Original Manuscript

Persuading the Implicit Mind: Changing Negative Implicit Evaluations With an 8-Minute Podcast

Social Psychological and Personality Science
1-10

DOI: 10.1177/19485506211037140
journals.sagepub.com/home/spp
SAGE

Benedek Kurdi[1], Thomas C. Mann[2], and Melissa J. Ferguson[1]

Abstract

Implicit evaluations can be malleable via reinterpretation of previously encountered evidence. Here, we report three studies ($N = 1{,}007$) investigating the robustness of this updating modality using ecologically realistic materials. Participants were first introduced to a target who killed an endangered black rhino in Namibia. They then listened to a real podcast providing counterattitudinal information on the benefits of trophy hunting. The podcast resulted in considerable revisions of initially negative implicit evaluations toward positivity, consistently across implicit measures (affect misattribution procedures vs. implicit association test), samples (American students vs. nonstudents from various countries), study settings (lab vs. online), and the presence versus absence of a memory retrieval manipulation prompting reflection on participants' views on trophy hunting. Taken together, these findings suggest that reinterpretation can shift implicit evaluations of even highly negative targets, including under conditions of external validity.

Keywords

attitude change, implicit evaluations, persuasion, reinterpretation

Meet Corey Knowlton. He is 41 years old and resides in Dallas, TX, with his wife and two children. In 2014, he paid $350,000 to kill an endangered black rhino in Namibia and, despite considerable controversy, did so in 2015.

If, like the most Americans (Kellert, 1993), you are unfavorably disposed toward trophy hunting, reading even the minimal passage above was probably sufficient for you to develop highly negative views of Knowlton. These negative views are likely strong enough to emerge not only on explicit (self-report) measures but even on their implicit counterparts assumed to reflect more automatic aspects of thought and feeling (Devine, 1989; Fazio et al., 1986; Greenwald & Banaji, 1995). But does the extremity of this negative initial impression necessarily imply that it will be resistant to persuasion? Or could it be malleable and, if so, under what conditions?

In this project, we use Knowlton's highly publicized story as a case study to test whether even highly negative implicit evaluations, such as the ones created via the short vignette above, are amenable to updating using strong counterattitudinal arguments. In doing so, we build and expand upon previous work on implicit attitude change in general (Cone et al., 2017; De Houwer et al., 2020; Kurdi & Dunham, 2020) and implicit attitude change via reinterpretation in particular (Mann & Ferguson, 2015, 2017; Wyer, 2010, 2016).

Implicit Attitude Change Via Reinterpretation

Most past work examining the possibility of implicit attitude change has used entirely novel counterattitudinal information to induce updating. In contrast, a small set of studies relying on the idea of reinterpretation have introduced counterattitudinal information that prompts a different interpretation of some old information. For instance, Mann and Ferguson (2015) had participants read a vignette about a target named Francis West who breaks into two adjacent family homes to remove "precious things" from them. Subsequently, it is revealed that West, initially presumed to be a burglar, entered the homes because they were on fire and the "precious things" were young children, resulting in significant revision, and often full reversal, of implicit evaluations.

Notably, reinterpretation is the only learning modality that has reliably produced rapid reversals of initially negative

[1] Department of Psychology, Yale University, New Haven, CT, USA
[2] Department of Psychology, Harvard University, Cambridge, MA, USA

Corresponding Author:
Benedek Kurdi, Department of Psychology, Yale University, New Haven, CT 06511, USA.
Email: benedek.kurdi@yale.edu

implicit evaluations. Given the diagnosticity of negative information, this direction of updating has consistently been shown to be more difficult to achieve than updating in the other direction (e.g., Reeder & Brewer, 1979; Trafimow & Schneider, 1994). However, so far, reinterpretation-based updating of implicit evaluations has been demonstrated only in artificial paradigms in which participants are initially led down the wrong path only to then encounter new information that completely reverses initial interpretation of the events. In this way, an initial evaluatively extreme behavior (e.g., burglary) is fully replaced with another evaluatively extreme behavior (e.g., heroism).

