By bootstrap, the ES condition created significantly stronger learning than the REP condition, $b_{diff} = 0.23$ [0.10; 0.37], $p < .001$. The combined condition outperformed the REP condition, $b_{diff} = 0.28$ [0.16; 0.41], $p < .001$, but not the ES condition, $b_{diff} = 0.05$ [−0.07; 0.16], $p > .05$.

Discussion

With a new set of social group stimuli, Study 2 replicated the results of Study 1 to the extent that we observed a fairly strong implicit attitude toward a novel social attitude object in each learning modality condition, confirming that the attitude induction was, again, successful. Moreover, the pattern of results obtained in the incongruent condition, which allowed ample room for movement, was similar to the pattern of results observed in Study 1 in that the REP condition produced the weakest learning effect, and the learning effects produced by ES and the combined intervention were on par with each other. Unlike in Study 1, however, where the learning effects created by ES were descriptively but not inferentially stronger than those created by REP, here REP was clearly outperformed by ES. This is a novel result that was not predicted by any of the theoretical perspectives reviewed. Moreover, informing participants that they would be exposed to certain stimulus pairings created an implicit attitude that was nearly as

[9] The learning effects created in the congruent condition were overall negligible. The intercept significantly differed from zero, $b = 0.48$ [0.41; 0.55], $t(502) = 13.95$, $p < .001$, indicating a strong baseline preference for Longfaces. The fact that baseline preferences differed across congruency conditions is because of the well-documented order effect on the IAT (Greenwald & Nosek, 2001; Nosek et al., 2005, 2007). REP did not produce a significant learning effect compared with baseline, $b = 0.06$ [−0.05; 0.16], $t(502) = 1.03$, $p = .302$. The learning effect produced by ES also failed to reach significance, $b = 0.09$ [−0.02; 0.19], $t(502) = 1.67$, $p = .096$. The combined condition produced a small but statistically significant learning effect, $b = 0.19$ [0.09; 0.30], $t(502) = 3.56$, $p < .001$. Using a bootstrapping method for additional contrasts, we found no significant difference between REP and ES, $b_{diff} = 0.03$ [−0.07; 0.14], $p > .05$. The combined condition added significant value to both the REP condition, $b_{diff} = 0.13$ [0.02; 0.25], $p < .05$, and the ES condition, $b_{diff} = 0.10$ [0.00; 0.20], $p < .05$.

strong as the attitude created by first informing participants of the stimulus pairings and then actually exposing them to the stimuli. Taken together, these results speak to the generalizability of the findings obtained in Study 1 to more ecologically valid targets where the baseline implicit preference is not neutral. Even minimal language-based manipulations involving nondiagnostic information seem to be robustly effective in shifting implicit attitudes and at least as powerful, if not more so, than repeated presentation of evaluative pairings.

Study 3

Studies 1 and 2 have demonstrated that ES are more effective than REP in shifting implicit attitudes. Moreover, REP did not seem to have any incremental effect over and above ES, suggesting that language provides solid cognitive scaffolding for social attitudes to be built. However, it remains to be directly investigated whether the social–nonsocial divide modulates the relative effectiveness of REP versus ES in shifting implicit attitudes. Different bodies of literature from social psychology suggest various possible hypotheses regarding this issue. According to one longstanding view, theories of social cognition can be derived from general theories about the physical world (Banaji & Bhaskar, 2000; see also Contreras, Banaji, & Mitchell, 2012). This perspective would predict that there should not be much difference between shifting implicit social and implicit nonsocial attitudes. On the other hand, social groups tend to be more complex than categories of nonsocial objects (Cantor & Mischel, 1979), and knowledge about social targets often evokes more emotion than knowledge about nonsocial targets (Norris, Chen, Zhu, Small, & Cacioppo, 2004). Moreover, work from social neuroscience suggests that different neural processes might underlie social and nonsocial learning (Contreras et al., 2012; J. P. Mitchell, Heatherton, & Macrae, 2002). The latter set of findings would predict that the acquisition of social and nonsocial attitudes might differ from each other. Specifically, according to Dunbar (1996), language may have evolved for the purpose of gossip, that is, strengthening social bonds by talking about shared acquaintances. If this is so, language-based learning mechanisms such as ES might be especially effective in creating attitudes toward social, rather than nonsocial, targets. Thus, in Study 3, we sought to test whether the separate and joint learning effects of REP and ES would remain the same for nonsocial targets (squares vs. rectangles) while keeping superficial visual features of the stimuli constant.

Method

Participants. We recruited 918 volunteers from the Project Implicit educational website to participate in the study. Fifty participants were excluded from all analyses for failing to complete the IAT and 5 participants were excluded because they had response latencies below 300 ms on more than 10% of all IAT trials. This resulted in a final sample size of $N = 863$ participants ($N = 606$ female, mean age = 37.87, $SD = 14.17$ years). This sample size, combined with the effect size obtained, yielded post hoc power of 1 within machine precision for the incongruent condition and .97 for the congruent condition. Two hundred thirty-six participants were assigned to the control condition, 199 to the REP condition, 209 to the ES condition, and 219 to the combined condition. Moreover, each participant was also assigned to one of two congruency conditions, that is, they learned either that squares are good and rectangles are bad (congruent condition, $N = 428$) or that rectangles are good and squares are bad (incongruent condition, $N = 435$). We refer to the squares–good condition as the congruent condition because in the control condition 164 participants showed an implicit preference in favor of squares over rectangles, and only 72 participants exhibited the opposite preference, $p < .001$, by binomial test.

Materials and procedure. The learning procedure (including the US) and the implicit and explicit attitude measures paralleled the ones used in Studies 1 and 2, with squares and rectangles serving as CS. The squares and rectangles were created by turning the pixels that formed the face stimuli in Study 2 into random noise and cutting them into form using square-shaped and rectangle-shaped cropping tools. This way, the CS used in Study 3 were similar to the CS used in Study 2 in terms of superficial visual features such as hue, saturation, and brightness. Pretesting had revealed a fairly high level of mean implicit preference in favor of squares over rectangles. In an attempt to reduce this implicit preference, we used sharp edges for squares and round edges for rectangles. Moreover, to keep the verbal labels maximally similar to the labels used in Study 2, squares were referred to as "square shapes" and rectangles were referred to as "long shapes" throughout the instructions and the words "square" and "long" were used as labels on the IAT.

Results

Manipulation check. Explicit and implicit attitudes were found to be moderately correlated with each other, $r = .32$ [.25; .37], $t(848) = 9.68$, $p < .001$, underscoring the soundness of our design and manipulation.

Implicit attitudes. Descriptive statistics by condition are reported in Table 1. To assess learning effects in the incongruent condition,[10] a linear regression was fit to the data, with learning modality as the sole predictor and the control condition as the reference category. The intercept, that is, the mean of the control condition, did not differ from zero, $b = 0.03$ [−0.06; 0.11], $t(431) = 0.66$, $p = .508$, indicating no baseline preference. Each learning modality produced a strong learning effect. In the REP condition, $b = 0.51$ [0.38; 0.63], $t(431) = 8.03$, $p < .001$; in the ES condition, $b = 0.52$ [0.39; 0.65], $t(431) = 8.05$, $p < .001$; and in the combined condition, $b = 0.52$ [0.39; 0.65], $t(431) = 7.99$, $p < .001$. We found no differences across learning conditions: For

[10] The learning effects created in the congruent condition were overall negligible. The intercept significantly differed from zero, $b = 0.49$ [0.42; 0.57], $t(424) = 12.96$, $p < .001$, indicating a strong baseline preference for squares. The difference in baseline preference across congruency conditions is because of an IAT order effect (Greenwald & Nosek, 2001; Nosek et al., 2005, 2007). As in Study 2, REP did not produce a significant learning effect compared with baseline, $b = 0.00$ [−0.11; 0.11], $t(424) = 0.05$, $p = .961$. The learning effects produced by ES, $b = 0.15$ [0.05; 0.26], $t(424) = 2.85$, $p = .005$, and the combined condition, $b = 0.19$ [0.09; 0.30], $t(424) = 3.64$, $p < .001$, were small but significant. The ES condition, $b_{diff} = 0.15$ [0.03; 0.27], $p < .05$, and the combined condition, $b_{diff} = 0.19$ [0.08; 0.30], $p < .01$, clearly outperformed REP. However, we found no significant difference between ES and combined, $b_{diff} = 0.03$ [−0.06; 0.13], $p > .05$, all three contrasts by bootstrap.

the REP versus ES contrast, $b_{diff} = 0.01$ [−0.11; 0.14], $p > .05$; for REP versus combined, $b_{diff} = 0.01$ [−0.12; 0.14], $p > .05$; and for ES versus combined, $b_{diff} = 0.00$ [−0.14; 0.13], $p > .05$, all by bootstrap.

