American Psychological Association

Journal of Personality and Social Psychology:
Attitudes and Social Cognition

0022-3514/19/$12.00

http://dx.doi.org/10.1037/pspa0000151

Attitude Change via Repeated Evaluative Pairings Versus Evaluative Statements: Shared and Unique Features

Benedek Kurdi and Mahzarin R. Banaji
Harvard University

When tested immediately, evaluative statements (ES; verbal information about upcoming categories and their positive/negative attributes) surprisingly shift implicit (IAT) attitudes more effectively than repeated evaluative pairings (REP; actual pairing of category members with positive/negative attributes). The present project (total $N = 5{,}317$) explored the shared and unique features of these two attitude change modalities by probing (a) commonalities visible in the extent to which propositional inferences created by ES infiltrate REP learning and (b) differences visible in performance of ES and REP learning over time. In REP, the number of stimulus pairings (varied parametrically from 4 to 24) produced no effect (Study 1), but verbally describing stimulus pairings as diagnostic versus nondiagnostic did modulate learning (Study 2), suggesting that even REP give rise to some form of propositional representation. On the other hand, learning from ES decayed quickly, whereas learning from REP remained stable over time both within an immediate session of testing (Study 3) and following a 15-min delay (Study 4), revealing a difference between these two forms of learning. Beyond their theoretical import, these findings may inform interventions designed to produce short- and long-term change in implicit attitudes.

Keywords: associative learning, attitude change, evaluative conditioning, implicit attitudes, propositional learning

Humans are confronted with a fundamental and consequential type of decision numerous times a day: They must decide whether to approach or to avoid each of the myriad stimuli that they encounter in their environment. According to an influential idea as old as the concept itself, choices about approach and avoidance are driven by *attitudes*, that is, evaluations of stimuli along a positive–negative valence dimension: Positive evaluations result in approach behaviors and negative evaluations result in avoidance behaviors (Allport, 1935; Cacioppo, Harkins, & Petty, 1981; Eagly & Chaiken, 1998; Lewin, 1935). If decisions about approach and avoidance are to adaptively guide behavior, they must be made not only accurately but also quickly and effectively. As such, approach–avoidance decisions may, at least in part, be driven by *implicit attitudes*, or evaluations that are activated automatically upon encountering a stimulus (Bargh, Chaiken, Govender, & Pratto, 1992; Devine, 1989; Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Greenwald & Banaji, 1995). Such automatically activated evaluations are usually measured via response interference tasks, as opposed to *explicit attitudes,* which are measured via self-report.[1]

Given their role in shaping affect, cognition, and behavior, the question of how implicit evaluations shift in the face of novel information is both of basic theoretical interest and of relevance to the more applied endeavor of producing enduring shifts in implicit evaluations of existing categories, such as social groups. In attempting to characterize implicit attitude change, the present work takes as its starting point two fundamentally different approaches that have been used to capture a wide range of phenomena in human learning, including concept learning (Goodman, Tenenbaum, Feldman, & Griffiths, 2008; Kruschke, 1992), causal learning (Dickinson, 2001; Gopnik et al., 2004), and language acquisi-

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

Parts of the data reported in this article have been presented at the Fifth European Meeting on the Psychology of Attitudes, Cologne, Germany, in July 2016; the 18th Annual Meeting of the Society for Personality and Social Psychology, San Antonio, TX, in January 2017; the Sixth Annual Symposium for Boston-Area Graduate Students in Psychology, Boston, MA, in April 2017; and the 18th General Meeting of the European Association of Social Psychology, Granada, Spain, in July 2017. All data files, analysis scripts, and stimuli used in this project are available for download from the Open Science Framework (https://osf.io/serq4/). We thank Sarah Ryan for assistance with study coding as well as the Dean's Competitive Fund for Promising Scholarship and the Institute for Quantitative Social Science at Harvard University for financial support.

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

[1] In line with Fazio and colleagues (e.g., Fazio & Olson, 2003), we see the distinction between explicit and implicit attitudes as located primarily at the level of measures and do not make a priori assumptions about the nature of the mental representations or learning processes underlying implicit evaluations. In fact, we believe that the view advocated by De Houwer (e.g., De Houwer, 2007, 2018), according to which evaluative conditioning should be conceptualized as an effect rather than a process, has significantly advanced evaluative conditioning research by allowing for the possibility that implicit measures may reflect propositional processes of attitude change. The present work continues in that tradition.

tion (Griffiths & Kalish, 2010; Rumelhart & McClelland, 1987): the *associative approach* and the *inferential approach*.[2]

The associative approach relies on the premise that human learning arises from the incremental updating of associative representations in long-term memory as a result of repeated experiences with the environment. The associative approach encompasses powerful and diverse ideas such as neural networks (Rumelhart, Hinton, & Williams, 1986), the Rescorla–Wagner model of Pavlovian conditioning (Rescorla & Wagner, 1972), and temporal difference learning (Sutton & Barto, 1998). Even though specific instantiations of this approach obviously differ in their details of implementation, they are unified by a few core ideas, including (a) relatively slow and iterative updating that progressively reduces the difference between an existing representation and a target representation based on prediction errors; (b) learning as an informationally encapsulated process that preferentially relies on certain kinds of input while ignoring others (see Fodor, 1983); and (c) long-term memory as a collection of associative strengths.

By contrast, the inferential approach relies on the radically different premise that human learning arises from a process of theory testing in which the learner entertains multiple competing hypotheses about some problem and revises her beliefs about the plausibility of each hypothesis upon encountering new data. Like the associative approach, the inferential approach encompasses powerful and diverse ideas such as the child as scientist (Gopnik, 1996), Bayesian models of human cognition (Tenenbaum, Kemp, Griffiths, & Goodman, 2011), and the pedagogical view of learning and teaching (Csibra & Gergely, 2009). Again, in spite of important differences in implementation, a few core ideas seem to unify the inferential approach, including (a) the possibility of drawing quick inferences from sparse data; (b) learning as an informationally promiscuous process that integrates information from all relevant sources (see Fodor, 1983); and (c) long-term memory as a store of symbolic representations.

Ever since its inception, the very concept of implicit evaluations has been intimately linked with the idea of associative learning and representation. Early investigations of implicit race attitudes drew heavily from the idea of spreading activation in associative networks (e.g., Fazio et al., 1986) and relied on methods that had been implemented to study conceptual associations in nonsocial category representation (Neely, 1976). In fact, one of the most widely applied measures of implicit evaluations, also used in the present studies, has the term "association" in its name (Greenwald, McGhee, & Schwartz, 1998). As such, it should not be surprising that the study of implicit social cognition has traditionally been dominated by associative theories (E. R. Smith & DeCoster, 2000; McConnell & Rydell, 2014; Rydell & McConnell, 2006; Strack & Deutsch, 2004). These theories may be regarded as applications of the associative approach to human learning to the more specific problem of implicit attitude acquisition and change. Even though these theories differ from each other in numerous details, they converge on a generally consistent view that implicit attitudes should shift (a) slowly and incrementally, (b) preferentially as a result of direct experience with the environment, and (c) by creating conceptual associations in long-term memory.

Specifically, under associative theories, if implicit attitudes toward a social group, such as the Laapians (Gregg, Seibt, & Banaji, 2006), are to be shifted in a positive direction, members of this group should be paired with intrinsically positive stimuli for a sufficient (large) number of trials, as a result of which the links between the conceptual nodes laapian and good will be strengthened in long-term memory. Associative theories describe implicit attitude change as slow and incremental: Smith and DeCoster (2000) mention a memory system that "build[s] up knowledge slowly and incrementally" (p. 111); Strack and Deutsch (2004) talk about a simple memory system that "slowly forms enduring, nonpropositional representations . . . over many learning trials" (pp. 223–224); and, finally, according to Rydell and McConnell (2006), "implicit attitudes form and change through the use of slow-learning, associative reasoning" (p. 995; see also McConnell & Rydell, 2014).

The idea of informational encapsulation is also reflected in associative theories of implicit evaluation. According to Smith and DeCoster (2000), "the two [associative vs. propositional] processing modes tap separate databases that represent knowledge in different formats," with the associative mode drawing "solely on patterns built up over time in the slow learning memory system" (p. 127). The same authors rule out the possibility that a newly learned rule may be able to affect associative processing. Similarly, McConnell and Rydell (2014) posit "two dissociable systems of knowledge" (p. 204): a slow-learning system relying on associative information and a fast-learning system relying on verbal information (Rydell & McConnell, 2006). In the same vein, Strack and Deutsch (2004) hypothesize that whereas "explicit measures tap into people's knowledge or beliefs, implicit measures tap into their associative structures" (p. 239).

Finally, and related to both the idea of slow learning and informational encapsulation, associative theories propose that implicit and explicit evaluations are subserved by separate memory systems. For instance, Smith and DeCoster (2000) argue that whereas implicit attitudes are "represented by associations built up in a connectionist distributed memory . . ., explicit beliefs are symbolically represented" (p. 122). Strack and Deutsch (2004) propose a different kind of memory dissociation, with implicit evaluations relying on long-term memory and explicit evaluations relying on a temporary memory store. However, in spite of such differences, all associative theories reviewed here share the basic premise that implicit attitudes arise from conceptual associations stored in long-term memory. (A more exhaustive list of verbatim quotations from relevant theoretical papers is provided on the Open Science Framework; https://osf.io/serq4/).

More recently, *propositional theories* (De Houwer, 2009, 2014, 2018; Mitchell, De Houwer, & Lovibond, 2009), which can be conceptualized as applying the inferential approach to human learning to implicit evaluation, have begun to challenge the primacy of the associative view. Propositional accounts are fundamentally different from associative accounts in the learning pro-

[2] We do not argue that the two approaches are mutually exclusive. For instance, a recent proposal posits that Pavlovian conditioning operates not over observable cues and outcomes but rather over latent causes inferred by the learner (Gershman, Norman, & Niv, 2015). Moreover, given that neural nets are universal function approximators, they can be used to implement symbolic (e.g., Bayesian) approaches to learning. However, these observations do not invalidate the fact that the two approaches have traditionally been seen as competing in providing a description and explanation of human learning.

cesses and mental representations posited to underlie implicit evaluations. Specifically, under the propositional view, implicit and explicit attitudes are hypothesized to differ from each other solely in terms of the level of automaticity with which evaluative representations are activated, with implicit measures such as the IAT (Greenwald et al., 1998) forcing participants to respond relatively automatically and explicit measures imposing no such requirement.

When it comes to learning, the propositional account, unlike its associative counterparts, posits that implicit evaluations can shift (a) quickly, without the need for incremental updating, (b) in an informationally promiscuous manner, taking into account all available evidence (including both stimulus pairings encountered in the environment and verbal information), and (c) by creating symbolic and propositional representations in long-term memory. Specifically, in a clear deviation from associative accounts, De Houwer (2009) suggests that "the generation and evaluation of propositions will [not] always be slow; if necessary, it can be done expeditiously" (p. 10), and such propositions may, in turn, be activated automatically on implicit measures of evaluation such as the IAT used in the present studies. Moreover, according to De Houwer (2014), "propositions about events can be formed not only on the basis of the repeated experience of those events but also as the result of a single instruction or inference concerning those events" (p. 344). This idea of informational promiscuity is in opposition to the core tenet of associative theories according to which implicit attitudes should be exclusively, or at least preferentially, responsive to stimulus pairings experienced in the environment. Finally, De Houwer (2018) also does away with the idea of a separate associative memory store driving responding on implicit measures. In fact, he proposes that even in situations where the procedure itself may be described as associative, including repeated presentations of stimulus pairings, such pairings "can influence liking only after a proposition has been formed about the relation between the stimuli" (p. 3).

What Kind of Information Is Most Effective in Shifting Implicit Evaluations?

Associative and propositional theories of implicit evaluation both predict that repeated evaluative pairings (REP) of category members with intrinsically valenced stimuli should shift implicit attitudes. Indeed, evidence in line with this prediction has been produced by multiple labs and involving many specific instantiations of the REP intervention (Gibson, 2008; Grumm, Nestler, & Collani, 2009; Hughes & Barnes-Holmes, 2011; Mitchell, Anderson, & Lovibond, 2003; Olson & Fazio, 2001, 2002, 2006; Prestwich, Perugini, Hurling, & Richetin, 2010; for a meta-analysis see Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). However, associative and propositional theories differ in their predictions about the effectiveness of evaluative statements (ES), that is, informing participants about upcoming stimulus pairings without ever presenting such stimulus pairings. Surprisingly, and contrary to the informational encapsulation idea central to many associative theories, a series of recent studies have demonstrated significant changes in implicit evaluation as a result of purely verbal instructions, without any direct experience involving intrinsically valenced stimuli (De Houwer, 2006; Gast & De Houwer, 2013; Van Dessel, De Houwer, Gast, & Smith, 2015; Van Dessel, De Houwer, Gast, Smith, & De Schryver, 2016; Van Dessel, Gawronski, Smith, & De Houwer, 2017; Van Dessel, Mertens, Smith, & De Houwer, 2017).

If REP and ES both have the ability to shift implicit attitudes, such results raise the question of relative effectiveness, which informs both the theoretical debate regarding the nature of the learning processes from which implicit evaluations emerge and efforts aimed at creating change in real-world evaluations. In line with such theoretical and practical interest, recent work has investigated whether stimulus pairings experienced in the environment or mere verbal instructions produce stronger evaluative learning (Kurdi & Banaji, 2017). In the REP condition of Kurdi and Banaji's studies, participants were exposed to pairings of members of each target category (conditioned stimuli [CS]) with intrinsically positive or negative images (unconditioned stimuli [US]). In the ES condition, participants were informed about upcoming stimulus pairings without actual exposure. Evaluative statements were consistently more effective in shifting implicit attitudes than repeated evaluative pairings, even in the face of procedural changes intended to enhance learning in the REP condition or weaken learning in the ES condition. Such results are difficult to explain within the framework of associative theories, which propose that implicit evaluations should rely exclusively (E. R. Smith & DeCoster, 2000; Rydell & McConnell, 2006; Strack & Deutsch, 2004) or at least preferentially (McConnell & Rydell, 2014) on stimulus pairings experienced in the environment.

By What Kind(s) of Process Is Evaluative Learning Produced?