Although this kind of "bait and switch" may happen from time to time under ecologically realistic conditions, it is far more common for individuals to learn new information that pushes them to reconsider initial evaluations by adding context, nuance, and detail. For example, one might learn about someone harming another person, but then find out that the harm was accidental or prompted by an altruistic motive (e.g., Young et al., 2007). These cases of updating are evaluatively complex in that new information does not fully invalidate the initial interpretation of the behavior but rather prompts reconsideration of its evaluative implications (e.g., Knobe, 2003). The additional complexity involved in inferring the evaluative implications of new information under these conditions raises the question of whether implicit evaluations are sensitive to such nuance.

Some theoretical perspectives that argue that implicit evaluations are immune, or at least less responsive, to high-level propositional reasoning (e.g., Rydell & McConnell, 2006; Smith & DeCoster, 2000) would predict that implicit evaluations may be relatively insensitive to such processes of updating. Moreover, such more complex cases of updating are characterized by the simultaneous presence of positive and negative information. As such, the same theoretical accounts would also likely predict that implicit evaluations should not be strongly moved given this evaluative complexity: After all, under these accounts, implicit evaluations are thought to reflect the sum total of positive and negative information encountered about an attitude object. By contrast, more recent theoretical perspectives have suggested that implicit evaluations can be sensitive to propositional reasoning emerging from persuasive messages and need not reflect the sum total of evaluative associations (e.g., Cone et al., 2017; De Houwer, 2014; Kurdi & Dunham, 2020). The studies reported below give us the opportunity to test these competing predictions.

The Present Work

In the present work, we probe whether implicit evaluations respond to ecologically realistic novel information with complex evaluative implications. Specifically, across three studies, we introduced participants to trophy hunter Corey Knowlton via a short vignette akin to the one included above. Participants then listened to excerpts from an episode of the *Radiolab* podcast that addresses the Knowlton case and provides convincing counterattitudinal information about it. For example, it is revealed that the trophy hunting program in Namibia targets only postreproductive rhinos that are likely to hurt and even kill conspecifics. Moreover, listeners learn that the funds raised through hunting licenses are spent on wildlife conservation and job creation.

This paradigm is an instance of the kind of complex updating discussed above. Participants first learn about Knowlton as a trophy hunter, and this fact is never challenged; rather, the new information introduces detail on previously unknown implications of the behavior (e.g., Knobe, 2003). In addition, the podcast, produced entirely outside the experimental context, contains ample negatively valenced information, whereas counterattitudinal information in past work was always consistently positive. Although some of this negative information may be seen as relatively incidental, the podcast repeatedly expresses ambivalence regarding the very practice of trophy hunting itself. Crucially, counterattitudinal material was produced by an award-winning radio show whose goal is to get listeners to consider another perspective on a matter that may initially seem straightforward. As such, in addition to probing whether updating occurs in response to attributionally and evaluatively nuanced information, we can establish whether ecologically valid materials can result in the updating of implicit evaluations via reinterpretation.

Finally, we also directly explore the replicability of reinterpretation effects and their robustness to variations of the paradigm. Specifically, in Study 2, we include a memory retrieval manipulation asking participants to reflect on their views regarding trophy hunting before being exposed to counterattitudinal information. Such a thought-listing manipulation may inoculate participants against counterattitudinal arguments (e.g., McGuire, 1961; McGuire & Papageorgis, 1961; Tannenbaum et al., 1966; but see Schiller et al., 2010), thus potentially undermining the effectiveness of the reinterpretation manipulation. Moreover, Studies 1–2 and Study 3 additionally differed from each other in terms of the implicit measure used, the type of sample recruited, and the setting in which the experiment was conducted, thus creating further ways to explore generalizability.