Discussion

In Study 3 we tested the effectiveness of REP versus ES in the context of nonsocial attitudes. The pattern of results obtained in Study 3 had both some important similarities and some marked differences compared with the previous nonsocial studies. In line with Studies 1 and 2, we observed strong learning effects. Moreover, ES produced robustly strong implicit attitudes and REP did not offer any incremental value over and above ES. Unlike in Study 2, however, the learning effects produced by REP in isolation were on par with the learning effects produced by ES. Overall, learning effects were somewhat larger than in Studies 1 and 2, suggesting that attitudes toward nonsocial stimuli might be more malleable than attitudes toward social stimuli, especially in the face of stimulus pairings experienced in the environment. Crucially, however, Study 3 further underscores the generalizability of our findings about the power of ES-based learning in shifting implicit attitudes.

Study 4

In Studies 1 and 2 we demonstrated that ES are not any less effective than REP in shifting implicit attitudes toward novel social groups created within the context of the experiment. Using such ad hoc social groups offers the undeniable advantage of complete experimental control over participants' learning; however, such social groups may elicit "nonattitudes" (Pierce & Rose, 1974), thus compromising the external validity of the findings. Moreover, prior research suggests that REP and ES might differ in terms of their effectiveness in shifting real-world attitudes. Several studies have found that REP can shift implicit attitudes toward real-world social groups like the elderly or Blacks (Karpinski & Hilton, 2001; Lai et al., 2014; Olson & Fazio, 2006). However, this might not be the case for ES. In a recent study (Van Dessel et al., 2015), approach/avoidance instructions shifted implicit attitudes toward fictitious social groups (Niffites and Luupites) but not toward Whites and Blacks. This finding suggests that our results on the power of ES might not generalize to existing social groups.

However, a potential limitation of the study conducted by Van Dessel et al. (2015) is that they tested the difference between existing and novel targets using only Whites versus Blacks as target groups. Results obtained with these two groups may or may not generalize to other existing social groups. Implicit attitudes toward Blacks are recalcitrant (Lai et al., 2014), and the attempt to change them might even result in ironic reactance effects (Eagly & Chaiken, 1993). Therefore, rather than using White–Black as the target contrast, we probed whether the results obtained in previous studies would generalize to shifting attitudes toward two other social contrasts subject to high levels of implicit bias in contemporary American society. To provide a robust test of the relative and combined effects of REP and ES, we selected two pairs of target categories, one of them a biologically and evolutionary relevant social category and the other one an arbitrary-set category (Pratto, Sidanius, & Levin, 2006): young versus elderly and American versus foreign. Negative explicit (Cuddy & Fiske, 2004) and implicit (Levy & Banaji, 2002) attitudes toward the elderly are highly prevalent. Similarly, Americans express high levels of explicit (Fiske, Cuddy, Glick, & Xu, 2002) and implicit (Carter, Ferguson, & Hassin, 2011) preference for their ingroup over foreigners. More specifically, high levels of implicit bias measured toward the elderly (Nosek et al., 2007) and foreigners (Cunningham, Nezlek, & Banaji, 2004) have been amply documented in the context of IAT research, making these two contrasts an ideal test case for a comparative investigation of REP and ES in real-world settings.

Method

Participants. We recruited 881 volunteers from the Project Implicit educational website to participate in the study. Twenty-nine participants were excluded from all analyses for failing to complete the IAT and 4 participants were excluded because they had response latencies below 300 ms on more than 10% of all IAT trials. This resulted in a final sample size of $N = 848$ participants ($N = 600$ female, mean age = 33.85, $SD = 14.39$ years). This sample size, combined with the effect size obtained, yielded post hoc power of 1 within machine precision for the incongruent condition and, given the small effect size, .51 for the congruent condition. Two hundred seventeen participants were assigned to the control condition, 221 to the REP condition, 199 to the ES condition, and 211 to the combined condition. Moreover, each participant was also assigned to one of two congruency conditions, that is, they learned either that young people are good and elderly people are bad (congruent condition, $N = 419$) or that elderly people are good and young people are bad (incongruent condition, $N = 429$). We refer to the young–good condition as the congruent condition on the basis of previous work that has revealed robust implicit bias in favor of young over elderly people in American society (Levy & Banaji, 2002; Nosek et al., 2007). In line with this previous work, 209 participants in the control condition showed an implicit preference in favor of young over the elderly, and only 8 participants exhibited the opposite preference, $p < .001$, by binomial test.

Materials and procedure. The learning procedure and the implicit and explicit attitude measures paralleled the ones used in previous studies. In this study we used faces of young and elderly people as CS and a novel set of photographs as US. Conditioned stimuli were black-and-white photographs of faces of young and elderly people (White, both female and male), with external facial features removed. These stimuli were obtained from the Project Implicit website (http://implicit.harvard.edu). The US were taken from the Open Affective Standardized Image Set (OASIS; Kurdi, Lozano, & Banaji, 2016), an open-access collection of color photographs with valence and arousal norms. The US were selected to be highly positive or negative and moderately arousing. The positive set, consisting of OASIS images I256, I335, I345, I456, and I462, had an overall mean valence of 6.29 and a mean arousal of 4.22 (both measured on a 7-point scale). The negative set, consisting of OASIS images I119, I227, I236, I276, and I382, had an overall mean valence of 1.74 and a mean arousal of 4.06. This novel set of images was used for two reasons. First, because, REP did not outperform ES in any of the previous studies and unexpectedly so, we sought to verify that the superiority of ES was not

because of the particular US included in the studies. Second, because the US used in Studies 1 through 3 were line drawings, we expected that they might contaminate the results of this study by virtue of their stereotypic association with the category "young" rather than the category "elderly." The images that we selected for Study 4 were not stereotypically associated with either category.

Results

Manipulation check. Unlike in previous studies, explicit and implicit attitudes were found to be only weakly correlated with each other, $r = .07$ [.00; .14], $t(834) = 1.99$, $p = .046$. This is often the case when explicit measures are subject to social desirability pressures (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Nosek, 2005, 2007).

Implicit attitudes. Descriptive statistics by condition are reported in Table 1. To assess learning effects in the incongruent condition,[11] a linear regression was fit to the data, with learning modality as the sole predictor and the control condition as the reference category. The intercept was significantly different from zero, $b = -0.61$ [−0.69; −0.53], $t(425) = -14.88$, $p < .001$, which indicates a strong baseline preference for young over elderly, replicating previous research (Levy & Banaji, 2002). All three conditions, including REP, $b = 0.26$ [0.15; 0.38], $t(425) = 4.52$, $p < .001$, ES, $b = 0.42$ [0.30; 0.54], $t(425) = 6.96$, $p < .001$, and combined, $b = 0.48$ [0.37; 0.60], $t(425) = 8.10$, $p < .001$, produced considerable learning effects relative to control. Compared with the REP condition, the learning effects created by the ES condition, $b_{diff} = 0.15$ [0.03; 0.27], $p < .05$, and by the combined condition, $b_{diff} = 0.22$ [0.09; 0.34], $p < .01$, were more robust, and comparable with each other in terms of magnitude, $b_{diff} = 0.06$ [−0.06; 0.19], $p > .05$, all three contrasts by bootstrap.