Beyond the relative effectiveness of REP versus ES in shifting implicit evaluations, Kurdi and Banaji (2017) also explored the learning effects produced by a combination of the ES and REP interventions. They found that a third learning condition in which descriptions of stimulus pairings were followed by the presentation of actual stimulus pairings (ES + REP) never significantly outperformed the ES condition in isolation, suggesting that direct experience with stimulus pairings did not add any value to purely verbal instructions. This result is unsurprising from the perspective of propositional theories: If both REP and ES give rise to the same propositional inference about the attitude object, their effects on implicit evaluations should be redundant. However, it is difficult to accommodate this result under associative theories given that these theories posit that implicit evaluations should not be influenced by purely verbal manipulations (E. R. Smith & DeCoster, 2000; Rydell & McConnell, 2006; Strack & Deutsch, 2004).

Kurdi and Banaji (2017) concluded that, in spite of reasonable expectations to the contrary, the data were more closely aligned with propositional theories of implicit evaluation. Specifically, their findings strongly suggest that in humans, higher-level reasoning can insert itself into learning even from repeated presentations of pairings between categories and attributes. In other words, even stringent conditions of (supraliminal) associative learning may involve inferential processes that give rise to propositional representations. Such a conclusion is all the more surprising given that learning from repeated evaluative pairings has often been characterized as mediated by low-level stimulus-driven processes (Baeyens, Eelen, Crombez, & van den Bergh, 1992; Baeyens, Vansteenwegen, & Hermans, 2009; Gawronski & Bodenhausen,

2006; Gawronski, Balas, & Creighton, 2014; Levey & Martin, 1975; Martin & Levey, 1978; Rydell & McConnell, 2006; Strack & Deutsch, 2004).

In light of this unexpected yet robust finding, replicated six times by Kurdi and Banaji (2017), the first part of the present paper reports two further tests of the process by which repeated evaluative pairings lead to changes in implicit evaluation. In Study 1, we investigated the effects of the number of stimulus pairings presented during learning. In Study 2, we probed whether a purely verbal manipulation (describing the stimulus pairings as diagnostic vs. nondiagnostic of the underlying nature of the target categories) significantly modulated learning effects from repeated evaluative pairings. We chose to study the effects of these two potential moderators of evaluative learning from REP for three reasons. First, whether learning is (a) fast or slow and (b) informationally promiscuous (i.e., drawing on all available data) or informationally encapsulated (i.e., restricted to a specific kind of informational input; see Fodor, 1983) is of basic theoretical interest. Findings regarding these features of the learning process stand to constrain any future theory of implicit attitude acquisition. Second, empirical support for both sets of predictions (quick and informationally promiscuous vs. slow and informationally encapsulated learning) seemed conceivable in light of existing evidence. Third, current associative and propositional theories of implicit evaluation differ in terms of the degree to which they are compatible with different patterns of empirical data.

Number of Stimulus Pairings Presented During Learning

Existing empirical evidence regarding the effect of number of CS–US pairings on evaluative learning is mixed. A number of studies have investigated this issue using explicit measures of attitude (Baeyens et al., 1992; Levey & Martin, 1987; Stuart, Shimp, & Engle, 1987). Baeyens and colleagues (1992) have found a quadratic effect of number of stimulus pairings for positively valenced USs and a linear effect for negatively valenced USs. By contrast, Levey and Martin (1987) concluded that number of stimulus pairings was inconsequential; however, no statistical evidence in favor of this proposition was put forth. Finally, Stuart and colleagues (1987) found nonsignificant linear trends using some measures of attitude but no effect using others. However, whether the effects obtained using explicit measures of evaluation (to the extent that such effects are, indeed, reliable) would generalize to implicit attitudes is an open empirical question, and in the present research we will have the opportunity to test it.

Evidence obtained using implicit measures of attitude has been similarly inconclusive. Rydell, McConnell, Strain, Claypool, and Hugenberg (2007) found that the number of counterattitudinal statements linearly predicted the amount of change in implicit attitudes toward a novel social target in an impression formation task. However, it should be noted that this finding was obtained using a learning intervention involving (a) verbal, rather than pictorial, stimuli, (b) a single novel target rather than a novel group, and (c) positive and negative feedback rather than passive exposure to stimuli. As such, the results obtained by Rydell et al. (2007) may not readily generalize to the REP manipulation implemented here. More relevant for the present purposes, Hu, Gawronski, and Balas (2017) observed no statistically significant difference in implicit attitudes created as a result of eight versus 24 stimulus pairings and the meta-analysis by Hofmann et al. (2010) found no significant influence of the number of stimulus pairings presented on evaluative conditioning effects. However, (a) the meta-analysis did not examine explicit and implicit attitudes separately, (b) by definition, meta-analysis can provide only correlational, but not causal, evidence, and (c) lack of statistical significance in a frequentist framework should not be accepted as positive evidence for the absence of an effect (Rouder, Speckman, Sun, Morey, & Iverson, 2009; Wilson & Miller, 1964).

When it comes to current associative and propositional theories of implicit evaluation, both sets of theories predict that repeated evaluative pairings should shift implicit attitudes. However, each set of theories makes unique predictions about how such learning should unfold as the number of stimulus pairings increases. As reviewed above, associative theories of implicit evaluation share the core idea, also present in more general associative approaches to human learning, that implicit attitudes should shift slowly and incrementally. If this is, indeed, the case, the number of stimulus pairings presented in the REP condition should modulate learning effects: A relatively small number of stimulus pairings may not produce any change, and increasing numbers of stimulus pairings should result in increasingly powerful learning effects. By contrast, in line with more general ideas about humans' ability to make quick and accurate inferences even from noisy and sparse observations, the propositional account allows for the possibility that asymptotic performance may be reached after only a few stimulus pairings. As such, repeated presentations of redundant evaluative information may not result in additional learning.

Diagnosticity

Prior work has demonstrated that a single piece of extremely negative information deemed highly diagnostic of a person's character (e.g., being a child molester or someone who enjoys torturing animals) is sufficiently powerful to reverse previously positive implicit attitudes formed as a result of a hundred behavioral statements (Cone & Ferguson, 2015). But can a purely verbal manipulation of diagnosticity (describing stimulus pairings as reflecting the underlying true nature of the target groups vs. having been randomly generated by the computer) have an effect on learning from repeated evaluative pairings? Existing empirical evidence on the issue is contradictory.

Some studies have provided evidence for the idea that implicit attitude acquisition from repeated evaluative pairings is informationally promiscuous in that it takes into account all available information, regardless of its format. As such, describing stimulus pairings as invalid (Peters & Gawronski, 2011) or as representing an opposition, rather than equivalency, relationship (Zanon, De Houwer, Gast, & Smith, 2014) has been found to significantly modulate evaluative learning as reflected by implicit attitude measures. Moreover, verbally describing stimulus pairings as unrelated, predictive, or causal also seems to modulate implicit attitude acquisition (Hughes, Ye, Van Dessel, & De Houwer, 2018). These results suggest that, just like validity or type of relationship, diagnosticity should modulate learning effects from repeated evaluative pairings.

However, in other studies, implicit attitude acquisition from repeated evaluative pairings appears to be relatively information-

ally encapsulated. For instance, Rydell, McConnell, Mackie, and Strain (2006) found that implicit attitudes toward a novel target uniquely reflected subliminally presented REP, with no effect of supraliminally presented ES of the opposite valence (but see Heycke, Gehrmann, Haaf, & Stahl, 2018). Similarly, implicit attitudes do not consistently reflect verbal information about the nature of the relationship between upcoming stimulus pairings in a REP paradigm (Hu et al., 2017; Moran & Bar-Anan, 2013), and even when they do, CS–US contingencies can still have an effect over and above relational information (Moran, Bar-Anan, & Nosek, 2016). This latter set of findings would suggest that REP-based evaluative learning effects should not be modulated by verbal information on diagnosticity.

These two sets of results align with the predictions of propositional versus associative theories, respectively. Propositional theories suggest that because verbal and nonverbal manipulations influence implicit evaluations via the same process of propositional inference, they should interact in producing evaluative learning. By contrast, associative theories posit that implicit attitude acquisition is relatively informationally encapsulated and, as such, should preferentially rely on stimulus pairings encountered in the environment, with little or no effect of verbal manipulations.

What Kind of Information Is Most Effective in Durably Shifting Implicit Evaluations?

To gain further insight into the nature of the evaluative learning processes arising from interventions based on stimulus pairings versus mere verbal instructions, the second set of experiments reported here probed the temporal stability of implicit attitudes created via REP versus ES. Specifically, Study 3 used a new analytic strategy to examine whether implicit attitude strength decayed differentially as a function of learning conditions during a single session of the Implicit Association Test (IAT; Greenwald et al., 1998), administered immediately after learning. Study 4 investigated the effects of time more directly by imposing a 15-min delay between the learning and test phases of the experiment.

Why might time differentially affect implicit attitudes acquired via REP versus ES and what are the theoretical implications of such potential differences? The associative perspective on implicit evaluation posits that implicit attitudes should shift preferentially in response to stimulus pairings experienced in the environment. Although the results of Kurdi and Banaji (2017) seem to contradict this idea, it could be argued that the immediate effects of learning manipulations observed there should be characterized as arising from temporary shifts in concept accessibility rather than as genuine conceptual change. As such, the present studies may offer a more stringent test of the relative effectiveness of evaluative learning based on REP versus ES.

By contrast, under the propositional perspective, implicit and explicit evaluations are generally hypothesized to rely on shared memory representations. As such, studies on differences in episodic memory for personally experienced versus verbally described events may be instructive regarding the effects that should emerge on implicit measures of evaluation.[3] For instance, in a study conducted by Toglia, Shlechter, and Chevalier (1992), some participants directly experienced a staged event, whereas others were exposed to a verbal description of the same event. When surprise tests of recall and recognition for details of the event were administered immediately, they revealed superior performance in the verbal description condition; however, following a delay, performance deteriorated sharply in the verbal description condition but remained unaffected in the direct experience condition (see also Baggett, 1979; Baggett & Ehrenfeucht, 1983; Beentjes & van der Voort, 1991; Larsen & Plunkett, 1987). To the extent that this effect generalizes from explicit measures of episodic memory to implicit measures of evaluative learning, REP should produce more durable changes in implicit evaluation than ES, despite the initial superiority of the latter (for initial evidence see Hughes & Barnes-Holmes, 2011).

To summarize, the associative and propositional views on implicit evaluation both predict that learning from repeated evaluative pairings should be superior to learning from evaluative statements provided that the test of learning is administered following a delay. The associative perspective makes this prediction on the basis of a posited match between the format of the information (stimulus pairings) and the format of the memory representation (conceptual associations). By contrast, the propositional perspective allows for reliance on work on episodic memory to make the same prediction; however, this prediction is made for a completely different reason. Specifically, a crucial difference between REP and ES is that the former requires participants to generate inferences about the valence of each target category. Propositions generated by participants, as opposed to propositions provided to them in verbal form, have been hypothesized to result in superior retention on explicit memory tasks (Baggett, 1979; for related findings see Greenwald & Banaji, 1989; MacLeod & Bodner, 2017; Slamecka & Graf, 1978).

As such, the associative and propositional perspectives differ in their predictions for the combined ES + REP condition: If the effect is based on preferential reliance of implicit evaluations on environmental associations, then the ES + REP condition should behave the same way as REP isolation given that both include the REP manipulation; however, if the effect is based on a difference between experimenter-provided versus participant-generated inferences, then the ES + REP condition should behave the same way as ES in isolation given that the combined condition involves no participant-generated inferences.

Finally, beyond their theoretical implications, Studies 3 and 4 may also shed new light on recent attempts to create durable change in implicit attitudes (Devine, Forscher, Austin, & Cox, 2012; Forscher, Mitamura, Dix, Cox, & Devine, 2017; Lai et al., 2016; Mann & Ferguson, 2015, 2017). Even though temporary malleability in implicit attitudes within a single experimental session has been demonstrated numerous times and using myriad different manipulations (Blair, 2002; Lai et al., 2014), ranging from social influence (Lowery, Hardin, & Sinclair, 2001) to context effects (Wittenbrink, Judd, & Park, 2001) and from exposure to counterstereotypic exemplars (Dasgupta & Greenwald, 2001) to implementation intentions (Stewart & Payne, 2008), successful demonstrations of long-term change in implicit evaluation have been conspicuously absent from the literature (for some notable exceptions in the context of implicit evaluations of single individ-

[3] However, as a reviewer of this work pointed out, propositional theories of implicit evaluation do not explicitly make this prediction. We concur with the reviewer's judgment on this point.

uals see Mann & Ferguson, 2015, 2017). For instance, when Lai et al. (2016) measured the effectiveness of the most immediately successful manipulations following a 24-hr delay, none of them had any appreciable impact on implicit attitudes toward African Americans. Given that the studies reported here used a timeframe of minutes rather than hours or days, they have the potential to offer some insight into the time course of decay in implicit attitudes before learning effects dissipate entirely.

Overview of the Present Project

Implicit attitudes have been robustly demonstrated to shift in response to both repeated evaluative pairings of category members with positive and negative attributes (Gibson, 2008; Grumm et al., 2009; Hughes & Barnes-Holmes, 2011; Kurdi & Banaji, 2017; Mitchell et al., 2003; Olson & Fazio, 2001, 2002, 2006; Prestwich et al., 2010) and evaluative statements, that is, purely verbal manipulations merely describing upcoming stimulus pairings without actual exposure (De Houwer, 2006; Gast & De Houwer, 2013; Kurdi & Banaji, 2017). In an attempt to improve understanding of the processes by which evaluative learning is achieved in response to repeated evaluative pairings, the present project investigated two possible moderators of this learning effect: In Study 1, we varied the number of stimulus pairings to which participants were exposed and, in Study 2, we manipulated whether stimulus pairings were described to participants as randomly generated or reflecting the true underlying character of the target categories. Moreover, in Studies 3 and 4, we probed the temporal stability of learning from REP and ES, as well as from a combined ES + REP condition.

These manipulations were chosen for a number of reasons. First, they provide evidence on basic features of the learning processes underlying implicit attitude acquisition, including its temporal unfolding, informational inputs, and stability over time. As such, the results of the present work have the ability to constrain any future theory of implicit evaluation. Second, existing empirical evidence on the effects of the chosen manipulations was mixed (Studies 1–2) or scarce (Studies 3–4). Third, although no single study or even set of studies may be sufficient to conclusively arbitrate between existing associative versus propositional theories of implicit evaluation, certain patterns of data are nonetheless easier or more difficult to reconcile with different theoretical positions. Fourth, the results of the current work stand to inform efforts that are aimed at designing interventions to produce durable change in implicit attitudes, with or without a desire to advance basic theory.