Studies 1–2

In Studies 1–2, we probe the effects of reinterpretation on implicit evaluations in a novel paradigm that, unlike relevant past work, uses materials created entirely outside the experimental context to relay counterattitudinal information to participants, thus improving ecological realism at the cost of experimental control. Crucially, this material contains ample negatively valenced information, expresses ambivalence regarding the actions performed by the target, and addresses a controversial issue. In addition, in Study 2, we examine whether a thought-listing manipulation instructing participants to reflect on their preexisting views regarding trophy hunting modulates the effectiveness of counterattitudinal material. As such, these studies provide us with multiple opportunities to

investigate the robustness of reinterpretation as a modality of updating implicit evaluations.

Method

Open Science Practices

All materials, data files, and analysis scripts are available for download from the Open Science Framework (https://osf.io/427mh/). We report how we determined the sample size, all data exclusions, all manipulations, and all measures in all studies. Study 3 was formally preregistered (https://aspredicted.org/dn3vi.pdf).

Participants and Design

In both studies, participants were undergraduate students ($N = 216$ and $N = 208$, respectively) recruited via the Cornell University study pool. They completed the study in the lab and received partial course credit. We made an a priori decision to recruit at least 200 participants and continue data collection until the end of the semester even if the sample size of 200 has been reached. Participants who did not complete both affect misattribution procedures (AMP; $n = 2$ in Study 1 and $n = 8$ in Study 2), pressed the same key on all AMP trials, suggesting noncompliance with instructions (ns = 7 and 10), or failed a manipulation check (ns = 3 and 2) were excluded from the analyses. The final sample size, following exclusions, was 204 in Study 1 and 208 in Study 2.

In Study 1, for the purposes of the Time 2 learning phase (see below), participants were randomly assigned to a control condition ($n = 102$) or an experimental condition ($n = 102$). In Study 2, no control condition was implemented. Instead, participants were randomly assigned to a no retrieval condition ($n = 104$), which was identical to the experimental condition of Study 1, or a retrieval condition ($n = 104$), in which participants were additionally prompted to reflect on their views regarding big-game hunting after reading the initial negative information about the target. In both studies, time of measurement (Time 1 vs. Time 2) and, for the implicit measure, type of prime (control vs. Knowlton) were additionally manipulated within participants.

Materials

Facial images. An image of Corey Knowlton was used to represent him throughout the study. An additional set of five images of White men with neutral facial expressions obtained from the Chicago face database (Ma et al., 2015) were used as control primes on the AMP.

Abstract images. Eighty computer-generated images were used as target stimuli on the AMP (Katz et al., 2020). These images were divided into two sets of 40 each (Set 1 and Set 2).

Podcasts. Two episodes of the *Radiolab* podcast were adapted for use in this project. Both episodes were edited down to a length of about 8 min to prevent participant fatigue. The episode entitled "≤ kg" was selected for use in the control condition (Study 1). This episode discusses the origins of the definitions of the units of measurement in the International System of Units. The episode is mildly evaluatively positive and does not contain any information related to Corey Knowlton or trophy hunting.

The episode entitled "The Rhino Hunter" was selected for use in the experimental condition and both conditions of Study 2. Featuring a combination of reporter narration and interviews with Knowlton and other individuals, this episode discusses the benefits of auctioning off permits to hunt black rhinos in countries such as Namibia. To summarize, such permits are granted in select cases to kill older, aggressive individuals that frequently endanger the black rhino population by injuring and even killing younger animals. Moreover, the money raised is used to protect endangered species from poachers and create new jobs for game wardens and trackers. In fact, according to the podcast, wildlife populations in Namibia have increased by 80% since the start of the program.

Procedure and Measures

Learning Phase 1. In both conditions of both studies, participants were first introduced to Corey Knowlton via a short vignette designed to induce negative evaluations of him, given general societal disapproval of trophy hunting (Kellert, 1993). In the retrieval condition of Study 2, after reading the vignette about Knowlton, participants were instructed to reflect on how they feel about trophy hunting and why they feel the way that they do. They were asked to list their thoughts in a textbox and were not allowed to proceed until they spent at least 2 min doing so.