Discussion

In Study 4, we tested the effectiveness of REP and ES in shifting implicit attitudes in the context of a real-world contrast, young versus elderly people. Consistent with previous work (Levy & Banaji, 2002; Nosek et al., 2007), we obtained strong implicit preference in favor of young over elderly at baseline. The learning effects observed in the incongruent condition, designed to move participants away from the prevailing implicit preference in favor of young over elderly, were robust. In line with Studies 1 and 3 (and unlike in Study 2), REP produced a significant learning effect in isolation, replicating previous work showing that implicit attitudes toward the elderly can be changed via REP (Karpinski & Hilton, 2001). However, like in Study 2, ES was more effective than REP and the learning effect observed in the combined condition was not superior to the learning effect observed in the ES condition. In other words, Study 4 seems to suggest that learning involving existing and ad hoc social groups might be less different from each other than previously assumed. The failure to find learning effects for mere instructions in previous work (Van Dessel et al., 2015) might have been the result of the particular contrast used (i.e., Whites vs. Blacks). Thus, the pattern of results obtained in Studies 1 and 2 does not seem to be confined to novel social groups created for the purpose of the experiment; rather, they seem to generalize to existing social groups subject to high levels of pre-existing implicit bias.

Study 5

As described in the introduction to Study 4, previous research has failed to find effects for ES in the context of real-world social groups (Van Dessel et al., 2015). Hence we sought to probe the robustness of the findings of Study 4 by replicating the same experiment using another real-world contrast subject to strong pre-existing explicit (Fiske et al., 2002) and implicit (Carter et al., 2011) attitudes—American versus foreign.

Method

Participants. We recruited 1035 volunteers from the Project Implicit educational website to participate in the study. Fifty-nine participants were excluded from all analyses for failing to complete the IAT and 18 participants were excluded because they had response latencies below 300 ms on more than 10% of all IAT trials. This resulted in a final sample size of $N = 958$ participants ($N = 672$ female, mean age = 30.79, $SD = 14.09$ years). This sample size, combined with the effect size obtained, yielded post hoc power of >.99 for the incongruent condition and .90 for the congruent condition. Participation was restricted to volunteers with both American citizenship and U.S. residence. Two hundred sixty-four participants were assigned to the control condition, 206 to the REP condition, 213 to the ES condition, and 275 to the combined condition. Moreover, each participant was also assigned to one of two congruency conditions, that is, they learned either that American is good and foreign is bad (congruent condition, $N = 478$) or that foreign is good and American is bad (incongruent condition, $N = 480$). We refer to the American–good condition as the congruent condition because in the control condition 254 participants showed an implicit preference in favor of American over foreign, and only 10 participants exhibited the opposite preference, $p < .001$, by binomial test.

Materials and procedure. The learning procedure and the implicit and explicit attitude measures paralleled the ones used in previous studies. In this study we used drawings and photographs as CS and the line drawings used in Studies 1 through 3 as US. CS were adapted from Devos and Banaji (2005) and Devos, Gavin, and Quintana (2010). For the American category, the stimuli included a 1-dollar bill, the Great Seal of the United States, the U.S. flag, the Golden Gate Bridge, and a map of the continental United States. For the foreign category, the stimuli included a 100-hryven bill, the coat of arms of Flanders, the flag of Djibouti, the Great Sphinx of Giza, and a map of Luxembourg rotated by 90 degrees to the left.

[11] Learning effects in the congruent condition were negligible. The intercept significantly differed from zero, $b = 0.61$ [0.54; 0.68], $t(415) = 17.11$, $p < .001$, revealing a strong baseline preference in favor of young over elderly. As in Studies 2 and 3, REP did not produce a significant learning effect, $b = 0.05$ [−0.05; 0.15], $t(415) = 0.05$, $p = .960$. The learning effects produced by ES, $b = 0.10$ [0.00; 0.20], $t(415) = 2.04$, $p = .041$, and the combined condition, $b = 0.11$ [0.01; 0.20], $t(415) = 2.14$, $p = .033$, were fairly small but significant and of comparable magnitude. None of the three remaining contrasts were significant, with REP versus ES at $b_{diff} = 0.05$ [−0.04; 0.15], $p > .05$; REP versus combined at $b_{diff} = 0.05$ [−0.04; 0.15], $p > .05$, and ES versus combined, $b_{diff} = 0.00$ [−0.09; 0.09], $p > .05$, all by bootstrap.

Results

Manipulation check. Explicit and implicit attitudes were found to be moderately correlated with each other, $r = .40$ [.34 .46], $t(944) = 13.54$, $p < .001$, underscoring the soundness of our design and manipulation.

Implicit attitudes. Descriptive statistics by condition are reported in Table 1. To assess learning effects in the incongruent condition,[12] a linear regression was fit to the data, with learning modality as the sole predictor and the control condition as the reference category. The intercept was significantly different from zero, $b = -0.63$ [−0.71; −0.56], $t(476) = -16.07$, $p < .001$, which indicates a strong baseline preference for American over foreign, replicating previous work (Nosek, 2005; Nosek et al., 2007). Unlike in Studies 1, 3, and 4, and like in Study 2, the REP condition did not produce a significant learning effect, $b = 0.06$ [−0.05; 0.18], $t(476) = 1.04$, $p = .299$. However, as in all previous studies, the ES condition, $b = 0.34$ [0.22; 0.46], $t(476) = 5.65$, $p < .001$ and the combined condition, $b = 0.32$ [0.22; 0.43], $t(476) = 5.89$, $p < .001$, created strong learning effects relative to baseline. Accordingly, we found statistically significant differences between the REP and ES conditions, $b_{diff} = 0.28$ [0.16; 0.40], $p < .001$, as well as the REP and combined conditions, $b_{diff} = 0.26$ [0.14; 0.38], $p < .001$. However, no significant difference was found between ES and combined, $b_{diff} = -0.02$ [−0.14; 0.11], $p > .05$, all contrasts by bootstrap.

Discussion

In Study 5 we tested whether the results obtained in Study 4 would generalize to another pre-existing social contrast. We replicated the major results from Study 4, with one notable exception. REP did not produce significant learning effects, suggesting that stimulus pairings experienced in the environment may not always be sufficient to move implicit attitudes away from a strong baseline. This is a useful result to have obtained because it represents a unique contrast to the strength of ES. Indeed, ES and the combined intervention produced learning effects significantly above baseline and on par with each other.

Study 6A

With the ES condition producing learning effects at least on par with the REP condition and the combined condition not adding any value to the ES condition, Studies 1–5 yielded some surprising yet consistent findings. Therefore, we conducted three follow-up studies in which we sought to test to what extent these findings are robust to procedural variations involving time on intervention (Study 6A), levels of construal (Study 6B), and expectancy effects (Study 6C), thus eliminating potential confounds. In Study 6A, we probed whether fixing time on intervention across the learning conditions was instrumental to the superiority of the ES manipulation, or the same effects would be obtained if learning unfolded in a self-paced manner. Moreover, we investigated whether the combined condition would produce stronger learning than the ES condition if it included the same number of conditioning trials as the REP condition.

Method

In Study 6A, conducted with 873 participants on Project Implicit and involving the same stimuli (i.e., line drawings of individuals with long and square faces) and the same four learning conditions (i.e., control, REP, ES, and combined) as Study 2, we did not equate the time spent on intervention across learning conditions. Rather, learning was self-paced in the ES condition, and the REP and combined conditions featured the same number of conditioning trials (37).

Results and Discussion[13]

In the incongruent (i.e., Squarefaces–good) condition,[14] the intercept, that is, the mean of the control condition, did not differ from zero, $b = -0.10$ [−0.20; 0.00], $t(429) = -1.90$, $p = .058$, indicating no baseline preference. Each learning modality produced a learning effect compared with control. In the REP condition, $b = 0.22$ [0.08; 0.36], $t(429) = 3.13$, $p = .002$; in the ES condition, $b = 0.46$ [0.33; 0.61], $t(429) = 6.52$, $p < .001$; and in the combined condition, $b = 0.59$ [0.44; 0.74], $t(429) = 7.77$, $p < .001$. The ES condition, $b_{diff} = 0.25$ [0.11; 0.38], $p < .001$, and the combined condition, $b_{diff} = 0.37$ [0.22; 0.51], $p < .001$, each outperformed REP, whereas the ES and combined conditions were on par with each other, $b_{diff} = 0.12$ [−0.03; 0.26], $p > .05$. Thus, we fully replicated the results of Study 2, suggesting that, at least in the present paradigm, time spent on intervention and number of conditioning trials may not be crucial in determining the magnitude of learning effects.