Study 1

Prior work has demonstrated that repeated evaluative pairings of category members with valenced images can produce shifts in implicit attitudes toward the categories involved (Gibson, 2008; Grumm et al., 2009; Hughes & Barnes-Holmes, 2011; Kurdi & Banaji, 2017; Mitchell et al., 2003; Olson & Fazio, 2001, 2002; Prestwich et al., 2010). The present study probed a key moderator of this effect, investigating whether the number of stimulus pairings presented to participants modulates the strength of the implicit attitudes created. On the one hand, updating may unfold slowly as a result of incrementally reducing the difference between a target representation and the information provided by each stimulus pairing. Alternatively, implicit evaluations may be updated via quick conceptual change, especially in the face of information that provides a clear learning signal.

Method

Participants and design. Volunteers (N = 1,243) were recruited from the Project Implicit educational website (http://implicit.harvard.edu) to participate in the study. In line with the standard scoring procedure (Greenwald, Nosek, & Banaji, 2003), participants who did not complete the IAT (Greenwald et al., 1998) and participants with a response latency of 300 ms or less on at least 10% of IAT trials (N = 43) were excluded from all further analyses. Moreover, consistent with established practice in research on evaluative learning (Van Dessel, De Houwer, Gast, & Smith, 2015; Van Dessel, De Houwer, Gast, et al., 2016; Van Dessel, Gawronski, et al., 2017), participants who provided an inaccurate response on a manipulation check probing explicit recollection of the learning phase (N = 55) were also eliminated from consideration. Participant exclusions resulted in a final sample size of N = 1,145 (N = 693 female and N = 417 male participants, mean age = 39.56, SD = 17.05). This sample size provides .80 power to detect a small effect of r = .08.

All participants underwent an attitude induction procedure using repeated evaluative pairings (REP) of category members with positive and negative images. In a between-subjects design, each participant was randomly assigned to be exposed to a certain number of stimulus pairings, varied parametrically from four to 24. The median number of participants in each cell was 53 (SD = 6.45). There was no evidence for differential attrition across number of stimulus pairing conditions, as indicated by a lack of correlation between number of stimulus pairings and number of participants in each cell of the design, $\rho = -.036$, $p = .875$.

Materials. Line drawings depicting the faces of young White men were used as conditioned stimuli (CS). To enable easy categorization, the faces were manipulated to vary along a salient perceptual dimension: *Long faces* had a length-to-width ratio of 2:1, whereas *square faces* had a length-to-width ratio of 4:3. Faces were selected to serve as CS because they communicate a wealth of social information (Cogsdill, Todorov, Spelke, & Banaji, 2014; Willis & Todorov, 2006) and have a demonstrated ability to serve as the targets of evaluative learning (Hehman, Flake, & Freeman, 2015). Positive unconditioned stimuli (US) included line drawings of a flower, a heart, a cone of ice cream, a sun, and a beach, and negative US included line drawings of a frowny face, a fleeing man, a snake, a terrorist, and an ant. Both the CS and US have been used successfully in the context of attitude induction in previous work (Kurdi & Banaji, 2017).[4]

Procedure and measures. The study was administered entirely online and consisted of (a) a *learning phase* in which implicit attitudes were induced toward the target categories via exposure to REP, with the number of stimulus pairings varied between participants, and (b) a *test phase* in which implicit attitudes and explicit attitudes toward the targets as well as memory for stimulus pairings were assessed.

[4] All stimuli are available for download from the Open Science Framework (OSF; https://osf.io/serq4/).

Learning phase. Participants were informed that in the course of the experiment they would see two types of faces and two types of drawings and were instructed to learn the general relationship between a certain type of face and a certain type of drawing without focusing on particular exemplars.[5] Crucially, the instructions did not use verbal labels to refer to the groups and did not mention stimulus valence. Next, participants were introduced to the full set of conditioned and unconditioned stimuli. Finally, participants were exposed to a standard evaluative conditioning paradigm (Levey & Martin, 1975; Martin & Levey, 1978). Each trial consisted of a CS (a line drawing of a long face or a square face) and a US (a line drawing of an intrinsically positive or negative object) presented side-by-side on the screen for 2,500 ms. Each trial was followed by an intertrial interval of 1,000 ms. Depending on their random assignment to condition, participants received different numbers of stimulus pairings, ranging from four to 24. Previous work revealed an overall implicit preference for long faces over square faces at baseline (Kurdi & Banaji, 2017). To allow for learning to unfold without ceiling effects, the attitude induction manipulation was designed to move participants away from this baseline, that is, square faces were always paired with positive stimuli and long faces were always paired with negative stimuli.[6]

Test phase. In the test phase, participants completed (a) an IAT (Greenwald et al., 1998) measuring their implicit attitudes toward long faces and square faces, followed by (b) two feeling thermometer measures used to construct a relative measure of explicit preferences, and finally (c) a manipulation check probing participants' recollection of their previous learning.

Implicit attitudes. Participants completed a standard five-block IAT (Greenwald et al., 1998) as a measure of implicit attitudes toward the two target groups (individuals with square vs. long faces). In the first practice block (20 trials), participants sorted positively and negatively valenced words.[7] In the second practice block (20 trials), participants sorted the images of square-faced and long-faced individuals that were used as CS in the learning phase. In the first critical block (40 trials), positive words and square-faced individuals were mapped onto one response key, whereas negative words and long-faced individuals were mapped onto the other response key. In the third practice block (20 trials), participants learned the new assignment of target groups to response keys. Finally, in the second critical block (40 trials), the combined task was repeated with the opposite assignment of target groups to valence (long-faced/good and square-faced/bad). Given our interest in relative attitude strengths across different numbers of conditioning trials without comparison to an absolute standard, the order of critical blocks was kept constant across participants, with the congruent (square faces–good/long faces–bad) block always presented first. Performance on the IAT was assessed using the scoring algorithm recommended by Greenwald et al. (2003), with positive D scores indicating implicit evaluations in line with the attitude induction. Using the split-half correlation method recommended by Kurdi et al. (2018), an estimate of $R = .77$ was obtained for the internal consistency of the IAT, which is comparable to but somewhat lower than the estimate reported by Bar-Anan and Nosek (2014).

Explicit attitudes. Participants responded to two feeling thermometer items, one of them asking how warmly they felt toward square-faced individuals and the other one asking how warmly they felt toward long-faced individuals. Responses were provided on 10-point scales, anchored by *extremely warmly* and *extremely coldly*. Given the focus of the current work on implicit attitudes, explicit attitudes are not discussed any further.[8]

Manipulation check. Given that contingency awareness is a major moderator of attitude change via repeated evaluative pairings (Hofmann et al., 2010), participants were asked to complete a manipulation check item, which instructed them to recall what they had learned at the beginning of the study. Response options included "square faces are good and long faces are bad" (accurate response), "long faces are good and square faces are bad" (inaccurate response), and "nothing that would have indicated whether the groups are good and bad" (intermediate response).[9]

Results[10]

Number of stimulus pairings. The focal question of Study 1 was whether the number of stimulus pairings presented to participants would modulate the strength of the implicit attitudes created as a result of a learning intervention involving repeated evaluative pairings of category members with valenced images. This question was probed by fitting a linear regression to the data, with IAT D score as the response variable and number of stimulus pairings presented as the sole predictor (see Figure 1). We obtained a significant intercept, $b_0 = .31$ [.25, .37], $t(1122) = 10.56$, $p < .001$, indicating considerable deviation from neutrality among participants who had been exposed to as few as four CS–US pairings. Crucially, evidence for linear change in attitude strength as a result of increasing the number of stimulus pairings would be provided by a significant slope parameter, which we failed to obtain, $b_1 = .00$ [−.00, .01], $t(1122) = 1.63$, $p = .102$. Visual inspection revealed no obvious nonlinearity in the data that could have accounted for this result.

Summary of supplementary analyses. In Supplement 1, we report additional analyses of the same data. To summarize, we demonstrate that (a) the evaluative learning effects reported here are significantly modulated by participants' conscious recollection of the learning manipulation (see Hofmann et al., 2010); (b) the focal result reported above, suggesting no effect of number of

[5] The verbatim text of the instructions is available for download from OSF (https://osf.io/serq4/).

[6] For exploratory purposes, the opposite pairings were also included in the study but are not discussed here. Interested readers may conduct their own analyses using the raw data made available on OSF (https://osf.io/serq4/).

[7] Positive words used as stimuli on the IAT included *love*, *peace*, *joy*, *happy*, *sweet*, *glory*, and *success*, and negative words included *hate*, *war*, *devil*, *bomb*, *bitter*, *agony*, and *failure*.

[8] Interested readers may conduct their own analyses using the data made available on OSF (https://osf.io/serq4/).

[9] This manipulation check item is not a pure measure of contingency awareness but rather a combined measure of contingency awareness and participants' ability and willingness to make inferences about the valence of the target categories from the visual (or, in later studies, verbal) information presented to them. As such, an intermediate response need not indicate lack of contingency awareness; it may reflect participants' inability or unwillingness to conclude that group A is good and group B is bad.

[10] Statistical analyses for this and subsequent studies were performed using the R statistical computing environment. The R code enabling interested researchers to reproduce all analyses reported in the article is available from OSF (https://osf.io/serq4/).

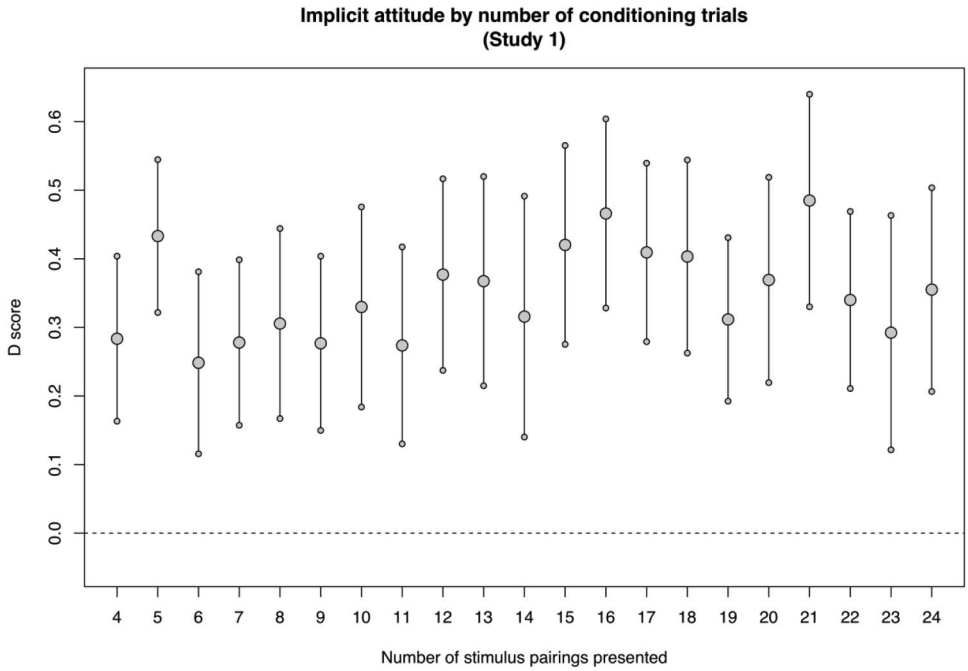

Figure 1. Implicit attitude strength (measured using IAT D scores) as a function of the number of conditioning trials presented (Study 1). Positive scores imply implicit attitudes in line with the learning manipulation. Error bars show 95% confidence intervals comparing each effect size to zero. The dashed line shows neutrality.

stimulus pairings, emerges in a Bayesian linear regression using uninformative priors; (c) in a model including all participants without exclusions based on contingency awareness, only contingency awareness but not number of stimulus pairings emerges as a significant predictor of implicit attitude strength; (d) a significant difference in implicit attitudes emerges between the incongruent learning condition reported here (square faces paired with positive and long faces paired with negative stimuli) and the congruent learning condition (long faces paired with positive and square faces paired with negative stimuli) and this difference is consistent across number of stimulus pairings; and (e) the effect of number of stimulus exposures is not significant in the congruent learning condition, thus mirroring the results reported here. Overall, these supplementary analyses demonstrate the robustness of the focal result of no effect of number of stimulus pairings to (a) participant exclusions, (b) stimulus effects, and (c) analytic frameworks.

Discussion

The present study probed the effects of the number of stimulus pairings on the strength of implicit attitudes created toward novel targets. Experiencing four stimulus pairings was sufficient to induce novel implicit attitudes, with no difference observed between implicit evaluations that emerged as a result of 24 compared with four CS–US pairings (for a similar result with two between-participants conditions, see Hu et al., 2017). Based on a Bayesian linear regression reported in more detail in Supplement 1, the most likely posterior value of the slope parameter is zero, thus providing positive evidence in favor of attitude stability across number of stimulus pairings, rather than incremental updating. Unlike the number of stimulus pairings, participants' conscious recollection of the content of the learning manipulation was strongly predictive of the strength of implicit attitudes (see also Gast & De Houwer, 2012; Van Dessel, De Houwer, & Gast, 2015; Van Dessel, De Houwer, Roets, & Gast, 2016).

Overall, these results seem easier to reconcile with an approach to human learning that allows for quick inferences from novel data than with an approach that characterizes learning as a process of incremental updating. At the same time, it should be noted that even though the associative approach to human learning generally relies on the idea of incremental updating, the same approach can also allow for high learning rates and, as such, accommodate episodes of single-shot associative learning, for example, fear conditioning from a single CS–US pairing (Drew, Denny, & Hen, 2010). However, the implementations of this approach in the context of implicit evaluation (e.g., McConnell & Rydell, 2014; Rydell & McConnell, 2006; E. R. Smith & DeCoster, 2000; Strack & Deutsch, 2004) all explicitly posit a slow-learning process, which should not be able to produce change of the magnitude observed here within as few as four learning trials, corresponding to only 10 seconds of learning.

Study 2

Following a test of the unfolding of implicit attitude acquisition from an increasing number of repeated evaluative pairings, Study 2 probed another key moderator of this learning effect. Specifically, we investigated whether implicit evaluations are updated in an informationally encapsulated manner, arising preferentially from stimulus pairings experienced in the environment, or in an informationally promiscuous manner, combining all relevant sources of evidence, including pairings and verbal information.

Existing evidence on this is issue is contradictory, which provides further impetus for another test: Some studies have found that verbal descriptions of the nature of stimulus pairings significantly modulate evaluative learning from REP (Moran et al., 2016; Moran, Bar-Anan, & Nosek, 2015; Peters & Gawronski, 2011; Zanon et al., 2014), whereas in others, similar manipulations produced no effect (Hu et al., 2017; Moran & Bar-Anan, 2013).