Test Phase 1. Implicit and explicit evaluations of Knowlton were measured. Given the theoretical focus of the present work, implicit evaluations were always measured first.

Implicit evaluations. Implicit evaluations were measured using an AMP (Payne et al., 2005). On each trial of the AMP, a prime image (75 ms) was first followed by a blank screen (125 ms), then by a target image (100 ms), and finally a noise image until the participant entered a response. Participants were instructed to indicate the pleasantness of the target while disregarding the prime using the "K" key (more pleasant than average) or the "D" key (less pleasant than average).

Twenty control trials and 20 Knowlton trials were administered. On control trials, the prime image was randomly selected from the five control faces such that each control face was shown four times. On Knowlton trials, the face of Knowlton served as the prime image. The target image was randomly selected such that each of the 40 targets was displayed once. Use of target image Sets 1 and 2 on the two AMPs was counterbalanced. Trial order was individually randomized.

Explicit evaluations. Six semantic differential items were administered in the following order: unlikeable–likeable, bad–good,

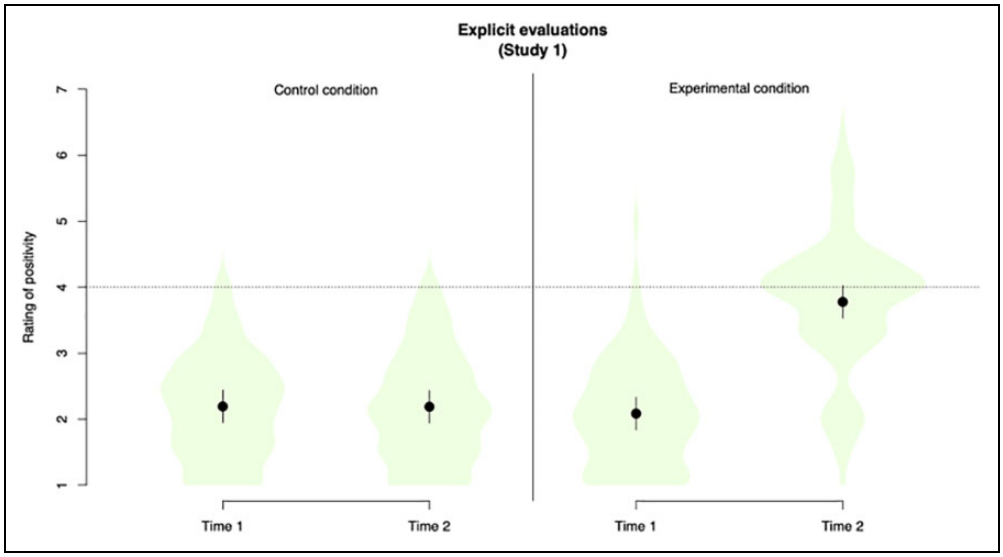

Figure 1. Distribution of explicit evaluations by condition and time of measurement (Study 1). The dashed horizontal line marks neutrality, and the solid dots show condition means. Error bars represent 95% confidence intervals derived from the mixed-effects model reported for the study. Positive scores indicate positive evaluations of Knowlton.

mean–pleasant, disagreeable–agreeable, uncaring–caring, cruel–kind. Responses were provided on a 7-point scale.

Learning Phase 2. Participants listened to one of the two podcast episodes: an attitudinally irrelevant podcast about the metric system (control condition of Study 1) or the counterattitudinal podcast about Corey Knowlton (experimental condition of Study 1 and both conditions of Study 2).

Test Phase 2. Implicit and explicit evaluations of Knowlton were measured anew in the same fixed order. A manipulation check item, reinterpretation items, and demographic items were also additionally administered.

Manipulation check. Participants selected a short description of the content of the podcast to which they had listened from four options.

Reinterpretation items. Participants completed five items measuring the extent to which information encountered during the second learning phase caused them to revise their views of Knowlton and big-game hunting. Responses were provided on a 9-point scale.