Study 6B

The differences across the REP and ES conditions may have emerged because of the fact that the instructions provided for each activated different levels of construal (Trope & Liberman, 2010). Whereas instructions in the ES condition explicitly referred to

[12] Learning effects in the congruent condition were negligible. The intercept significantly differed from zero, $b = 0.75$ [0.68; 0.81], $t(474) = 23.44$, $p < .001$, indicating a strong baseline preference for American over foreign. As in Studies 2, 3, and 4, REP did not produce a significant learning effect, $b = 0.03$ [−0.06; 0.12], $t(474) = 0.66$, $p = .508$. The learning effects produced by ES, $b = 0.14$ [0.05; 0.23], $t(474) = 3.03$, $p = .003$, and the combined condition, $b = 0.13$ [0.05; 0.22], $t(474) = 3.00$, $p = .003$, were fairly small but statistically significant and of similar magnitude. Accordingly, both ES, $b_{diff} = 0.11$ [0.01; 0.21], $p < .05$, and combined, $b_{diff} = 0.10$ [0.01; 0.20], $p < .05$, created a stronger learning effect than REP. We found no difference between ES and combined, $b_{diff} = -0.01$ [−0.09; 0.07], $p > .05$, all by bootstrap.

[13] Raw data for Studies 6A–6C, permitting further analyses not reported here, are available for download from OSF (https://osf.io/jyk8c/).

[14] The learning effects created in the congruent condition were overall smaller than in the incongruent condition. The intercept significantly differed from zero, $b = 0.37$ [0.28; 0.46], $t(436) = 8.18$, $p < .001$, indicating a strong baseline preference for Longfaces. Both the REP condition, $b = 0.23$ [0.10; 0.35], $t(436) = 3.64$, $p < .001$, and the combined condition, $b = 0.28$ [0.16; 0.41], $t(436) = 4.38$, $p < .001$, produced significant learning effects compared with control, whereas the ES condition did not, $b = 0.12$ [−0.01; 0.25], $t(436) = 1.87$, $p = .061$. Using a bootstrapping method for additional contrasts, we found no significant difference between REP and ES, $b_{diff} = -0.10$ [−0.22; 0.01], $p > .05$. The combined condition added significant value only to the ES condition, $b_{diff} = 0.16$ [0.04; 0.29], $p < .01$, but not to the REP condition, $b_{diff} = 0.06$ [−0.05; 0.16], $p > .05$, all by bootstrap.

social groups (e.g., "you will learn that a certain group of people is associated with pleasant things or unpleasant things"), encouraging high-level construal, instructions in the REP condition referred to the stimuli themselves (e.g., "you will see two types of drawings"), encouraging low-level construal. Even though this difference mirrors the way in which the two types of learning unfold under more ecologically realistic conditions, we probed whether removing it from the procedure would eliminate the superiority of the ES condition.

Method

In Study 6B, conducted with 603 participants on Project Implicit, we used the same stimuli as in Study 1 (Laapians vs. Niffians) but modified the instructions[15] such that in the REP+ condition they explicitly referred to social groups and attributes ("you will see two types of names, representing two groups, and two types of drawings, representing two traits"), encouraging high-level construal, whereas in the ES- condition, they referred exclusively to stimuli ("you will learn that certain names are associated with pleasant drawings or unpleasant drawings"), encouraging low-level construal.

Results and Discussion

Even though the present study sought to tilt the balance in favor of the REP condition over the ES condition by facilitating abstract construal of the stimuli in the former and concrete construal of the stimuli in the letter, the ES- condition still produced a significantly stronger learning effect than the REP+ condition, $t(599.9) = 2.39$, $p = .017$, Cohen's $d = 0.19$. Thus, different levels of construal created by the two sets of instructions are unlikely to account for the superiority of the ES condition observed in the previous studies.

Study 6C

The evaluative statements used in Studies 1–5 generated the expectation that the stimulus pairs would be shown to the participant later, whereas no such unfulfilled expectation was created in the REP condition. This feature of the ES condition raises the possibility that its superiority over the REP condition may be because of a Zeigarnik effect (Rajagopal, Raju, & Unnava, 2006; Savitsky, Medvec, & Gilovich, 1997; Zeigarnik, 1927), that is, a memory advantage as a result of the interrupted nature of the task. In Study 6C, we sought to eliminate this potential confound.

Method

Study 6C was conducted on Project Implicit with 429 participants and using the same stimuli as Study 1 (Laapians vs. Niffians). In this study, we tested whether removing the expectation of upcoming stimulus pairings would modulate the learning effect created by the ES condition. Thus, participants were assigned to one of two conditions. The first learning condition was identical to the ES condition of Study 1 (expectation condition), whereas the other one did not create the expectation of upcoming stimulus pairings but rather simply informed participants that one group was good and the other group was bad (no expectation condition).[16]

Results and Discussion

Despite the high-powered design, we found no difference across the expectation condition and the no expectation condition in terms of the strength of the implicit attitudes created, $t(426.94) = 1.21$, $p = .226$, Cohen's $d = 0.03$. Thus, overall, Studies 6A–6C demonstrate that time spent on intervention, levels of construal, or a Zeigarnik effect cannot account for the results of Studies 1–5 reported above.

Combined Analyses Across Experiments

Because Studies 1–5 shared the same basic design, we aggregated the results across these studies meta-analytically (see Figure 1). First, we calculated Cohen's d effect size measures for each learning modality condition (compared with the control condition of the given study). Second, to arrive at an estimate of the uncertainty around the effect sizes aggregated across studies, we adjusted D scores for mean differences across studies and obtained 95% CIs around the effect size estimates for each learning modality compared with control on the basis on 10,000 bootstrap samples.[17] Finally, we investigated whether effect sizes are significantly different from each other by using an indirect testing approach. We drew 10,000 bootstrap samples for each pairwise comparison, that is, REP versus ES, REP versus combined, and ES versus combined, separately in the congruent and incongruent conditions.[18] Next, we used the indirect testing approach described in more detail in the Results section of Study 1 to establish significant differences at the $p < .05$, $p < .01$, or $p < .001$ levels.

Overall, learning effects were considerably stronger in the incongruent condition, Cohen's $d = 0.71$ [63; 80], than in the congruent condition, $d = 0.27$ [17; 38], $p < .05$ by bootstrap. However, despite the difference in average effect sizes, the patterns of data by learning modality were similar across the congruent versus incongruent conditions.

REP produced the smallest learning effect relative to baseline both in the congruent condition, $d = 0.13$ [0.02; 0.27], and in the incongruent condition, $d = 0.53$ [0.42; 0.65]. Moreover, in the incongruent condition, effect sizes for REP were variable across studies, ranging from 0.16 in Study 2 to 1.04 in Study 3. Overall, the learning effects produced by ES were larger and less variable than the learning effects produced by REP, with $d = 0.27$ [0.14; 0.40] in the congruent condition and $d = 0.84$ [0.72; 0.96] in the incongruent condition, ps $< .01$ by bootstrap for the REP–ES contrasts. The combined condition also produced reliably large learning effects, with $d = 0.39$ [0.27; 0.52] in the congruent condition and $d = 0.83$ [0.72; 0.95] in the incongruent condition. The combined condition produced stronger learning effects than REP both in the congruent and the incongruent conditions, ps $<$ .001 by bootstrap, but the ES–combined contrast did not reach

[15] The full set of instructions is available for download from OSF (https://osf.io/jyk8c/).

[16] The full set of instructions is available on OSF (https://osf.io/jyk8c/).

[17] Because of the larger number of observations, this procedure yielded more stable estimates than bootstrapping Cohen's d values.