Recent work has demonstrated that *diagnosticity*, that is, whether a piece of information is perceived as being reflective of an attitude object's true underlying character, can have sizable effects on implicit evaluations acquired on the basis of purely verbal manipulations (Cone & Ferguson, 2015). Inspired by this work, Study 2 investigated whether manipulations of diagnosticity can influence the acquisition of implicit attitudes from repeated evaluative pairings: In one condition, participants were informed that stimulus pairings were highly diagnostic (i.e., revealing the true nature of the target categories), whereas in another condition they were told that stimulus pairings were randomly generated and thus completely nondiagnostic. Given that the time at which verbal information about stimulus pairings is provided seems to be a crucial moderator of its effectiveness in influencing implicit attitudes (Hu et al., 2017; Peters & Gawronski, 2011; Zanon et al., 2014), we also varied whether participants received diagnosticity information before or after exposure to REP.

Method

Participants and design. Volunteers (N = 2,097) were recruited from the Project Implicit educational website (http://implicit.harvard.edu) to participate in the study.[11] Consistent with standard scoring procedures (Greenwald et al., 2003), participants who did not complete the IAT and participants with a response latency of 300 ms or less on at least 10% of trials (N = 246) were excluded from all further analyses. Moreover, in line with standard exclusion rules used in the field in general (Van Dessel, De Houwer, Gast, & Smith, 2015; Van Dessel, De Houwer, Gast, et al., 2016; Van Dessel, Gawronski, et al., 2017) and in Study 1 in particular, participants who provided an inaccurate response on a manipulation check probing explicit recollection of the learning phase (N = 483) were also eliminated from consideration. This resulted in a final sample size of N = 1,365 (N = 896 female and N = 428 male participants, mean age = 38.32, SD = 15.39). This sample size provides .80 power to detect a small effect of r = .08.

All participants underwent an attitude induction procedure using repeated evaluative pairings (REP) of category members with positive and negative images. In a 2 × 2 between-subjects design, each participant was randomly assigned to a *diagnosticity condition* (stimulus pairings described as diagnostic vs. nondiagnostic) as well as to a *timing condition* (diagnosticity information received before vs. after exposure to stimulus pairings). Similar to Study 1, we failed to obtain evidence for differential attrition across conditions, $\chi^2(3) = 7.27$, $p = .064$.

Materials. Names from two fictitious social groups served as conditioned stimuli (CS). Names were pronounceable nonsense words constructed two conform to one of two phonological patterns: Laapian names ended on the syllable *lap* (e.g., Caalap) and Niffian names ended on the syllable *nif* (Gregg et al., 2006). The unconditioned stimuli (US) were the same as in Study 1. Both the CS and US have been used successfully for attitude induction in previous work (Gregg et al., 2006; Kurdi & Banaji, 2017).[12]

Procedure and measures. The study was administered entirely online and consisted of (a) a *learning phase* in which implicit attitudes were induced toward the target categories via exposure to REP, with a diagnosticity manipulation implemented either before or after the presentation of stimulus pairings, and (b) a *test phase* in which implicit attitudes and explicit attitudes toward the targets as well as memory for stimulus pairings were assessed.

Learning phase. The learning phase was identical to the learning phase of Study 1, with three exceptions. First, as mentioned above, Laapian and Niffian names, rather than individuals with long and square faces, served as CS. Because of a slight baseline preference in favor of Laapians (Kurdi & Banaji, 2017), Niffians were always paired with positive stimuli and Laapians were always paired with negative stimuli to allow for learning to unfold without ceiling effects. Second, the number of stimulus pairings presented to participants was not varied but was rather held constant at 20. Finally, and crucially, the diagnosticity of stimulus pairings was manipulated using a simple verbal instruction.[13] Participants in the diagnostic condition were informed that the pairings had been created to teach them something fundamental about the nature of the groups, whereas participants in the nondiagnostic condition were informed that the pairings had been randomly generated by the computer. Depending on their assignment to timing condition, this information was communicated to participants either as part of the initial instructions and before the repeated evaluative pairings were displayed (*before condition*) or at the end of the learning phase, after the repeated evaluative pairings had already been presented (*after condition*).

Test phase. The test phase was identical to the test phase of Study 1, with the exception that the target categories were Laapians and Niffians rather than individuals with long faces and square faces. Participants completed (a) an IAT (Greenwald et al., 1998) measuring their implicit attitudes toward the targets (internal consistency R = .77), followed by (b) two feeling thermometer measures used to construct a relative measure of explicit preferences, and finally (c) a manipulation check probing participants' recollection of the stimulus pairings presented in the learning phase.

Results

Diagnosticity and timing. The focal question of Study 2 was whether the diagnosticity of stimulus pairings, manipulated via minimal verbal information presented either before or after exposure to stimulus pairings, would influence the magnitude of implicit attitudes. The effects of the diagnosticity and timing conditions were probed by fitting a linear regression to the data, with IAT D score as the response variable, and timing condition (0 = after and

[11] The initial data collection included 833 participants. In response to a reviewer's concern about the strength of the effect given the sample size, an additional round of data collection, involving 1,264 new participants, was conducted. Given that the new data showed a similar pattern for the crucial interaction, $\chi^2(9) = 4.25$, $p = .894$, the analyses reported in the paper are based on the data from the full sample.

[12] All stimuli are available for download from OSF (https://osf.io/serq4/).

[13] The verbatim text of the instructions is available for download from OSF (https://osf.io/serq4/).

1 = before), diagnosticity condition (0 = nondiagnostic and 1 = diagnostic), as well as their interaction, entered as predictors (for a visual display of the results, see Figure 2).

We obtained a significant intercept, $b_0 = .26$ [.21, .32], $t(1326) = 10.11$, $p < .001$, suggesting considerable deviation from neutrality in the reference category, that is, among the group of participants who received information about the stimulus pairings being nondiagnostic after being exposed to them. However, the main effects of timing condition, $b_1 = .04$ [−.04, .11], $t(1326) = 0.93$, $p = .350$, and of diagnosticity condition, $b_2 = .02$ [−.06, .09], $t(1326) = 0.41$, $p = .682$, were both nonsignificant. Thus, the present data provide no evidence for an effect of (a) timing (before vs. after stimulus presentation) among participants in the diagnostic information condition or (b) diagnosticity information (stimulus pairings diagnostic vs. nondiagnostic) when it was communicated to participants *after* the stimulus pairings have already been presented.

Crucially, the Timing × Diagnosticity interaction emerged as significant, $b_3 = -.12$ [−.23, −.01], $t(1326) = -2.19$, $p = .029$, indicating a larger effect of diagnosticity in the information before condition than in the information after condition. Taken together, these results suggest that whereas the diagnosticity manipulation had considerable influence on the extent of evaluative learning when it was implemented *before* exposure to stimulus pairings, it did not produce any reliable effect *after* stimulus pairings had already been presented. The same impression can also be confirmed by fitting separate linear regressions to the data in the before versus after conditions, with diagnosticity as the predictor. In the *before* condition, a significant effect of diagnosticity emerged, $t(669) = 2.70$, $p = .007$; however, in the *after* condition, no evidence of the same effect could be obtained, $t(657) = 0.41$, $p = .682$.

Summary of supplementary analyses. In Supplement 1, we report additional analyses of the same data. To summarize, we demonstrate that (a) the evaluative learning effects reported here are significantly modulated by participants' conscious recollection of the learning manipulation (see Hofmann et al., 2010 and current Study 1); (b) the focal result reported above, yielding a significant Timing × Diagnosticity interaction in predicting implicit attitudes, also emerges in a Bayesian linear regression using uninformative priors; and (c) in a model including all participants without exclusions based on contingency awareness, only contingency awareness emerges as significant predictors of the strength of implicit attitudes, whereas the Timing × Diagnosticity interaction is reduced to nonsignificance. Overall, these supplementary analyses (a) demonstrate the robustness of the focal Timing × Diagnosticity interaction analytic frameworks and (b) suggest that the effects of the experimental manipulation on implicit evaluations may not be independent of the effects of contingency awareness.

Discussion

Study 2 probed whether (a) implicit attitude acquisition preferentially relies on stimulus pairings encountered in the environment, without taking into account verbal information accompanying such stimulus pairings, or (b) verbal information about the diagnosticity of such stimulus pairings influences the inferences made by participants and thus the strength of implicit attitudes toward the categories. We found that implicit attitudes were stronger in the condition in which stimulus pairings were described as diagnostic than in the condition in which they were described as randomly generated. In combination with previous work (Moran et al., 2015, 2016; Peters & Gawronski, 2011; Zanon et al., 2014), this finding provides evidence that implicit attitude acquisition from REP is not impervious to the influence of verbal instructions. Moreover, it demonstrates that in addition to validity (Moran et al., 2015; Peters & Gawronski, 2011) and type of relationship such as opposition (Zanon et al., 2014) or causation (Moran et al., 2015, 2016), information about diagnosticity can also modulate learning from valenced stimulus pairings.

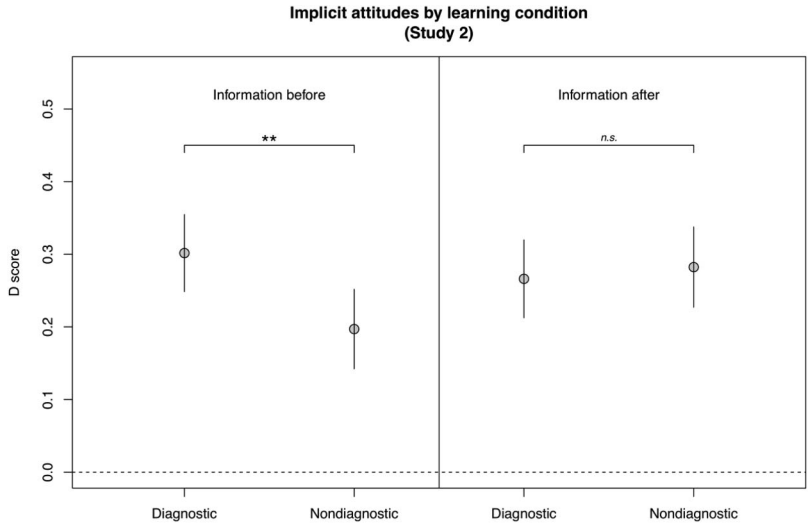

Figure 2. Implicit attitude strength (measured using IAT D scores) as a function of learning conditions (Study 2). The left pane shows the information before condition, whereas the right pane shows the information after condition. Positive scores imply implicit attitudes in line with the learning manipulation. Error bars show 95% confidence intervals comparing each effect size to zero and differences significant at $p < .005$ are marked **.

However, the effects of the diagnosticity manipulation are subject to at least two qualifications. First, implicit attitudes significantly differed from neutrality even in the condition in which stimulus pairings were described as nondiagnostic before actual exposure. This result is in line with previous findings demonstrating that even purely verbal information presented to participants as fully nondiagnostic can influence implicit evaluations (Van Dessel, De Houwer, Gast, et al., 2016). Such findings may be construed as evidence that REP-based interventions shift implicit attitudes via a mix of associative and propositional processes (e.g., Gawronski & Bodenhausen, 2018; see General Discussion). Alternatively, they may be parsimoniously accounted for within a propositional framework. Specifically, human communication ubiquitously relies on an assumption of relevance (Grice, 1975) and a single manipulation of diagnosticity may not be sufficient to fully override this assumption. Moreover, as pointed out by De Houwer (2018), the time constraints imposed by implicit measures may result in "quick and dirty" propositional reasoning in the course of which not all relevant pieces of information may be fully integrated with each other. Indeed, in a recent study using a process dissociation framework, Hütter and De Houwer (2017) provided evidence for automatic (memory-independent) effects of purely verbal manipulations on evaluative learning, suggesting that associative processes need not underlie the influence of nondiagnostic information on implicit evaluations.

Second, the manipulation of diagnosticity was ineffectual in the condition in which it was implemented after exposure to stimulus pairings (see also Hu et al., 2017; Peters & Gawronski, 2011; Zanon et al., 2014). This finding was not, a priori, predicted by propositional theories, although it is compatible with a Bayesian model of evaluative conditioning assuming a relatively strong prior expectation of stimulus pairings being diagnostic. Further implications of the latter qualification on the effect of the diagnosticity manipulation, including potential computational models that may be able to accommodate them, are addressed in the General Discussion.

Study 3

The effect of the passage of time on the retention of information is among the earliest and most widely studied phenomena in the psychology of learning and memory (e.g., Bouton, 1993; Ebbinghaus, 1913/2011; Rubin & Wenzel, 1996). In the initial investigations probing the relative effectiveness of repeated evaluative pairings versus evaluative statements in shifting implicit evaluations, attitude change was assessed immediately after the learning phase (Kurdi & Banaji, 2017). However, in the context of implicit attitude acquisition, the rate of decay in learning effects as a result of different interventions has far-reaching practical and theoretical implications.

Attempts at creating long-term change in implicit evaluations are rare and usually unsuccessful (for some notable exceptions see Mann & Ferguson, 2015, 2017). Failures to achieve durable change may, at least in part, be attributable to the fact that most studies to date have used long delays of days or at least hours between learning and test (Devine et al., 2012; Lai et al., 2016; Mann & Ferguson, 2017) during which (a) forgetting may have occurred and (b) participants may have spontaneously encountered sources of interference. To offer some insight into the decay of learning effects before the return of implicit attitudes to their to baseline levels, the present studies used shorter delays than are customary in this literature, probing the effects of time within a single IAT session (Study 3) and imposing a delay of 15 min, rather than hours or days, between the learning and test phases of the experiment (Study 4).

Crucially, rates of decay in implicit attitudes induced via repeated evaluative pairings (REP) versus evaluative statements (ES) also have the potential to illuminate the processes by which each intervention produces evaluative learning. As discussed above, associative theories predict that REP should create more enduring change in implicit evaluations than ES because of a proposed fit between stimulus associations experienced in the environment and conceptual associations stored in long-term memory.

Conversely, existing propositional accounts of implicit evaluation do not make a strong prediction regarding the memory effects that should emerge. However, to the extent that they posit explicit and implicit evaluations to be subserved by the same memory systems, propositional accounts seem to be compatible with the idea that findings regarding episodic memory should be applicable to the present study. Specifically, episodic memory for verbally described events has been demonstrated to be initially superior to episodic memory for actually experienced events (Baggett, 1979; Baggett & Ehrenfeucht, 1983; Beentjes & van der Voort, 1991; Larsen & Plunkett, 1987; Toglia et al., 1992); however, this pattern is reversed if a delay is imposed between the event and the test of memory. If the same finding extends to the effects of evaluative learning, as reflected by implicit attitude measures, then the effects of ES should dissipate quickly, whereas learning in the REP condition should remain relatively stable.