Demographic items. Participants reported their gender, age, political orientation, and race.

Analytic Strategy

Implicit evaluations were investigated using a generalized linear mixed-effects model with binary response (pleasant vs. unpleasant) as the dependent variable. Explicit evaluations were investigated using a linear mixed-effects model with responses on the 7-point semantic differential items as the dependent variable. In both cases, random and fixed effects were added stepwise to the model, and incremental gains in model fit were investigated using a likelihood ratio test. The best fitting model was retained and interpreted in each case.

In Study 1, we conducted an exploratory analysis to probe whether the reinterpretation items mediated the effects of condition on implicit evaluations. Mediation models were fit using 10,000 nonparametric bootstrap simulations with bias-corrected and accelerated confidence intervals. Based on considerations of internal consistency, the mean of the first three reinterpretation items was used as the mediator variable.

Results

Evaluations of Knowlton are shown in Figures 1 and 2 (Study 1) and Figures 3 and 4 (Study 2). Explicit and implicit evaluations of Knowlton were markedly negative at Time 1 in both conditions of both studies. In Study 1, evaluations remained equally negative at Time 2 in the control condition but shifted toward positivity and reached neutrality in the experimental condition. In Study 2, initially negative explicit and implicit evaluations of Knowlton shifted toward positivity at Time 2 in both conditions, suggesting that updating unfolded robustly irrespective of the retrieval manipulation.

Explicit Evaluations

In Study 1, the best fitting model included random intercepts for participants and items, and a Condition × Time interaction as a fixed effect. In the control condition, the difference between Time-1 and Time-2 explicit evaluations was not significant, $t(11) = 0.04$, $p = .972$. By contrast, in the

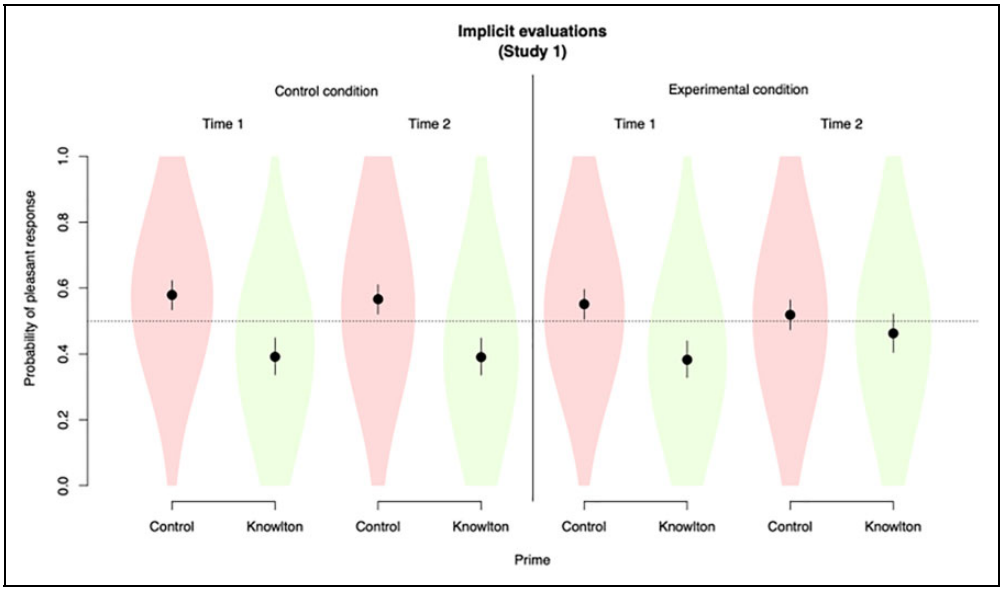

Figure 2. Distribution of implicit evaluations by condition, time of measurement, and prime (Study 1). The dashed horizontal line marks neutrality, and the solid dots show condition means. Error bars represent 95% confidence intervals derived from the mixed-effects model reported for the study. Positive scores indicate positive evaluations of the control primes or Knowlton, respectively.