[18] For the purpose of this analysis, all participants from Study 1 were included in the incongruent condition because of the neutral baseline, which allowed ample room for movement in both the Laapian–good and the Niffian–good condition.

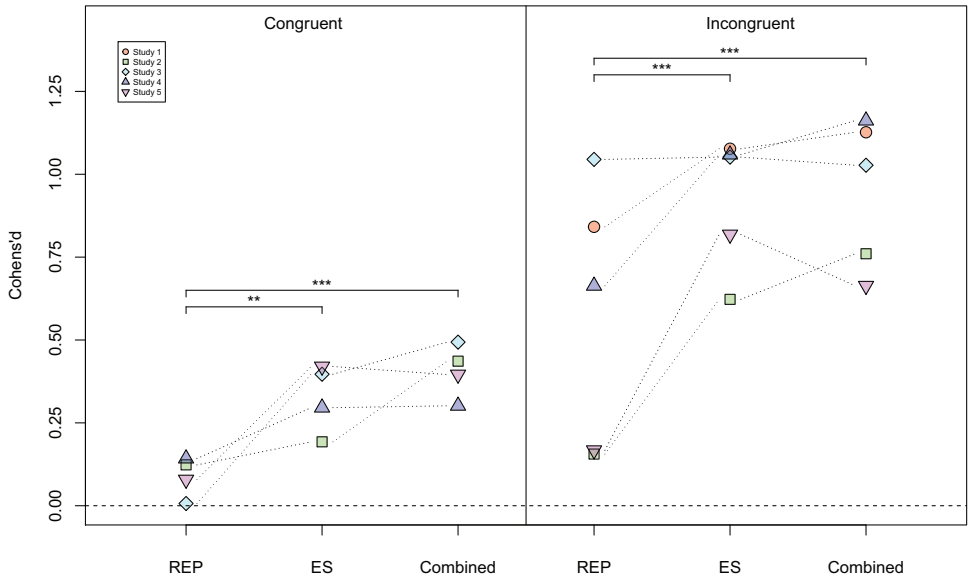

Figure 1. Meta-analytic results from Studies 1–5. The *x*-axis shows learning modality conditions (REP = repeated evaluative pairings; ES = evaluative statements; Combined) and the *y*-axis shows standardized effect sizes (Cohen's *d*s) comparing each learning modality to the control condition. Statistically significant differences between effect sizes (adjusted for mean D score differences across studies) are marked with asterisks, $^{**}p < .01$ and $^{***}p < .001$, based on 10,000 stratified bootstrap samples. See the online article for the color version of this figure.

significance in either case, *p*s $> .05$ by bootstrap, suggesting that even in these tests with greater power to detect differences, the combined condition reflects the contribution of ES with no additional boost from REP.

Finally, we probed whether these findings were mediated by the type of attitude object (novel social, including Studies 1 and 2; nonsocial, including Study 3; and existing social, including Studies 4 and 5). We found no differences across novel and existing social groups, with the ES and combined conditions outperforming the REP condition in both cases, *p*s $< .05$ by bootstrap and the combined condition offering no added value over the ES condition, $p > .05$. For nonsocial targets, the pattern of results was slightly different, with the only significant difference emerging between the REP and combined conditions, $p < .05$ by bootstrap.

General Discussion

In this project we conducted six experiments to answer two, as yet unexplored, questions about the acquisition and change of implicit attitudes. Building on large numbers of previous studies demonstrating that both REP and ES can lead to changes in implicit evaluation, we investigated whether the two learning modalities would create implicit attitudes on par with each other. Additionally, our experiments were designed to probe whether the learning effects created by REP and ES are additive or redundant.

Regarding the first question of relative effectiveness, we made two contradictory predictions based on existing literature. On the basis of association formation theories positing that implicit attitudes should shift gradually over time (Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006; Strack & Deutsch, 2004) and empirical demonstrations of the superiority of REP in determining implicit attitudes when REP and ES are contradictory (DeCoster et al., 2006; Gregg et al., 2006; Moran & Bar-Anan, 2013; Rydell et al., 2006; Rydell & McConnell, 2006), we expected that REP might create stronger implicit attitudes than ES. Conversely, propositional theories (De Houwer, 2009, 2014; Hughes et al., 2011; C. J. Mitchell et al., 2009) and some empirical findings (Gast & De Houwer, 2013) suggested that REP and ES should produce learning effects of similar magnitude.

In fact, focusing on the incongruent conditions that allowed ample opportunity for attitudinal shift, we found a robust pattern of data that we had not a priori predicted based on either theoretical perspective. Unexpectedly and despite unusually high-powered designs, REP created significant learning effects in only three out of five experiments, and the effect sizes produced by REP displayed considerable variability across studies. Even more surprisingly, ES resulted in consistently large shifts in implicit attitudes that were never inferior to and, in three out of five studies, even significantly exceeded the attitude change created by REP. These findings were unmediated by the domain of learning (social vs. nonsocial) and the status of the targets (well-known vs. novel). Moreover, they remained robust in the face of procedural variations designed to remove potential confounds, including time on intervention, levels of construal, and expectancy effects (Studies 6A–6C).

This pattern of results is surprising for a number of reasons. First, unlike in previous experiments where information about the attitude objects was highly diagnostic (Cone & Ferguson, 2015;

Mann & Ferguson, 2015), participants in the present studies were merely informed that the targets would be paired with pleasant or unpleasant drawings, which is hardly on par with learning that a person is a child molester, the manipulation used by Cone and Ferguson (2015). Second, with new stimuli appearing every few seconds, if anything, the REP condition should have been more attention-grabbing, and thus more powerful, than the ES condition. Finally, our findings are difficult to reconcile with association formation theories of implicit attitudes, which have dominated the field ever since the beginnings of implicit social cognition research (Bargh, 1989; Fazio et al., 1986; Greenwald & Banaji, 1995) well into the present (Gawronski & Bodenhausen, 2006; Rydell et al., 2006; Strack & Deutsch, 2004).

For the second issue concerning the combined effects of REP and ES, we had envisaged three possible outcomes. Strict association formation theories (Rydell et al., 2006; Strack & Deutsch, 2004) predicted that ES should, at best, have minimal effects on implicit attitudes, rendering this question irrelevant; dual-process theories allowing for interactions between the explicit and implicit systems (Gawronski & Bodenhausen, 2006) predicted that the combination of REP and ES should result in additive, or at least subadditive, effects of each isolated learning modality; and propositional models (De Houwer, 2014; C. J. Mitchell et al., 2009) predicted that the combination of both attitude induction modalities should produce learning on par with their separable effects. Regarding this second question, the prediction made by propositional models was clearly borne out: The ES and combined modalities were statistically indistinguishable from each other across all five studies, and the combined condition did not produce larger learning effects than ES even when (a) REP created sizable learning in isolation (Studies 1, 3, and 4) and (b) when the combined condition included the same number of stimulus pairings as the REP condition (Study 6A). Ceiling effects do not offer a convincing explanation for this finding; it emerged consistently irrespective of the strength of the prevailing attitude, including in cases where learning occurred in the opposite direction from a strong initial evaluation (e.g., for participants learning that elderly people or foreigners are good).

These findings are considerably easier to reconcile with propositional than with association formation theories of implicit attitudes. According to single-process theories (De Houwer, 2009, 2014; C. J. Mitchell et al., 2009), both REP and ES give rise to the same mental representation (e.g., "Squares are good and rectangles are bad"), and given the redundancy observed between the REP and the ES conditions, our data are clearly in line with this prediction. Even though the prediction of equal effects across REP and ES that we derived from single-process theories was not empirically supported, one might argue that the superiority of ES should, after all, not be that surprising from their perspective. For REP and ES to be able give rise to propositional representations about the valence of the attitude objects, participants need to make inferences about the stimuli to which they are exposed. However, the number of inferential steps is not the same across the REP and the ES conditions. In the former case, (a) participants are exposed to stimulus pairings from which (b) they deduce a rule ("Squares are always paired with pleasant; rectangles are always paired with unpleasant"), from which (c) they can infer the intended proposition ("Squares are good and rectangles are bad"). The ES condition, by contrast, does not require participants to undertake step (b), because the rule has already been provided to them. This difference could explain the lower average effectiveness of the REP modality and the greater variability in its success.[19] Even though our data do not conclusively demonstrate the validity of this reasoning, prior work has shown that the ability to correctly recall the instructions is a precondition for successful instructed conditioning effects (Gast & De Houwer, 2013; Van Dessel, De Houwer, Gast, Smith, & De Schryver, 2016). The lack of conscious recollection, of course, precludes any further inferential steps.