As such, more durable learning effects of REP compared with ES are generally compatible with both associative and propositional theories. However, the predictions of both perspectives diverge when it comes to a combined ES + REP intervention. Specifically, if the superiority of REP over ES is attributable to the presence of stimulus pairings, as argued by associative theories, the ES + REP combined condition should be indistinguishable from the REP condition given that both involve exposure to stimulus pairings. By contrast, experienced events, such as the REP condition of the present studies, may result in superior long-term retention because they require participants to generate inferences about the event rather than merely encoding information already provided to them in propositional form (Baggett, 1979). Should this be the case, then the combined ES + REP condition should exhibit the same patterns of decay as the ES condition given that participants in the combined condition, like in the ES condition, need not generate any inferences about the attitude objects. Study 3 provides an initial test of these predictions by conducting a reanalysis of the data reported by Kurdi and Banaji (2017), focusing on the effects of time on implicit attitude strength as a function of different learning interventions (REP, ES, and ES + REP combined).

Method

Participants and design. A final sample of 2,201 volunteers from the Project Implicit educational website (http://implicit.harvard

.edu) participated in the study. This sample size provides .80 power to detect a small effect of $r = .06$. Exclusion criteria and demographic characteristics are described in Kurdi and Banaji (2017). In a between-subjects design, participants were randomly assigned to one of four learning conditions: *control* (involving exposure to only US–US pairings), *repeated evaluative pairings* (REP; exposure to pairings of category members with positive and negative images), *evaluative statements* (ES; exposure to verbal information about upcoming stimulus pairings without actual stimulus presentations), and *combined* (ES + REP; verbal information followed by exposure to stimulus pairings). Unlike in Studies 1 and 2, we observed differential attrition across learning conditions, $\chi^2(3) = 16.68$, $p = .001$. Specifically, differential attrition affected primarily the ES condition, which included $N = 498$ participants (as opposed to the expected N of 550 under H_0).

Materials. Conditioned stimuli included Laapian and Niffian names (Study 3A; $N = 329$), individuals with long faces and square faces (Study 3B; $N = 528$), squares and rectangles (Study 3C; $N = 435$), faces of young and elderly individuals (Study 3D; $N = 429$), and photographs of American and foreign symbols (Study 3E; $N = 480$). The line drawings used in Studies 1 and 2 served as unconditioned stimuli, with the exception of Study 3D in which valenced photographs drawn from the Open Affective Standardized Image Set (OASIS; Kurdi, Lozano, & Banaji, 2017) were used.[14]

Procedure and measures. The study was administered entirely online and consisted of (a) a *learning phase* in which implicit attitudes toward the target categories were induced using one of three learning manipulations, described in detail below, and (b) a *test phase* in which implicit and explicit attitudes toward the targets were assessed.[15]

Learning phase. Participants were randomly assigned to one of four conditions: a control condition and three learning conditions (REP, ES, ES + REP combined), with time on task kept constant across all four conditions. In the *control condition*, participants were exposed to US–US parings to control for stimulus exposure. In the *REP condition*, attitudes were induced via exposure to pairings of members of one target category with positive images and members of the other target category with negative images. Crucially, in this condition, no verbal labels were used for the groups and no mention was made of stimulus valence. In the *ES condition*, participants were informed that they would see pairings of members of one category with positive images and members of the other category with negative images (De Houwer, 2006). In fact, no stimulus pairings were presented. Finally, in the *combined ES + REP condition*, verbal descriptions of stimulus pairings were followed by exposure to actual stimulus pairings.[16] To avoid ceiling effects, the attitude induction procedure was designed to move participants away from the prevalent baseline. That is, participants were taught that (a) Niffians are good and Laapians are bad (Study 3A), (b) square faces are good and long faces are bad (Study 3B), (c) rectangles are good and squares are bad (Study 3C), (d) elderly are good and young are bad (Study 3D), and (e) foreign is good and American is bad (Study 3E).[17] Further procedural details are described in Kurdi and Banaji (2017).

Test phase. The test phase was identical to the test phases of Studies 1 and 2, with the exception that no manipulation check probing for contingency awareness was administered to participants. Participants completed (a) an IAT (Greenwald et al., 1998) measuring their implicit attitudes toward the target categories, followed by (b) two feeling thermometer measures used to construct a relative measure of explicit preferences.

Analytic strategy. Traditional analyses of these data relying on D scores (Greenwald et al., 2003) to measure IAT performance were reported by Kurdi and Banaji (2017). Both the original scoring algorithm producing a simple mean difference (Greenwald et al., 1998) and the improved scoring algorithm producing a standardized mean difference across critical blocks (Greenwald et al., 2003) treat each IAT trial as equivalent, without taking into account the temporal aspect of IAT data. The focal question of the present study was whether IAT D scores unfolded differently over time as a result of the different learning manipulations administered to participants. We investigated this question using a moving time window method, that is, we recalculated the IAT D score for overlapping subsets of 10 trials, resulting in mini D scores. The first mini D score was calculated from Trials 1 through 10 of both the consistent and inconsistent IAT blocks, the second mini D score was calculated from Trials 2 through 11, and so on. Because both critical blocks consisted of 40 trials, 31 mini D scores were calculated for each participant and then submitted to statistical analyses.[18]

Results

Descriptive statistics. The temporal unfolding of mini D scores by learning condition is visually displayed in Figure 3. Adjusting for mean differences across studies, D scores in the control condition approached neutrality as the IAT session progressed, decreasing in absolute value from −0.37 (first mini D score involving Trials 1 through 10) to −0.22 (last mini D score involving Trials 31 through 40). D scores in the REP condition remained relatively stable, with an initial mini D score of −0.05 and a last mini D score of −0.01. The decline of mini D scores in the ES and combined conditions was considerably steeper, shifting from .20 to .09 in ES and from .23 to .12 in combined.

Model fitting. To formally test whether the time course of the decay in evaluative learning effects significantly differed across learning conditions, we fit a linear mixed-effects model (Baayen, Davidson, & Bates, 2008; Judd, Westfall, & Kenny, 2012) to the data using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in the R statistical computing environment, with mini D

[14] All stimuli are available for download from OSF (https://osf.io/serq4/).

[15] Study 3 was conducted before Studies 1, 2, and 4 and thus did not contain a manipulation check item.

[16] Because time on task was kept constant across conditions, participants in the control and combined conditions were exposed to 20 stimulus pairings, whereas participants in the REP condition were exposed to 37 stimulus pairings. However, in line with the results of Study 1, the pattern of results did not change when the combined condition involved the same number of stimulus pairings as the REP condition (Kurdi & Banaji, 2017).

[17] For exploratory purposes, the opposite associations were also reinforced for some participants. Data from those participants are not discussed here but are available for download from OSF (https://osf.io/serq4/).

[18] Because this method is new, we sought to ascertain its robustness to the size of the time window chosen. Therefore, we recalculated mini D scores based on five, rather than 10, trials each and repeated all statistical analyses reported below. We obtained inferentially equivalent results, suggesting that the method is relatively robust to the number of trials chosen.

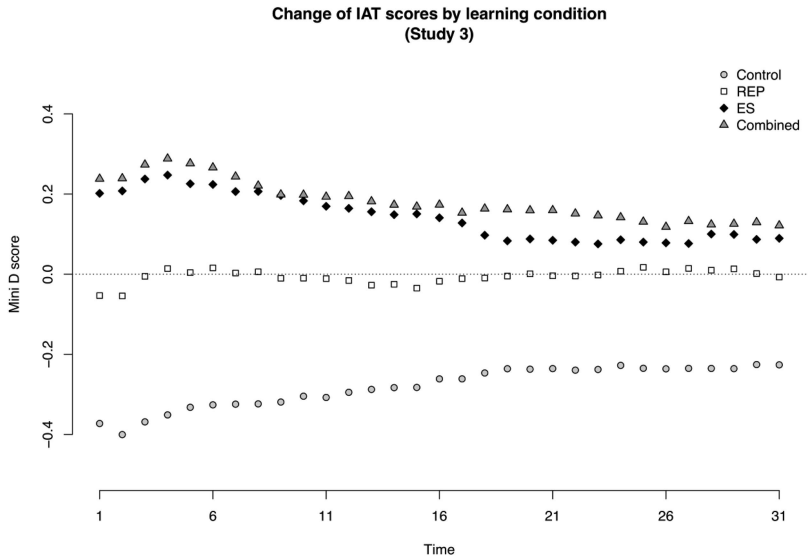

Figure 3. Implicit attitude strength (measured using IAT mini D scores) as a function of learning conditions and the passage of time within IAT sessions (Study 3). Positive scores imply implicit attitudes in line with the learning manipulation. The *x* axis marks the passage of time within the IAT session, with each number showing the first trial taken into account for calculating each 10-trial mini D score. Light gray dots show the control condition, white squares show the repeated evaluative pairings (REP) condition, black diamonds show the evaluative statements (ES) condition, and dark gray triangles show the ES + REP combined condition.

scores as the response variable. Model fitting proceeded in a stepwise fashion, with a predictor added in each step and a likelihood ratio test conducted to probe whether the new predictor significantly improved model fit. After starting with a null model that contained only random intercepts for participants to account for the fact that mini D scores were nested within participants, random intercepts for target categories were added to adjust for mean differences in implicit attitudes, $\chi^2(1) = 817.62$, $p < .001$. Subsequently, we entered a fixed effect for time as a continuous variable coded as the first trial included in the given mini D score, $\chi^2(1) = 36.46$, $p < .001$. In the next step, a fixed effect for learning condition (control vs. REP vs. ES vs. combined) was added, $\chi^2(3) = 253.75$, $p < .001$. Finally, as a test the focal question of the present study, we entered a Time × Learning Condition interaction, $\chi^2(3) = 759.58$, $p < .001$. Taken together, these model fitting steps suggest that, as expected, evaluative learning decays differentially depending on learning conditions even within a single IAT session. The nature of this interaction is discussed under Model interpretation below.

Model interpretation. The regression coefficients from the best-fitting mixed-effects model, including random intercepts for participants and targets, as well as a Time × Learning Condition interaction, are presented in Table 1. Dummy coding with the control condition as the reference category was used to represent the learning condition variable and time was entered as a continuous variable scaled such that 0 represents the first mini D score (and, correspondingly, 30 represents the last mini D score).

We observed a significant and positive main effect of time, suggesting that in the control condition, mini D scores approached neutrality as time progressed. In addition, we observed main effects for each learning condition compared with control, which is in line with previous work but irrelevant from the perspective of the present study. Crucially, we obtained significant Time × Learning Condition interactions, suggesting that the effects of time within the IAT session changed depending on the manner in which implicit attitudes had been acquired. Whereas in the REP condition implicit attitudes were predicted to stay quite stable throughout the IAT session (predicted change per trial: $d_{\text{pred}} = 0.001$), in the ES ($d_{\text{pred}} = -0.006$) and combined ($d_{\text{pred}} = -0.006$) conditions, implicit attitudes were predicted to approach neutrality at a considerably higher pace. According to the regression model, D scores in the REP condition should change (away from neutrality) by 0.026 D score units throughout the entire IAT session. By contrast, the overall predicted decrease toward neutrality was 0.178 units in the ES condition and 0.157 units in the combined condition.

Summary of supplementary analyses. A Bayesian reanalysis of the same data, reported in Supplement 1, confirms the main inference reported above: Implicit attitudes measured in the in the ES and combined conditions approached neutrality at a considerably faster rate than implicit attitudes measured in the REP condition.

Discussion

Study 3 provides initial evidence that, in line with predictions derived from associative theories of implicit evaluation and from studies on episodic memory for personally experienced versus verbally described events, the passage of time differentially affects implicit attitudes acquired from REP versus ES. Prior work has robustly demonstrated that when implicit evaluations are assessed immediately following the learning intervention, the learning effects produced by ES are at least on par with, and often superior to, the learning effects produced by REP (Kurdi & Banaji, 2017). However, when the passage of time within the IAT session is taken

Table 1
Regression Table Showing the Best-Fitting Mixed-Effects Model from Study 3

Groups	Random effects				
	Variance				*SD*
Participants (intercept)	.284				.533
Target (intercept)	.186				.431
Residual	.157				.396
	Fixed effects				
	Estimate	*SE*	*df*	*t*	*p*
Intercept	−.335	.194	5	−1.73	.15
Time	.005	.000	66027	16.47	<.001
Learning condition (REP)	.341	.033	2414	10.32	<.001
Learning condition (ES)	.591	.033	2414	17.78	<.001
Learning condition (Combined)	.621	.032	2414	19.43	<.001
Time × Learning condition (REP)	−.004	.000	66027	−9.28	<.001
Time × Learning condition (ES)	−.011	.000	66027	−23.35	<.001
Time × Learning condition (Combined)	−.011	.000	66027	−22.77	<.001

Note. Random effects include random intercepts for participants and targets. Fixed effects include time (with the mini *D* score calculated from Trials 1–10 coded as 0), learning condition (control, REP, ES, and ES + REP combined, with the control condition serving as the reference category), and a Time × Learning condition interaction. *SD* = standard deviation, *SE* = standard error, *df* = degrees of freedom.

into account, the interpretation of the same data changes substantially: Whereas attitudes created via ES were found to decay considerably even within a single IAT session, attitudes created via REP remained stable, suggesting that any differences in favor of the ES condition observed in previous work were driven by early trials of the IAT and dissipated quickly over time.

Furthermore, the combined condition exhibited the same rapid decay as the ES condition, indicating that exposure to stimulus pairings (as opposed to purely verbal information) may, in and of itself, not produce stronger evaluative learning unless participants are required to make inferences about those stimulus pairings (Toglia et al., 1992). This result, unlike the superior long-term effects of REP compared with ES, is difficult to reconcile with associative theories of implicit evaluation. Specifically, differences in the behavior of the ES + REP versus REP condition are not easily explained by an account suggesting that the primary determinant of the durability of evaluative learning is whether the relevant manipulation involves exposure to actual stimulus pairings. Rather, the REP condition may create more durable representations due to participant-generated propositions being more memorable than experimenter-provided ones, possibly because the additional cognitive effort that they require may result in deeper processing of evaluative information (Craik & Lockhart, 1972).

Study 4

Study 3 provided initial evidence for the differential decay of evaluative learning effects as a function of the learning manipulation used: When implicit attitude strength was measured dynamically within a single IAT session, learning from stimulus pairings remained stable, whereas learning produced as a result of a purely verbal manipulation or a combination of verbal instructions and stimulus pairings exhibited precipitous decay. Taken together, these results suggest that the stability in implicit attitudes in the REP condition is most likely due to the memory benefit created by participant-generated inferences (Baggett, 1979; Craik & Lockhart, 1972; Toglia et al., 1992). In Study 4, we created another test of the same idea by imposing a 15-min delay between the learning and test phases of the experiment. If the findings obtained in Study 3 extend to this relatively longer time scale, the advantage of the ES and ES + REP conditions over the REP condition, as observed when implicit attitudes are measured immediately after learning (Kurdi & Banaji, 2017), should be reduced or possibly even eliminated.