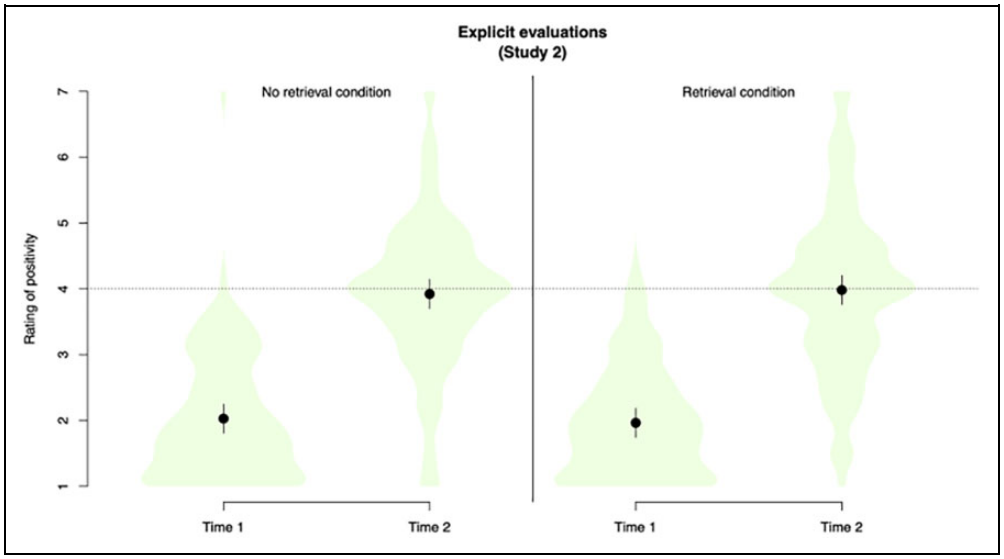

Figure 3. Distribution of explicit evaluations by condition and time of measurement (Study 2). The dashed horizontal line marks neutrality, and the solid dots show condition means. Error bars represent 95% confidence intervals derived from the full mixed-effects model containing a Condition × Time interaction. Positive scores indicate positive evaluations of Knowlton.

experimental condition, explicit evaluations shifted significantly in the positive direction, $t(11) = 12.60$, $p < .001$.

In Study 2, the best fitting model included random intercepts for participants and items, and a fixed effect for time but no main effect for condition and no interaction. In line with the results of Study 1, explicit evaluations shifted significantly in the positive direction from Time 1 to Time 2 across both retrieval conditions, $t(10) = 18.70$, $p < .001$.

Implicit Evaluations

In Study 1, the best fitting model included random intercepts for participants, target stimuli, and prime stimuli, and a Condition × Time × Prime type interaction as a fixed effect. At Time 1, we observed a strong implicit preference for control over Knowlton primes both in the control condition, $z = 6.63$, $p < .001$, and in the experimental condition, $z = 5.96$, $p < .001$.