The pattern of data observed in our studies is harder to reconcile with association formation theories of implicit attitudes. Strict association formation theories (Rydell & McConnell, 2006; Strack & Deutsch, 2004) posit that explicit and implicit attitudes are subserved by different learning mechanisms. Whereas explicit (i.e., propositional) attitudes can be acquired in one-shot learning episodes, implicit (i.e., associative) attitudes emerge from protracted exposure to contingencies in the environment. Contrary to these theories, our data show that not only are one-shot language-based learning episodes capable of creating implicit attitudes (De Houwer, 2006; Gast & De Houwer, 2013; Gregg et al., 2006; Van Dessel et al., 2015), these implicit attitudes are at least on par with, and usually even stronger than, implicit attitudes created by exposure to stimulus pairings. Thus, these data seem to militate against distinguishing between explicit and implicit attitudes on the basis of learning processes. Association formation theories that allow for interactions between the explicit and the implicit system, such as the associative–propositional evaluation (APE; Gawronski & Bodenhausen, 2006) model, also posit different underlying processes for explicit and implicit evaluation. Under this model, REP shift implicit attitudes directly, whereas ES first create explicit attitudes, and those explicit attitudes, in turn, influence implicit evaluation. Although this reasoning is not fully incompatible with our results, three particular aspects of the data seem to pose somewhat of an explanatory challenge for it. First, if REP and ES create implicit attitudes via separate cognitive pathways, why are their effects redundant? Second, how does the APE model explain the fact that overall, ES outperformed REP, even though for the former, an additional cognitive step is posited to be intervening between learning and evaluation? Finally, how is it possible for implicit attitudes to change without corresponding changes in explicit attitudes (Study 4; see also Van Dessel, De Houwer, Gast, Smith, & De Schryver, 2016)?

It should be noted, however, that the results reported here do not directly speak to the content or format of the mental representations created by each of the two learning interventions under consideration. That is, although the redundancy of the two interventions suggests that ES and REP may give rise to

[19] It is possible that the variability in the effectiveness of REP may have to do with the ease with which participants are able to spontaneously generate verbal labels for the two groups of stimuli. For instance, it is conceivable that REP did not cerate significant attitude change in Study 5 because participants failed to impose a verbal label on the quite disparate stimuli representing the foreign category. We thank an anonymous reviewer for pointing out this possibility. In a similar vein, it has been shown that providing verbal labels for social groups greatly facilitates social inference in 4-year-old children (Baron, Dunham, Banaji, & Carey, 2014).

the same mental representations, our studies are agnostic about the question of whether (a) both strengthen unqualified links between conceptual nodes in long-term memory (e.g., between *elderly* and *good*), (b) both form the basis of propositional representations with compositional structure and referential meaning (e.g., "Elderly people are good."), or (c) some other possibility. Examples of one-trial associative learning abound both in the domain of animal learning (Dufort, Guttman, & Kimble, 1954; Jarvik & Essman, 1960) and human learning (Öhman, Eriksson, & Olofsson, 1975; Rock & Heimer, 1959). Thus, even though our results seem to militate against current associative theories from social cognition that posit that the updating of implicit attitudes can only happen in a piecemeal manner via incremental strengthening of unqualified links between conceptual nodes, they are by no means inconsistent with a potential theory of implicit attitudes that would allow for the possibility of the immediate creation or updating of associations based on information presented in either verbal or nonverbal form. To date, however, no such theory of implicit evaluation has been proposed.

Somewhat unexpectedly, our results were not modulated by the novelty of the attitude objects. ES was more effective than REP and the two methods were redundant for both novel and existing social groups. Thus, it seems plausible that Van Dessel et al. (2015) may have failed to obtain implicit attitude change based on approach/avoidance instructions because of the particular outgroup that they selected for investigation (i.e., Blacks). Because of the long history of racial animosity in the United States (Nosek, Banaji, & Jost, 2009), interventions that mitigate bias against other outgroups can easily backfire when used in this particular context. More generally, our results show that ES-based interventions could be more effective than previously thought. Moreover, they caution against treating all outgroups as alike and suggest that different interventions might be differentially effective for different implicit outgroup biases. For instance, the hierarchy of interventions established by Lai et al. (2014) may change if one were to examine a different target group.

Even though our studies unequivocally demonstrate the effectiveness of ES and suggest that learning on the basis of stimulus pairings and verbal information may share the same underlying cognitive mechanism, they do not offer an immediate explanation for why ES are so powerful or through what (shared) mental processes they exert their effect. It is possible that the power of ES may have to do with the explanation outlined above in relation with propositional theories of implicit evaluation. Unlike repeated evaluative pairings, which are experienced over time and might leave room for interpretational ambiguities, ES are compact and, at least in our studies, express unequivocal rules (such as "squares = good" and "rectangles = bad"). Alternatively, the power of ES might not have been because of their compact propositional nature but rather because of the fact that upon reading the instructions, participants mentally simulated (Markman, Klein, & Suhr, 2012) the upcoming stimulus pairings, thus making the learning experience more associative in nature. This hypothesis does not explain, however, (a) why simulated stimulus pairings were more effective in shifting implicit attitudes than experienced stimulus pairings and (b) why we observed the same level of learning across the self-paced (Study 6A) and the fixed-time (Study 2) ES condition, even though if this account is accurate, participants spent almost three times longer simulating stimulus pairings in the latter than in the former.

In summary, this project has demonstrated that, in opposition to predictions, ES are more effective than REP in shifting implicit attitudes and the effects of both learning modalities are fully redundant. These findings have a number of theoretical and possible practical implications. Crucially, they call into question the idea of a match between more conscious modes of thought as using language-based learning on the one hand and less conscious modes of thought being more sensitive to experience-based learning on the other hand. They also show that whatever the driving force behind REP-based learning, the effects of REP in shifting implicit attitudes are subsumed under the effects of ES-based learning. Finally, we hope that our results might inform future work probing the mechanisms behind successful debiasing interventions, both by drawing attention to the potential of language in mitigating implicit bias and by highlighting the importance of taking into account subtle differences among outgroups.

At the same time, one might wonder to what extent these results would have turned out differently given alternative interpretations of some key concepts and concomitant variations in procedures. Stimulus pairings in the world and instructions expressed through language seem to be fundamentally incommensurable, that is, finding a common metric for the amount of information conveyed by each is far from a trivial task. In our studies, we overcame this problem by fixing time spent on intervention across learning conditions. Future studies might fix the strength of implicit attitude to be achieved and measure the amount of learning necessary to reach that prespecified goal. In addition, given the strong psychometric properties of the IAT both in general (Bar-Anan & Nosek, 2014) and specifically in the context of attitude induction via evaluative conditioning (Hofmann et al., 2010; Nosek et al., 2007), all studies reported here used the IAT as their dependent measure. However, considering the substantial amount of nonoverlapping variance observed among implicit attitude measures (Bar-Anan & Nosek, 2014), future work may wish to seek corroborating evidence for these results using a different implicit task. Moreover, attitude strength as measured immediately after learning is not the only possible measure of effectiveness, and other operationalizations could lead to substantive conclusions different from the ones reached here. For instance, one would expect effective interventions to bring about durable learning effects, and there is reason to suspect that attitudes created by ES might decay faster than attitudes created by REP (Hughes & Barnes-Holmes, 2011). Finally, an effective intervention should result in attitudes that are resistant to counterinformation and can be activated effortlessly even in the absence of available cognitive resources. How REP and ES fare against each other along these dimensions remains to be investigated.

References

Baeyens, F., & De Houwer, J. (1995). Evaluative conditioning is a qualitatively distinct form of classical conditioning: A reply to Davey (1994).