Method

Participants and design. A sample of 506 undergraduates from the study pool of a private university in the Northeastern United States and a sample of 189 participants from the Digital Lab for the Social Sciences (DLABSS; http://dlabss.harvard.edu/) were recruited for the study, the former in exchange for partial course credit and the latter as volunteers. In line with standard scoring procedures (Greenwald et al., 2003), participants who did not complete the IAT ($N = 22$) as well as participants with a response latency of 300 ms or less on at least 10% of trials ($N = 9$) were excluded from all further analyses. Moreover, consistent with Studies 1 and 2 and standard exclusion rules used in the field in general (e.g., Van Dessel, De Houwer, Gast, & Smith, 2015; Van Dessel, De Houwer, Gast, et al., 2016; Van Dessel, Gawronski, et al., 2017), participants who provided an inaccurate response on a manipulation check probing explicit recollection of the learning phase ($N = 56$) were also eliminated from consideration. This resulted in a combined final sample size of $N = 606$. Because the effects of learning manipulations did not differ across the undergraduate and DLABSS participants (see below), analyses collapse across the two samples. The combined sample size provides .80 power to detect a small effect of $r = .11$.

Like in Study 3, participants were randomly assigned to one of four learning conditions in a between-subjects design: *control*

(involving exposure to only US–US pairings), *repeated evaluative pairings* (REP; exposure to pairings of category members with positive and negative images), *evaluative statements* (ES; exposure to verbal information about upcoming stimulus pairings without actual stimulus presentations), and *combined* (ES + REP; verbal information followed by exposure to stimulus pairings). However, unlike in Study 3, a 15-min delay was imposed between the learning phase and the test phase of the experiment. Unlike in Studies 1 and 2, but similar to Study 3, we observed differential attrition across learning conditions, $\chi^2(3) = 45.34$, $p < .001$. However, unlike in Study 3, differential attrition affected primarily the control condition, which included $N = 87$ participants (as opposed to the expected N of 152 under H_0).

Materials. Conditioned stimuli included Laapian and Niffian names (undergraduate participants) and individuals with long faces and square faces (DLABSS participants). Because of different baseline attitudes across the two stimulus sets and samples, mixed-effects models (see below) included random intercepts for participant pools. The line drawings used in Studies 1–3 and in Kurdi and Banaji (2017) served as unconditioned stimuli for all participants.[19]

Procedure and measures. The study was administered entirely online and consisted of (a) a *learning phase* in which implicit attitudes toward the target categories were induced using one of three learning manipulations mentioned above (REP, ES, or ES + REP combined), with some additional participants measured at baseline in the control condition, and (b) a *test phase* in which implicit and explicit attitudes toward the targets as well as memory for stimulus pairings were assessed. Similar to previous studies, internal consistency of the IAT was acceptable, $R = .76$. Both the learning phase and the test phase were identical to Study 3, with two exceptions: (a) following the explicit attitude measures, the test phase included a measure of memory for stimulus pairings and, crucially, (b) a delay of 15 min was imposed between the learning and test phases of the experiment. During this delay, participants were allowed to spend their time in any way they wished. Once the delay was over, participants received an e-mail with a link to the second part of the study.

Results

Learning conditions. Study 4 examined the relative effectiveness of different learning manipulations in shifting implicit attitudes, as assessed following a 15-min delay between the learning and test phases of the experiment. Mean implicit attitudes by learning condition (control, REP, ES, and combined) are displayed in Figure 4. Descriptively, Figure 4 suggests that implicit attitudes in all three learning conditions remained different from control in spite of the delay. Moreover, unlike in previous work involving immediate testing of the effects of evaluative learning, implicit attitudes in the REP and combined conditions seem descriptively higher than in the ES condition; however, this impression may not be confirmed using inferential testing.

To formally test whether implicit attitudes significantly differed across learning conditions, we fit a linear mixed-effects model (Baayen et al., 2008; Judd et al., 2012) to the data using the lme4 package (Bates et al., 2015) in the R statistical computing environment, with IAT D scores as the response variable. As in Study 3, model fitting proceeded in a stepwise fashion. After starting with a null model that contained only random intercepts for participant pools to account for different baseline attitudes, a fixed effect for learning conditions was entered and found to produce significant improvement in model fit, $\chi^2(3) = 54.54$, $p < .001$. Finally, we added random slopes for learning conditions across participant pools, accounting for the possibility that the effect of learning manipulations may have differed across undergraduate and DLABSS participants; however, we obtained no evidence for such differences, $\chi^2(9) = 3.42$, $p = .945$. Therefore, below we interpret the best-fitting model, which included random intercepts for participant pools and a fixed effect for learning conditions.

As in Study 3, the control condition served as the reference category. Therefore, the intercept represents mean implicit attitude strength in the control condition and slope parameters represent differences between the control condition and each learning condition. Participants in the control condition exhibited a mean implicit attitude level of $b_0 = .07$ [−.18, .32], $t(2.90) = 0.60$, $p = .589$, suggesting a neutral baseline. Compared with the control condition, each learning condition produced significant learning, $b_1 = .40$ [.28, .52], $t(601.00) = 6.46$, $p < .001$ in the REP condition, $b_2 = .34$ [.22, .47], $t(600.60) = 5.34$, $p < .001$ in the ES condition, and $b_3 = .44$ [.32, .56], $t(600.70) = 7.25$, $p < .001$ in the combined condition. This suggests that each of the three learning manipulations was sufficiently powerful to persist over a 15-min delay between the learning and test phases of the experiment.

Given that potential differences across learning conditions are also of interest, we conducted indirect testing using bootstrap samples (see Kurdi & Banaji, 2017) to probe for such differences. To control for the familywise error rate, a Bonferroni-corrected alpha level of $\alpha = .05/6 = .008$ was used. We obtained no significant differences between learning conditions, including between REP and ES $b_{REP–ES} = -.11$, 99.2% CI [−.26, .05],[20] $p > .008$, between REP and combined $b_{REP–COMB} = .02$, 99.2% CI [−.13, .17], $p > .008$, or between ES and combined $b_{ES–COMB} = .12$, 99.2% CI [−.02, .26], $p > .008$.[21] Thus, unlike in previous experiments in which implicit attitudes were measured immediately, neither the ES nor the ES + REP condition outperformed REP in isolation, indicating that evaluative learning involving verbal manipulations decays faster than learning from stimulus pairings alone.

Summary of supplementary analyses. In Supplement 1, we report additional analyses of the same data. To summarize, we demonstrate that (a) the evaluative learning effects reported here are significantly modulated by participants' conscious recollection of the learning manipulation (see Hofmann et al., 2010 and current Studies 1 and 2); (b) the focal result reported above, yielding significant learning effects in all learning conditions compared with control but no differences across learning conditions, also emerges in a Bayesian linear regression using uninformative priors; and (c) in a model including all participants without exclusions based on contingency awareness, both contingency aware-

[19] All stimuli are available for download from OSF (https://osf.io/serq4/).

[20] Given that a Bonferroni correction was used, we report the confidence interval corresponding to the Bonferroni-corrected alpha level of .008 rather than the more usual 95% confidence interval.

[21] The differences between the control condition and each learning condition survived Bonferroni correction.

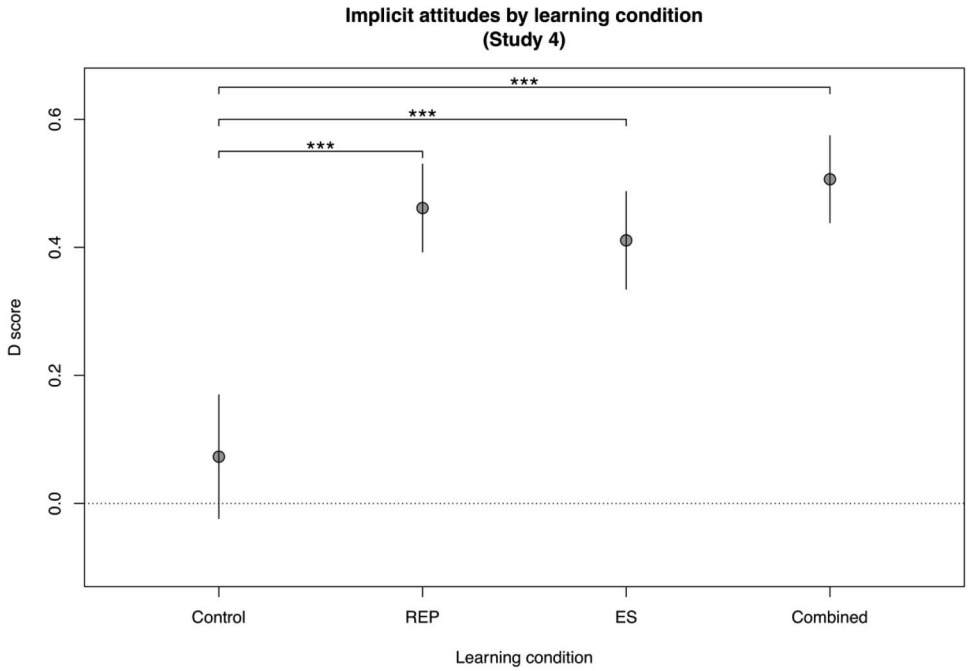

Figure 4. Implicit attitude strength (measured using IAT D scores) as a function of learning conditions (Study 4), including control, repeated evaluative pairings (REP), evaluative statements (ES), and ES + REP combined. Positive scores imply implicit attitudes in line with the learning manipulation. Error bars show 95% confidence intervals comparing each effect size to zero and differences significant at $p < .001$ are marked ***.

ness and the learning condition variable emerge as significant predictors of implicit attitude strength. Overall, these supplementary analyses demonstrate the robustness of the focal result of overall learning but no difference across learning conditions to (a) participant exclusions and (b) analytic frameworks.

Discussion

Study 4 constituted a second test of the idea that implicit evaluations created using different learning manipulations may decay differentially. In this study, unlike in previous work comparing the relative effectiveness of REP and ES in producing attitude change (Kurdi & Banaji, 2017) and most studies probing the effects of experimentally induced evaluative learning, implicit attitudes were assessed following a 15-min delay rather then immediately. Three findings from this study are noteworthy. First, implicit attitudes were significantly different from baseline in all three learning conditions, suggesting that the evaluative learning reported in previous work by Kurdi and Banaji (2017) was sufficiently powerful to persist over a 15-min delay. Second, whereas immediate measurement of attitudes produced a robust advantage of the learning conditions involving verbal manipulations (ES and ES + REP), the 15-min delay between learning and test used in the present study completely eliminated this advantage such that implicit attitudes across all three conditions were equivalent. This result demonstrates that, in line with the findings from Study 3, as well studies on episodic memory for personally experienced versus verbally described events (Baggett, 1979; Baggett & Ehrenfeucht, 1983; Beentjes & van der Voort, 1991; Larsen & Plunkett, 1987; Toglia et al., 1992), the memory advantage of verbal instructions over direct experience with stimulus pairings can be subject to fast decay. Third, again in line with Study 3, the ES and ES + REP conditions behaved similarly. Given that both the REP and the ES + REP conditions involve actual experience with stimulus pairings, this result indicates that REP may create more stable implicit attitudes due to the self-generated nature of the propositional inference that is required for evaluative learning to be successful in this condition (Craik & Lockhart, 1972; Toglia et al., 1992) rather than because stimulus pairings are inherently more effective in shifting implicit evaluations than verbal instructions.

General Discussion

In recent work, Kurdi and Banaji (2017) provided evidence that evaluative statements (ES) merely informing participants of upcoming stimulus pairings without actual exposure are more effective in shifting implicit attitudes than repeated evaluative pairings (REP) of category members with valenced images. Moreover, a combined condition of ES + REP did not result in more pronounced attitude change than ES in isolation, suggesting that exposure to stimulus pairings did not produce added value. Taken together, these findings indicated that, surprisingly, REP and ES may shift implicit attitudes via the same learning process, although ES seems to be more effective on an immediate test of implicit attitudes. To further explore the nature of these effects and the learning processes giving rise to them, the present work implemented tests of (a) the unfolding of learning as a result of increasing the number of stimulus exposures (Study 1), (b) the interaction between stimulus pairings and verbal information in producing implicit attitude change (Study 2), and (c) the temporal stability of

implicit attitude change (Studies 3–4). In addition to their theoretical relevance, these studies also have implications for the more applied issue of producing enduring change in preexisting implicit attitudes.

In Study 1, the number of stimulus pairings presented in the REP condition did not modulate the strength of implicit attitudes created. Four stimulus pairings, presented over a duration of only 10 seconds, were sufficient to produce attitude change and were found to be just as effective as 24 stimulus pairings. These results suggest that no protracted experience with stimulus pairings is necessary for implicit attitude change to occur. Moreover, they seem difficult to reconcile with the idea of a slow and incremental learning process.

Study 2 provides evidence that, much like other kinds of verbal information specifying the nature of the relationship between the stimuli presented (Moran et al., 2015, 2016; Peters & Gawronski, 2011; Zanon et al., 2014), manipulations of diagnosticity (describing stimulus pairings as randomly generated vs. revealing the underlying true nature of the target categories) can influence the magnitude of learning from stimulus pairings experienced in the environment. In combination with previous work, this finding shows that such learning is informationally promiscuous. Moreover, it demonstrates, for the first time, that manipulations of diagnosticity can be used to modulate the effects of evaluative learning not only from verbal statements (Cone & Ferguson, 2015) but also from stimulus pairings.

Third, across Studies 3 and 4, we have shown that implicit attitude change created by exposure to stimulus pairings is more durable than implicit attitude change created by verbal instructions merely signaling upcoming stimulus pairings. Moreover, the combined ES + REP condition exhibited the same precipitous decay in evaluative learning effects as the condition using ES alone. As such, this finding suggests that whatever feature of the REP manipulation leads to superior retention over time, this feature is eliminated by the presence of verbal instructions specifying the nature of stimulus pairings to which the participant will be exposed.

Theoretical Implications

We believe that the present results are primarily of interest because they have the ability to constrain any future theory of implicit attitude acquisition, be they associative, propositional, an associative/propositional hybrid, or of any other flavor. Specifically, based on the findings of the recent work by Kurdi and Banaji (2017), combined with the current studies, any theory of implicit evaluation must be able to explain why and how (a) a combined intervention of ES + REP produces equivalent learning to ES alone; (b) stimulus pairings presented beyond the initial four do not produce any further learning in REP; (c) learning in REP is integrated with verbal information on the nature of stimulus pairings; and (d) REP in isolation, but not a combination of ES + REP, produces more durable evaluative learning than ES alone. Although these results seem generally more compatible with a quick, inferential, and informationally promiscuous learning process that stores propositions in long-term memory (Csibra & Gergely, 2009; Gopnik, 1996; Tenenbaum et al., 2011) than with a slow, stimulus-driven, and informationally encapsulated that stores conceptual associations in long-term memory (Rescorla & Wagner, 1972; Rumelhart et al., 1986; Sutton & Barto, 1998), we remain open to the possibility that a fundamentally different kind of account may provide a better explanation of the same data in the future.

When it comes to currently available associative theories of implicit evaluation (McConnell & Rydell, 2014; Rydell & McConnell, 2006; E. R. Smith & DeCoster, 2000; Strack & Deutsch, 2004), these theories seem to be difficult to reconcile with several aspects of the present results. First, these theories posit that implicit evaluations should shift as a result of slow and incremental updating processes. However, in Study 1, learning emerged and stabilized quickly. Second, these theories suggest that implicit evaluations should rely exclusively, or at least preferentially, on associative rather than verbal interventions. And yet, in Study 2, implicit evaluations emerged from a combination of both kinds of information. Finally, the fact that stimulus pairings gave rise to more enduring attitude change than verbal statements (Studies 3 and 4) seems to be well accounted for by associative theories; however, such theories do not appear to be well equipped to explain why the benefit of stimulus pairings is eliminated when they are preceded by verbal information. As such, the present results also seem to call into question models of evaluative conditioning as a purely stimulus-driven and bottom-up process that bypasses high-level reasoning (Baeyens et al., 1992, 2009; Gawronski et al., 2014; Gawronski & Bodenhausen, 2006; Levey & Martin, 1975; Martin & Levey, 1978; Rydell & McConnell, 2006; Strack & Deutsch, 2004). At the same time, new associative theories of implicit evaluation may be able to account for the findings of Study 1 to the extent that, unlike currently available associative theories, they allow for single-shot associative learning. However, it is more difficult to see how purely associative theories would be able to deal with findings of interaction between verbal information and stimulus pairings in giving rise to implicit attitudes (Kurdi & Banaji, 2017; current Studies 2–4).

Overall, current propositional and symbolic theories of implicit evaluation (De Houwer, 2009, 2014, 2018; De Houwer & Hughes, 2016; Mitchell et al., 2009) seem to be better suited to account for the present results. Specifically, although these theories do not require that the number of stimulus pairings never modulate the strength of learning from REP, they allow for the possibility of quick propositional inferences following limited exposure (Study 1). Second, these theories posit that propositional learning should productively combine all available evidence, whether it be presented in pictorial or verbal format, in shifting implicit attitudes; the results of Study 2 are largely in line with this idea. Finally, even though propositional theories did not a priori predict the results of Studies 3 and 4, they suggest that REP should shift implicit attitudes via propositional processes and, as such, can easily incorporate past findings on the memory advantage of self-generated propositions (Craik & Lockhart, 1972; Toglia et al., 1992). At the same time, as currently formulated, propositional theories do not seem to be well-equipped to explain why diagnosticity information should affect implicit attitude acquisition from REP only if presented before, but not if presented after, exposure to stimulus pairings (Study 2). Below we provide a potential explanation of this finding within the broad framework of inferential approaches to human learning.

Although the current findings are difficult to explain under associative theories of implicit evaluation, they may be compatible with more recent versions of a hybrid model of implicit evaluation,

specifically the Associative–Propositional Evaluation model (APE; Gawronski & Bodenhausen, 2006, 2011, 2018). Indeed, since its inception, the APE model has recognized that implicit evaluations may be updated as a result of associative processes, propositional processes, or a combination of both. Moreover, after initially introducing REP as "[t]he prototypical case for implicit attitude changes resulting from changes in associative structure" (Gawronski & Bodenhausen, 2006, p. 697), more recent versions of the APE model allow for the possibility that REP-based learning may be "driven by associative learning under some conditions, but by propositional learning under other conditions" (Gawronski & Bodenhausen, 2011, p. 108). Gawronski and Bodenhausen (2018) go even further in theorizing that "repeated pairings of a CS and a US can influence mental representations via associative learning, propositional learning, or both" (p. 2). However, given that the propositional perspective is (a) able to accommodate the bulk of present findings using one, rather than two, learning processes and (b) more restricted in its predictions than the APE model, we believe that propositional models are preferable as the most parsimonious and well-specified account of the present results.

Furthermore, it should be noted that some may wish to characterize the theories that we refer to as associative in the present work (McConnell & Rydell, 2014; Rydell & McConnell, 2006; Smith & DeCoster, 2000; Strack & Deutsch, 2004) as hybrid theories given that these theories incorporate certain limited interactions between the posited explicit and implicit systems. In fact, Smith and DeCoster (2000) allow for one kind of explicit–implicit interaction in explaining that repeated use of a step-by-step rule-based process may give rise to associative representations (see also Strack & Deutsch, 2004). This stipulation does not apply to the present work given that none of the present studies involved the kind of protracted practice described by Smith and DeCoster. Strack and Deutsch (2004) further allow for the possibility that elements of propositions represented in the explicit system may be retrieved from the implicit system. This, again, seems irrelevant to the present experiments.

Finally, at the end of their chapter, McConnell and Rydell (2014) speculate that the explicit and implicit systems may interact with each other (a) by taking the same information as input, (b) sequentially, with one system taking the output from the other system as its input, or (c) by relying on past information processing involving both systems. We see this speculation as fundamentally inconsistent with the rest of McConnell and Rydell's systems of evaluation model, which describes explicit and implicit systems of evaluation as "dissociable systems of knowledge" (p. 204), with "each system becom[ing] more fully engaged with information to which it is most sensitive and neglect[ing] (at least in part) information to which it is less sensitive" (p. 213). However, to the extent that McConnell and Rydell wish to posit the same kind of intricate pattern of explicit–implicit interactions as the APE model described above, then it is our view that their theory will face similar challenges in terms of its falsifiability.

Methodological Implications

In addition to these substantive findings, we also introduced a new measure of performance on the IAT (Greenwald et al., 1998). This new measure is based on the improved scoring algorithm proposed by Greenwald et al. (2003) but, unlike the improved scoring algorithm, takes into account the temporal unfolding of response latencies within the IAT session. This new scoring method can be used to obtain mini D scores that index changes in participants' IAT performance as the critical blocks progress. Given that mini D scores rely on data obtained within a single IAT session, they can be used to probe the temporal stability of implicit attitudes without the need for recruiting participants for studies that extend across multiple measurement occasions. Importantly, the results that we obtained using mini D scores were (a) robust to the size of the moving average window chosen and (b) consistent with the findings of an experiment that relied on a multisession design (Study 4), thus providing evidence for the face validity of this method. We hope that future work will (a) subject the validity of mini D scores to further empirical scrutiny and (b) more generally probe whether the results reported in the present paper generalize to measures of implicit evaluation other than the IAT.

A Call for Computational Modeling

Further, the present project suggests two additional areas for exploration that we believe may be crucial to understanding how implicit attitudes change. First, current associative (McConnell & Rydell, 2014; Rydell & McConnell, 2006; Smith & DeCoster, 2000; Strack & Deutsch, 2004), propositional/symbolic (De Houwer, 2009, 2014, 2018; De Houwer & Hughes, 2016), and hybrid (Gawronski & Bodenhausen, 2006, 2011, 2018) accounts of implicit evaluation are best understood as general conceptual frameworks rather than as theories sufficiently specific to make quantitative predictions about the computations that the human mind should perform in response to different kinds of evaluative information. Given that current theories of implicit evaluation are formulated in a purely verbal manner, without making reference to potential computational instantiations of the learning processes that they posit, they are difficult to falsify.

Even though all computational models are wrong (Box, 1976) and, quite possibly, even stupid (Smaldino, 2017), they provide the tools necessary for addressing questions left open by the current experiments as well as by this body of work as a whole. Such open questions include (a) how exactly verbal information and actually experienced stimulus pairings interact with each other in producing implicit attitude change (Studies 1 and 2), (b) why more confirming instances of Group A paired with positive stimuli and Group B paired with negative stimuli sometimes do not lead to more belief revision (Study 1), and (c) why diagnosticity information presented before but not after exposure to stimulus pairings has an effect on evaluative learning (Study 2).

A possible answer to these open questions may be found within the framework of hierarchical Bayesian models of cognition (Tenenbaum et al., 2011), which provide a computational account of human reasoning under conditions of uncertainty. Bayesian computational modeling encompasses both (a) the selection of a generative model that can explain data encountered in the world and (b) the estimation of the parameters of that model, both via updating prior beliefs in the face of new information. In this particular case, Bayesian modeling may facilitate understanding how different kinds of information, such as verbal information about stimulus pairings and information provided by the stimulus pairings themselves, are productively combined with each other to

arrive at the appropriate understanding of the causal processes through which the stimulus pairings were generated.

More specifically, the verbal instructions provided prior to exposure to stimulus pairings can be understood as influencing participants' prior beliefs about the upcoming task by constraining the hypothesis space that they should reasonably entertain. For instance, in the REP condition of the current studies, participants were informed that they would see two types of targets (CS) and two types of drawings (US) and were instructed to learn the general relationship between a certain type of target and a certain type of drawing. Given this prior information, once participants are able to infer that the two types of images are instances of the general categories "good" and "bad," there are only two potential hypotheses remaining: (H_1) Group A always paired with positive images and Group B always paired with negative images or (H_2) Group A always paired with negative images and Group B always paired with positive images. At this point, a single US–CS pairing already provides conclusive evidence to help decide between H_1 and H_2. This may be the reason why number of stimulus pairings did not produce an effect in Study 1: Because the hypothesis space was heavily constrained by initial instructions, four stimulus pairings were sufficient for participants to make the valence inference and form posterior beliefs decisively in favor of either H_1 or H_2.

Such a Bayesian perspective may also be able to account for why the diagnosticity manipulation affected evaluative learning when verbal information was provided to participants before, rather than after, exposure to stimulus pairings. If participants receive prior information that they are about to learn something fundamental about the nature of the target categories, this implies a highly structured hypothesis space in which Group A and B are each paired with a unique set of USs. By contrast, if participants believe that the stimulus pairings are randomly generated, that would imply that any CS may be paired with any US. Given that the hypothesis space is considerably more constrained in the former case, the same amount of data should create more learning than in the latter case. However, when participants do not receive prior instructions about diagnosticity, they may reasonably assume that the stimulus pairings provided by the experimenter are relevant and informative (Grice, 1975), that is, the hypothesis space that they consider may be quite limited. For instance, participants may be unlikely to spontaneously entertain the possibility that, in a learning experiment, stimuli will be paired with each other in a fully randomized manner. If this is the case and, therefore, participants' learning from stimulus pairings is effective, post hoc information about diagnosticity may not eliminate the robust learning effects already in place.

Regardless of whether these specific conjectures are accurate or not, future work involving Bayesian modeling will be able to provide a computational account of how information presented in different formats (e.g., verbally vs. pictorially) is integrated in updating implicit evaluations. As such, Bayesian modeling may be seen as a computationally tractable instantiation of propositional and symbolic accounts of implicit evaluation (De Houwer, 2009, 2014, 2018; De Houwer & Hughes, 2016), which posit that implicit attitudes are updated in an inferentially promiscuous, rather than in a purely bottom-up and stimulus-driven, manner. Models of causal reasoning implemented in a Bayesian framework (Gopnik et al., 2004; Lu, Rojas, Beckers, & Yuille, 2016) may serve as a useful starting point for implementing computational models of the propositional processes resulting in the updating of (implicit) evaluations.

Implications for Long-Term Change

Although the present studies were conducted within a relatively constrained time span between learning and test, our findings have two kinds of implications for long-term change in implicit attitudes: (a) general implications for how the issue of long-term change should be conceptualized and (b) specific implications for interventions that might be successful in shifting implicit attitudes in a durable way.

Implicit attitudes are often characterized as recalcitrant and resistant to change (Bargh, 1999; Gregg et al., 2006; Rydell & McConnell, 2006). Accordingly, attempts at creating long-term change in implicit attitudes toward social groups are often unsuccessful (Forscher et al., 2017; Lai et al., 2016). However, at the same time, a recent analysis of data from more than 4 million participants from the Project Implicit educational website (http://implicit.harvard.edu/) has demonstrated sizable long-term shifts toward neutrality in implicit attitudes toward race, skin tone, and sexual orientation (Charlesworth & Banaji, 2019). If the effects of targeted interventions created in the lab dissipate within a few hours, how is it possible for implicit attitudes to change in the long run? The present project may provide some pointers toward a potential explanation.

The results of Study 4 suggest that the effects of interventions like repeated evaluative pairings and evaluative statements are sufficiently robust to survive a 15-min delay between the learning and test phases of the experiment. As such, this project demonstrates that changes in implicit attitudes created via such interventions are not fully ephemeral and do not depend on the information being actively rehearsed between learning and testing. Therefore, it seems that studies probing the effects of relatively minimal manipulations after relatively long delays (Forscher et al., 2017; Lai et al., 2016) may have been unsuccessful in creating long-term change not because the interventions themselves are ineffective. Rather, participants encounter members of social groups in evaluatively consequential contexts, as well as evaluatively relevant verbal information about them, thousands of times every day. Such information may have acted as spontaneous sources of interference and may have eliminated the effects of one-shot interventions.

Thus, if progress toward understanding long-term shifts in implicit evaluations is of interest, future studies may want to address this issue in one of two ways. First, they might examine the potential for long-term change through administering interventions that seem effective in the medium term (such as within a timeframe of hours) multiple times a day to counteract forgetting and interference. Second, a recent theoretical piece by Payne, Vuletich, and Lundberg (2017) suggests that, in addition to exploring implicit attitude change at the individual level, investigating change in the aggregate may also be a fruitful avenue of research. As such, one might use archival data to correlate changes or differences in potential antecedents to implicit attitudes across times or geographic areas with aggregate changes or differences across the same times or areas.