Behaviour Research and Therapy, 33, 825–831. http://dx.doi.org/10.1016/0005-7967(95)00021-O

Ballew, C. C., II, & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 17948–17953. http://dx.doi.org/10.1073/pnas.0705435104

Banaji, M. R., & Bhaskar, R. (2000). Implicit stereotypes and memory: The bounded rationality of social beliefs. In D. L. Schacter & E. Scarry (Eds.), *Memory, brain, and belief* (pp. 139–175). Cambridge, MA: Harvard University Press.

Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American Psychologist, 44,* 1185–1193. http://dx.doi.org/10.1037/0003-066X.44.9.1185

Banaji, M. R., & Crowder, R. G. (1991). Some everyday thoughts on ecologically valid methods. *American Psychologist, 46,* 78–79. http://dx.doi.org/10.1037/0003-066X.46.1.78

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods, 46,* 668–688. http://dx.doi.org/10.3758/s13428-013-0410-6

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3–51). New York, NY: Guilford Press.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology, 62,* 893–912. http://dx.doi.org/10.1037/0022-3514.62.6.893

Baron, A. S., Dunham, Y., Banaji, M. R., & Carey, S. (2014). Constraints on the Acquisition of Social Category Concepts. *Journal of Cognition and Development, 15,* 238–268. http://dx.doi.org/10.1080/15248372.2012.742902

Cacioppo, J. T., Petty, R. E., & Morris, K. J. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology, 45,* 805–818. http://dx.doi.org/10.1037/0022-3514.45.4.805

Cantor, N., & Mischel, W. (1979). Prototypicality and personality: Effects on free recall and personality impressions. *Journal of Research in Personality, 13,* 187–205. http://dx.doi.org/10.1016/0092-6566(79)90030-8

Carter, T. J., Ferguson, M. J., & Hassin, R. R. (2011). Implicit nationalism as system justification: The case of the United States of America. *Social Cognition, 29,* 341–359. http://dx.doi.org/10.1521/soco.2011.29.3.341

Cogsdill, E. J., & Banaji, M. R. (2015). Face-trait inferences show robust child–adult agreement: Evidence from three types of faces. *Journal of Experimental Social Psychology, 60,* 150–156. http://dx.doi.org/10.1016/j.jesp.2015.05.007

Cogsdill, E. J., Todorov, A. T., Spelke, E. S., & Banaji, M. R. (2014). Inferring character from faces: A developmental study. *Psychological Science, 25,* 1132–1139. http://dx.doi.org/10.1177/0956797614523297

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. http://dx.doi.org/10.1037/pspa0000014

Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2012). Dissociable neural correlates of stereotypes and other forms of semantic knowledge. *Social Cognitive and Affective Neuroscience, 7,* 764–770. http://dx.doi.org/10.1093/scan/nsr053

Cuddy, A. J. C., & Fiske, S. T. (2004). Doddering but dear: Process, content, and function in stereotyping of older persons. In T. D. Nelson (Ed.), *Ageism. Stereotyping and prejudice against older persons* (pp. 3–26). Cambridge, MA: MIT Press.

Cunningham, W. A., Nezlek, J. B., & Banaji, M. R. (2004). Implicit and explicit ethnocentrism: Revisiting the ideologies of prejudice. *Personality and Social Psychology Bulletin, 30,* 1332–1346. http://dx.doi.org/10.1177/0146167204264654

DeCoster, J., Banner, M. J., Smith, E. R., & Semin, G. R. (2006). On the Inexplicability of the Implicit: Differences in the Information Provided by Implicit and Explicit Tests. *Social Cognition, 24,* 5–21. http://dx.doi.org/10.1521/soco.2006.24.1.5

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. http://dx.doi.org/10.1016/j.lmot.2005.12.002

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

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

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

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127,* 853–869. http://dx.doi.org/10.1037/0033-2909.127.6.853

Delgado, M. R., Olsson, A., & Phelps, E. A. (2006). Extending animal models of fear conditioning to humans. *Biological Psychology, 73,* 39–48. http://dx.doi.org/10.1016/j.biopsycho.2006.01.006

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18. http://dx.doi.org/10.1037/0022-3514.56.1.5

Devos, T., & Banaji, M. R. (2005). American = White? *Journal of Personality and Social Psychology, 88,* 447–466. http://dx.doi.org/10.1037/0022-3514.88.3.447

Devos, T., Gavin, K., & Quintana, F. J. (2010). Say "adios" to the American dream? The interplay between ethnic and national identity among Latino and Caucasian Americans. *Cultural Diversity and Ethnic Minority Psychology, 16,* 37–49. http://dx.doi.org/10.1037/a0015868

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33,* 510–540. http://dx.doi.org/10.1006/jesp.1997.1331

Dufort, R. H., Guttman, N., & Kimble, G. A. (1954). One-trial discrimination reversal in the white rat. *Journal of Comparative and Physiological Psychology, 47,* 248–249. http://dx.doi.org/10.1037/h0057856

Dunbar, R. I. M. (1996). *Grooming, gossip, and the evolution of language*. Cambridge, MA: Harvard University Press.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Orlando, FL: Harcourt Brace Jovanovich College Publishers.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69,* 1013–1027. http://dx.doi.org/10.1037/0022-3514.69.6.1013

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. http://dx.doi.org/10.1037/0022-3514.50.2.229

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82,* 878–902. http://dx.doi.org/10.1037/0022-3514.82.6.878

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

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

Gibson, B. (2008). Can evaluative conditioning change attitudes toward mature brands? New evidence from the Implicit Association Test. *Journal of Consumer Research, 35,* 178–188. http://dx.doi.org/10.1086/527341

Glaser, J., & Banaji, M. R. (1999). When fair is foul and foul is fair: Reverse priming in automatic evaluation. *Journal of Personality and Social Psychology, 77,* 669–687. http://dx.doi.org/10.1037/0022-3514.77.4.669

Goren, C. C., Sarty, M., & Wu, P. Y. K. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics, 56,* 544–549. Retrieved from http://pediatrics.aappublications.org/content/56/4/544

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

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

Greenwald, A. G., & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. *Experimental Psychology, 48,* 85–93. http://dx.doi.org/10.1026//0949-3946.48.2.85

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. http://dx.doi.org/10.1037/0022-3514.85.2.197

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. http://dx.doi.org/10.1037/a0015575

Greenwald, A. G., Pratkanis, A. R., Leippe, M. R., & Baumgardner, M. H. (1986). Under what conditions does theory obstruct research progress? *Psychological Review, 93,* 216–229. http://dx.doi.org/10.1037/0033-295X.93.2.216

Gregg, A. P. (2000). *The hare and the tortoise: The origins and dynamics of explicit and implicit attitudes* (Doctorial dissertation). Yale University, New Haven, CT.

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. http://dx.doi.org/10.1037/0022-3514.90.1.1

Grumm, M., Nestler, S., & von Collani, G. (2009). Changing explicit and implicit attitudes: The case of self-esteem. *Journal of Experimental Social Psychology, 45,* 327–335. http://dx.doi.org/10.1016/j.jesp.2008.10.006

Hehman, E., Flake, J. K., & Freeman, J. B. (2015). Static and dynamic facial cues differentially affect the consistency of social evaluations. *Personality and Social Psychology Bulletin, 41,* 1123–1134. http://dx.doi.org/10.1177/0146167215591495

Hill, T., Lewicki, P., Czyzewska, M., & Schuller, G. (1990). The role of learned inferential encoding rules in the perception of faces: Effects of nonconscious self-perpetuation of a bias. *Journal of Experimental Social Psychology, 26,* 350–371. http://dx.doi.org/10.1016/0022-1031(90)90044-M

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

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31,* 1369–1385. http://dx.doi.org/10.1177/0146167205275613

Hughes, S., & Barnes-Holmes, D. (2011). On the formation and persistence of implicit attitudes: New evidence from the Implicit Relational Assessment Procedure (IRAP). *The Psychological Record, 61,* 391–410.

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

Jarvik, M. E., & Essman, W. B. (1960). A simple one-trial learning situation for mice. *Psychological Reports, 6,* 290. http://dx.doi.org/10.2466/pr0.1960.6.2.290

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81,* 774–788. http://dx.doi.org/10.1037/0022-3514.81.5.774

Kim, D.-Y. (2003). Voluntary controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly, 66,* 83–96. http://dx.doi.org/10.2307/3090143

Kurdi, B., Lozano, S., & Banaji, M. R. (2016). Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods, 48,* 1–14. http://dx.doi.org/10.3758/s13428-016-0715-3

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J.-E. L., Joy-Gaba, J. A., . . . Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General, 143,* 1765–1785. http://dx.doi.org/10.1037/a0036260

Levey, A. B., & Martin, I. (1975). Classical conditioning of human 'evaluative' responses. *Behaviour Research and Therapy, 13,* 221–226. http://dx.doi.org/10.1016/0005-7967(75)90026-1

Levy, B. R., & Banaji, M. R. (2002). Implicit ageism. In T. D. Nelson (Ed.), *Ageism: Stereotyping and prejudice against older persons* (pp. 49–75). Cambridge, MA: MIT Press.

Lewicki, P. (1986). Processing information about covariations that cannot be articulated. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 135–146. http://dx.doi.org/10.1037/0278-7393.12.1.135

Mann, T. C., & Ferguson, M. J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *Journal of Personality and Social Psychology, 108,* 823–849. http://dx.doi.org/10.1037/pspa0000021

Markman, K. D., Klein, W. M., & Suhr, J. A. (Eds.). (2012). *Handbook of imagination and mental simulation*. New York, NY: Psychology Press.

Mitchell, C. J., Anderson, N. E., & Lovibond, P. F. (2003). Measuring evaluative conditioning using the Implicit Association Test. *Learning and Motivation, 34,* 203–217. http://dx.doi.org/10.1016/S0023-9690(03)00003-1

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

Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 15238–15243. http://dx.doi.org/10.1073/pnas.232395699

Moran, T., & Bar-Anan, Y. (2013). The effect of object-valence relations on automatic evaluation. *Cognition and Emotion, 27,* 743–752. http://dx.doi.org/10.1080/02699931.2012.732040

Norris, C. J., Chen, E. E., Zhu, D. C., Small, S. L., & Cacioppo, J. T. (2004). The interaction of social and emotional processes in the brain.

Journal of Cognitive Neuroscience, 16, 1818–1829. http://dx.doi.org/10.1162/0898929042947847

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134,* 565–584. http://dx.doi.org/10.1037/0096-3445.134.4.565

Nosek, B. A. (2007). Implicit–explicit relations. *Current Directions in Psychological Science, 16,* 65–69. http://dx.doi.org/10.1111/j.1467-8721.2007.00477.x

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics, 6,* 101–115. http://dx.doi.org/10.1037/1089-2699.6.1.101

Nosek, B. A., Banaji, M. R., & Jost, J. T. (2009). The politics of intergroup attitudes. In J. T. Jost, A. C. Key, & H. Thorisdottir (Eds.), *Social and psychological bases of ideology and system justification* (pp. 480–506). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195320916.003.020

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31,* 166–180. http://dx.doi.org/10.1177/0146167204271418

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., . . . Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18,* 36–88. http://dx.doi.org/10.1080/10463280701489053

Öhman, A., Eriksson, A., & Olofsson, C. (1975). One-trial learning and superior resistance to extinction of autonomic responses conditioned to potentially phobic stimuli. *Journal of Comparative and Physiological Psychology, 88,* 619–627. http://dx.doi.org/10.1037/h0078388

Olivola, C. Y., & Todorov, A. T. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior, 34,* 83–110. http://dx.doi.org/10.1007/s10919-009-0082-1

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

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin, 32,* 421–433. http://dx.doi.org/10.1177/0146167205284004

Olsson, A., & Phelps, E. A. (2004). Learned fear of "unseen" faces after Pavlovian, observational, and instructed fear. *Psychological Science, 15,* 822–828. http://dx.doi.org/10.1111/j.0956-7976.2004.00762.x

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience, 10,* 1095–1102. http://dx.doi.org/10.1038/nn1968

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 11087–11092. http://dx.doi.org/10.1073/pnas.0805664105

Pavlov, I. P. (1927). *Conditioned reflexes. An investigation of the physiological activity of the cerebral cortex*. Oxford, United Kingdom: H. Milford.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31,* 109–130. http://dx.doi.org/10.1017/S0140525X08003543

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. http://dx.doi.org/10.1177/0146167211400423

Petty, R. E., & Briñol, P. (2010). Attitude change. In R. F. Baumeister & E. J. Finkel (Eds.), *Advanced social psychology: The state of the science* (pp. 217–259). Oxford, United Kingdom: Oxford University Press.

Petty, R. E., & Cacioppo, J. T. (1979). Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses. *Journal of Personality and Social Psychology, 37,* 1915–1926. http://dx.doi.org/10.1037/0022-3514.37.10.1915

Petty, R. E., Wells, G. L., & Brock, T. C. (1976). Distraction can enhance or reduce yielding to propaganda: Thought disruption versus effort justification. *Journal of Personality and Social Psychology, 34,* 874–884. http://dx.doi.org/10.1037/0022-3514.34.5.874

Pierce, J. C., & Rose, D. D. (1974). Nonattitudes and American public opinion: The examination of a thesis. *American Political Science Review, 68,* 626–649. http://dx.doi.org/10.2307/1959509

Pratto, F., Sidanius, J., & Levin, S. (2006). Social dominance theory and the dynamics of intergroup relations: Taking stock and looking forward. *European Review of Social Psychology, 17,* 271–320. http://dx.doi.org/10.1080/10463280601055772

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

Rajagopal, P., Raju, S., & Unnava, H. R. (2006). Differences in the cognitive accessibility of action and inaction regrets. *Journal of Experimental Social Psychology, 42,* 302–313. http://dx.doi.org/10.1016/j.jesp.2005.05.003

Ranganath, K. A., & Nosek, B. A. (2008). Implicit attitude generalization occurs immediately; explicit attitude generalization takes time. *Psychological Science, 19,* 249–254. http://dx.doi.org/10.1111/j.1467-9280.2008.02076.x

Rescorla, R. A., & Holland, P. C. (1982). Behavioral studies of associative learning in animals. *Annual Review of Psychology, 33,* 265–308. http://dx.doi.org/10.1146/annurev.ps.33.020182.001405

Rock, I., & Heimer, W. (1959). Further evidence of one-trial associative learning. *The American Journal of Psychology, 72,* 1–16. http://dx.doi.org/10.2307/1420207

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. http://dx.doi.org/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. http://dx.doi.org/10.1111/j.1467-9280.2006.01811.x

Savitsky, K., Medvec, V. H., & Gilovich, T. (1997). Remembering and regretting: The Zeigarnik effect and the cognitive availability of regrettable actions and inactions. *Personality and Social Psychology Bulletin, 23,* 248–257. http://dx.doi.org/10.1177/0146167297233004

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

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308,* 1623–1626. http://dx.doi.org/10.1126/science.1110589

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences, 12,* 455–460. http://dx.doi.org/10.1016/j.tics.2008.10.001

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117,* 440–463. http://dx.doi.org/10.1037/a0018963

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. http://dx.doi.org/10.1016/j.jesp.2015.11.002

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. *Experimen-*

tal Psychology, 62, 161–169. http://dx.doi.org/10.1027/1618-3169/a000282

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17,* 592–598. http://dx.doi.org/10.1111/j.1467-9280.2006.01750.x

Wood, W. (2000). Attitude change: Persuasion and social influence. *Annual Review of Psychology, 51,* 539–570. http://dx.doi.org/10.1146/annurev.psych.51.1.539

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

Zeigarnik, B. (1927). Über das Behalten von erledigten und unerledigten Handlungen [On finished and unfinished tasks]. *Psychologische Forschung, 9,* 1–85.

Received January 19, 2016
Revision received September 11, 2016
Accepted October 25, 2016 ■