Finally, the present project also has implications for the kinds of interventions whose effects should be relatively durable in creating implicit attitude change. Specifically, we found that repeated eval-

uative pairings resulted in more robust shifts in implicit attitudes over time than evaluative statements. However, this was only true when participants were required to make extensive inferences about the stimulus pairings themselves, without those inferences being provided to them in verbal form. In line with research exploring episodic memory for personally experienced versus merely verbally described events (Baggett, 1979; Baggett & Ehrenfeucht, 1983; Beentjes & van der Voort, 1991; Larsen & Plunkett, 1987; Toglia et al., 1992), this finding suggests that evaluative learning may be made more memorable by engaging participants in deeper processing of the stimuli. In addition, exposure to category members paired with positive images may create less reactance than verbal messages with the obvious intent to push participants toward change (De Houwer & Hughes, 2016).

However, long-term effectiveness of primarily nonverbal manipulations presupposes that participants have the ability to appropriately categorize stimuli and to detect spatiotemporal contingencies between them. This assumption may not always be warranted, as evidenced by high degrees of variability in initial learning effects from repeated evaluative pairings (Kurdi & Banaji, 2017). Thus, future attempts at shifting implicit attitudes may be particularly effective if interventions are designed to be sufficiently challenging to require participants' attention and active engagement but not as challenging as to make it impossible for participants to make the intended inferences.

References

Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *A handbook of social psychology* (pp. 798–844). Worcester, MA: Clark University Press.

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

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

Baeyens, F., Vansteenwegen, D., & Hermans, D. (2009). Associative learning requires associations, not propositions. *Behavioral and Brain Sciences, 32,* 198–199. http://dx.doi.org/10.1017/S0140525X09000867

Baggett, P. (1979). Structurally equivalent stories in movie and text and the effect of the medium on recall. *Journal of Verbal Learning & Verbal Behavior, 18,* 333–356. http://dx.doi.org/10.1016/S0022-5371(79)90191-9

Baggett, P., & Ehrenfeucht, A. (1983). Encoding and retaining information in the visuals and verbals of an educational movie. *Educational Communication & Technology, 31,* 23–32.

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. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). 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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48. http://dx.doi.org/10.18637/jss.v067.i01

Beentjes, J. W. J., & van der Voort, T. H. A. (1991). Recall and language usage in retellings of televised and printed stories. *Poetics, 20,* 91–104. http://dx.doi.org/10.1016/0304-422X(91)90035-N

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6,* 242–261. http://dx.doi.org/10.1207/S15327957PSPR0603_8

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin, 114,* 80–99. http://dx.doi.org/10.1037/0033-2909.114.1.80

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association, 71,* 791–799. http://dx.doi.org/10.1080/01621459.1976.10480949

Cacioppo, J. T., Harkins, S. G., & Petty, R. E. (1981). The nature of attitudes and cognitive responses and their relationships to behavior. In R. Petty, T. Ostrom, & T. Brock (Eds.), *Cognitive responses in persuasion* (pp. 31–54). Hillsdale, NJ: Erlbaum.

Charlesworth, T. E. S., & Banaji, M. R. (2019). Patterns of implicit and explicit attitudes I. Long-term change and stability from 2007 – 2016. *Psychological Science*. Advance online publication. http://dx.doi.org/10.1177/0956797618813087

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

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11,* 671–684. http://dx.doi.org/10.1016/S0022-5371(72)80001-X

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13,* 148–153. http://dx.doi.org/10.1016/j.tics.2009.01.005

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81,* 800–814. http://dx.doi.org/10.1037/0022-3514.81.5.800

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. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology, 10,* 230–241. http://dx.doi.org/10.1017/S1138741600006491

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior, 37,* 1–20. 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. (2018). Propositional models of evaluative conditioning. *Social Psychological Bulletin, 13,* e28046. http://dx.doi.org/10.5964/spb.v13i3.28046

De Houwer, J., & Hughes, S. (2016). Evaluative conditioning as a symbolic phenomenon: On the relation between evaluative conditioning, evaluative conditioning via instructions, and persuasion. *Social Cognition, 34,* 480–494. http://dx.doi.org/10.1521/soco.2016.34.5.480

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

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking

intervention. *Journal of Experimental Social Psychology, 48,* 1267–1278. http://dx.doi.org/10.1016/j.jesp.2012.06.003

Dickinson, A. (2001). The 28th Bartlett Memorial Lecture. Causal learning: An associative analysis. *The Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology, 54,* 3–25. http://dx.doi.org/10.1080/02724990042000010

Drew, M. R., Denny, C. A., & Hen, R. (2010). Arrest of adult hippocampal neurogenesis in mice impairs single- but not multiple-trial contextual fear conditioning. *Behavioral Neuroscience, 124,* 446–454. http://dx.doi.org/10.1037/a0020081

Eagly, A. H., & Chaiken, S. (1998). Attitude structure and function. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 269–322). New York, NY: McGraw-Hill.

Ebbinghaus, H. (2011). *Memory: A contribution to experimental psychology*. Eastford, CT: Martino Publishing. (Original work published 1913)

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

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

Fodor, J. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Forscher, P. S., Mitamura, C., Dix, E. L., Cox, W. T. L., & Devine, P. G. (2017). Breaking the prejudice habit: Mechanisms, timecourse, and longevity. *Journal of Experimental Social Psychology, 72,* 133–146. http://dx.doi.org/10.1016/j.jesp.2017.04.009

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

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

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

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132,* 692–731. 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. In J. M. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (Vol. 44, 1st ed., pp. 59–127). Waltham, MA: Elsevier. http://dx.doi.org/10.1016/B978-0-12-385522-0.00002-0

Gawronski, B., & Bodenhausen, G. V. (2018). Evaluative conditioning from the perspective of the Associative–Propositional Evaluation Model. *Social Psychological Bulletin, 13*(3), e28024. Advance online publication. http://dx.doi.org/10.5964/spb.v13i3.28024

Gershman, S. J., Norman, K. A., & Niv, Y. (2015). Discovering latent causes in reinforcement learning. *Current Opinion in Behavioral Sciences, 5,* 43–50. http://dx.doi.org/10.1016/j.cobeha.2015.07.007

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

Goodman, N. D., Tenenbaum, J., Feldman, J., & Griffiths, T. (2008). A rational analysis of rule-based concept learning. *Cognitive Science, 32,* 108–154. http://dx.doi.org/10.1080/03640210701802071

Gopnik, A. (1996). The scientist as child. *Philosophy of Science, 63,* 485–514. http://dx.doi.org/10.1086/289970

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111,* 3–32. http://dx.doi.org/10.1037/0033-295X.111.1.3

Greenwald, A. G., & Banaji, M. R. (1989). The self as a memory system: Powerful, but ordinary. *Journal of Personality and Social Psychology, 57,* 41–54. http://dx.doi.org/10.1037/0022-3514.57.1.41

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., & 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

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

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41–58). New York, NY: Academic Press.

Griffiths, T. L., & Kalish, M. L. (2010). Language evolution by iterated learning with Bayesian agents. *Cognitive Science, 31,* 441–480. http://dx.doi.org/10.1080/15326900701326576

Grumm, M., Nestler, S., & 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

Heycke, T., Gehrmann, S., Haaf, J. M., & Stahl, C. (2018). Of two minds or one? A registered replication of Rydell et al. (2006). *Cognition and Emotion, 32,* 1708–1727. http://dx.doi.org/10.1080/02699931.2018.1429389

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

Hu, X., Gawronski, B., & Balas, R. (2017). Propositional versus dual-process accounts of evaluative conditioning: I. The effects of co-occurrence and relational information on implicit and explicit evaluations. *Personality and Social Psychology Bulletin, 43,* 17–32.

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. http://dx.doi.org/10.1007/BF03395768

Hughes, S., Ye, Y., Van Dessel, P., & De Houwer, J. (2018). When people co-occur with good or bad events: Graded effects of relational qualifiers on evaluative conditioning. *Personality and Social Psychology Bulletin, 45,* 196–208. http://dx.doi.org/10.1177/0146167218781340

Hütter, M., & De Houwer, J. (2017). Examining the contributions of memory-dependent and memory-independent components to evaluative conditioning via instructions. *Journal of Experimental Social Psychology, 71,* 49–58. http://dx.doi.org/10.1016/j.jesp.2017.02.007

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution

to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103,* 54–69. http://dx.doi.org/10.1037/a0028347

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44. http://dx.doi.org/10.1037/0033-295X.99.1.22

Kurdi, B., & Banaji, M. R. (2017). Repeated evaluative pairings and evaluative statements: How effectively do they shift implicit attitudes? *Journal of Experimental Psychology: General, 146,* 194–213. http://dx.doi.org/10.1037/xge0000239

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods, 49,* 457–470. http://dx.doi.org/10.3758/s13428-016-0715-3

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., . . . Banaji, M. R. (2018). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist*. Advance online publication. http://dx.doi.org/10.1037/amp0000364

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

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

Larsen, S. F., & Plunkett, K. (1987). Remembering experienced and reported events. *Applied Cognitive Psychology, 1,* 15–26. http://dx.doi.org/10.1002/acp.2350010104

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

Levey, A. B., & Martin, I. (1987). Evaluative conditioning: A case for hedonic transfer. In H. J. Eysenck & I. Martin (Eds.), *Theoretical foundations of behavior therapy* (pp. 113–131). Boston, MA: Springer U.S. http://dx.doi.org/10.1007/978-1-4899-0827-8_5

Lewin, K. (1935). *A dynamic theory of personality*. New York, NY: McGraw-Hill.

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology, 81,* 842–855. http://dx.doi.org/10.1037/0022-3514.81.5.842

Lu, H., Rojas, R. R., Beckers, T., & Yuille, A. L. (2016). A Bayesian theory of sequential causal learning and abstract transfer. *Cognitive Science, 40,* 404–439. http://dx.doi.org/10.1111/cogs.12236

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science, 26,* 390–395. http://dx.doi.org/10.1177/0963721417691356

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

Mann, T. C., & Ferguson, M. J. (2017). Reversing implicit first impressions through reinterpretation after a two-day delay. *Journal of Experimental Social Psychology, 68,* 122–127. http://dx.doi.org/10.1016/j.jesp.2016.06.004

Martin, I., & Levey, A. B. (1978). Evaluative conditioning. *Advances in Behaviour Research and Therapy, 1,* 57–101. http://dx.doi.org/10.1016/0146-6402(78)90013-9

McConnell, A. R., & Rydell, R. J. (2014). The systems of evaluation model: A dual-systems approach to attitudes. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 204–217). New York, NY: Guilford 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

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

Moran, T., Bar-Anan, Y., & Nosek, B. A. (2015). Processing goals moderate the effect of co-occurrence on automatic evaluation. *Journal of Experimental Social Psychology, 60*(C), 157–162. http://dx.doi.org/10.1016/j.jesp.2015.05.009

Moran, T., Bar-Anan, Y., & Nosek, B. A. (2016). The assimilative effect of co-occurrence on evaluation above and beyond the effect of relational qualifiers. *Social Cognition, 34,* 435–461. http://dx.doi.org/10.1521/soco.2016.34.5.435

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & Cognition, 4,* 648–654. http://dx.doi.org/10.3758/BF03213230

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. (2002). Implicit acquisition and manifestation of classically conditioned attitudes. *Social Cognition, 20,* 89–104. http://dx.doi.org/10.1521/soco.20.2.89.20992

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

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry, 28,* 233–248. http://dx.doi.org/10.1080/1047840X.2017.1335568

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

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

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). East Norwalk, CT: Appleton-Century-Crofts.

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

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103,* 734–760. http://dx.doi.org/10.1037/0033-295X.103.4.734

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323,* 533–536. http://dx.doi.org/10.1038/323533a0

Rumelhart, D. E., & McClelland, J. L. (1987). Learning the past tenses of English verbs: Implicit rules or parallel distributed processing. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 195–248). Hillsdale, NJ: Erlbaum.

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

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

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 592–604. http://dx.doi.org/10.1037/0278-7393.4.6.592

Smaldino, P. E. (2017). Models are stupid, and we need more of them. In R. R. Vallacher, A. Nowak, & S. J. Read (Eds.), *Computational models in social psychology* (pp. 311–331). New York, NY: Routledge. http://dx.doi.org/10.4324/9781315173726-14

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

Stewart, B. D., & Payne, B. K. (2008). Bringing automatic stereotyping under control: Implementation intentions as efficient means of thought control. *Personality and Social Psychology Bulletin, 34,* 1332–1345. http://dx.doi.org/10.1177/0146167208321269

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

Stuart, E. W., Shimp, T. A., & Engle, R. W. (1987). Classical conditioning of consumer attitudes: Four experiments in an advertising context. *Journal of Consumer Research, 14,* 334–349. http://dx.doi.org/10.1086/209117

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science, 331,* 1279–1285. http://dx.doi.org/10.1126/science.1192788

Toglia, M. P., Shlechter, T. M., & Chevalier, D. S. (1992). Memory for directly and indirectly experienced events. *Applied Cognitive Psychology, 6,* 293–306. http://dx.doi.org/10.1002/acp.2350060403

Van Dessel, P., De Houwer, J., & Gast, A. (2015). Approach–avoidance training effects are moderated by awareness of stimulus–action contingencies. *Personality and Social Psychology Bulletin, 42,* 81–93. http://dx.doi.org/10.1177/0146167215615335

Van Dessel, P., De Houwer, J., Gast, A., & Smith, C. T. (2015). Instruction-based approach–avoidance effects. *Experimental Psychology, 62,* 161–169. http://dx.doi.org/10.1027/1618-3169/a000282

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., Roets, A., & Gast, A. (2016). Failures to change stimulus evaluations by means of subliminal approach and avoidance training. *Journal of Personality and Social Psychology, 110*(1), e1–e15. http://dx.doi.org/10.1037/pspa0000039

Van Dessel, P., Gawronski, B., Smith, C. T., & De Houwer, J. (2017). Mechanisms underlying approach–avoidance instruction effects on implicit evaluation: Results of a preregistered adversarial collaboration. *Journal of Experimental Social Psychology, 69*(C), 23–32. http://dx.doi.org/10.1016/j.jesp.2016.10.004

Van Dessel, P., Mertens, G., Smith, C. T., & De Houwer, J. (2017). The mere exposure instruction effect: Mere exposure instructions influence liking. *Experimental Psychology, 64,* 299–314. http://dx.doi.org/10.1027/1618-3169/a000376

Willis, J., & Todorov, A. T. (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

Wilson, W. R., & Miller, H. (1964). A note on the inconclusiveness of accepting the null hypothesis. *Psychological Review, 71,* 238–242. http://dx.doi.org/10.1037/h0046217

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology, 81,* 815–827. http://dx.doi.org/10.1037/0022-3514.81.5.815

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

Received January 29, 2018
Revision received December 18, 2018
Accepted January 1, 2019 ■