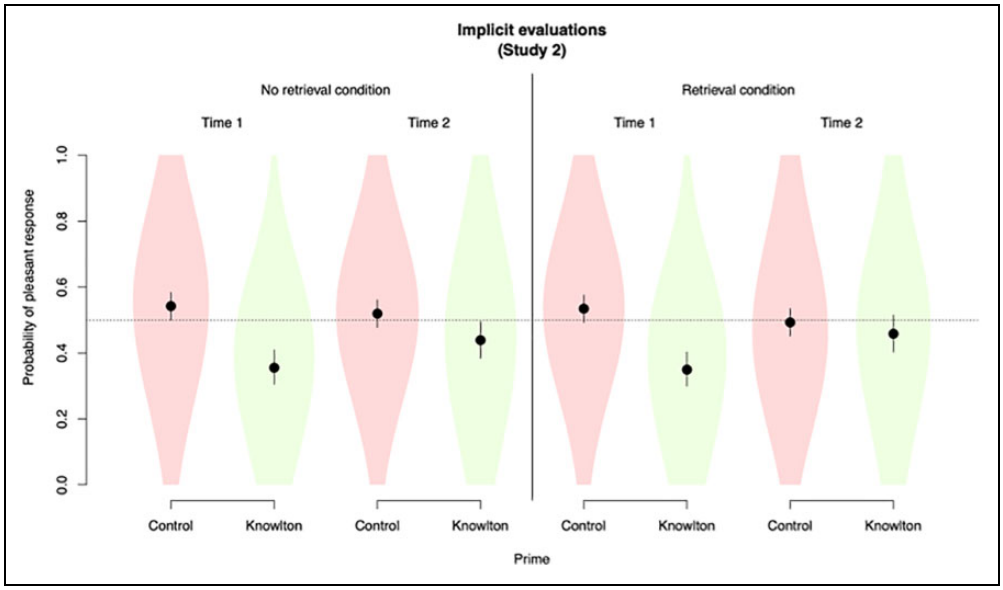

Figure 4. Distribution of implicit evaluations by condition, time of measurement, and prime (Study 2). The dashed horizontal line marks neutrality, and the solid dots show condition means. Error bars represent 95% confidence intervals derived from the full mixed-effects model containing a Condition × Time × Prime type interaction. Positive scores indicate positive evaluations of the control primes or Knowlton, respectively.

In the control condition, we found no evidence for any shift in the evaluation of control primes, $z = –0.80$, $p = .422$, or of Knowlton primes, $z = –0.06$, $p = .951$, from Time 1 to Time 2. By contrast, in the experimental condition, the evaluation of control primes may have shifted in the negative direction, $z = –1.96$, $p = .050$, and, crucially, the evaluation of Knowlton primes clearly shifted in the positive direction, $z = 4.90$, $p < .001$. Accordingly, at Time 2, control primes were still strongly preferred to Knowlton primes in the control condition, $z = 6.20$, $p < .001$, and to a considerably attenuated degree in the experimental condition, $z = 1.97$, $p = .049$.

In Study 2, the best fitting model included random intercepts for participants, target stimuli, and prime stimuli, and a Time × Prime type interaction as a fixed effect. Retrieval condition had no main effect and did not participate in any significant interactions. At Time 1, we observed a strong implicit preference for control over Knowlton primes, $z = 7.19$, $p < .001$; at Time 2, this preference was considerably attenuated, $z = 2.19$, $p = .029$, thus replicating the main finding from Study 1.

Mediation Via Reinterpretation

In Study 1, exploratory analyses suggested full mediation of implicit attitude change via reinterpretation. The total effect of experimental condition on implicit evaluations was significant, $\beta = 0.36$, 95% CI: [0.09, 0.63], $p = .007$. The average causal mediation effect was significant, $\beta = 0.54$, 95% CI: [0.11, 1.06], $p = .019$, whereas the average direct effect was not, $\beta = –0.19$, 95% CI: [–0.72, 0.28], $p = .447$.

In Study 2, given that condition had no effect on implicit or explicit evaluations, no mediation analyses involving the reinterpretation variable were conducted. However, across both conditions, updating of implicit evaluations was significantly associated with reinterpretation scores, $r = .15$, $t(206) = 2.23$, $p = .027$.

Discussion

In Studies 1–2, we found clear evidence for the revision of implicit evaluations via reinterpretation under a previously unexamined set of conditions. Specifically, the information that participants initially learned was in line with their preexisting negative views of a controversial subject. Moreover, the subsequently presented counterattitudinal material was ecologically realistic, contained information of mixed valence, and expressed ambivalence regarding the target's actions. In Study 2, this effect was additionally found to persist in the face of a manipulation that prompted participants to retrieve arguments in favor of their (overwhelmingly negative) initial evaluations of the focal issue. Finally, exploratory analyses suggested that the effects of counterattitudinal information on implicit evaluations were predicted by the extent to which participants were willing to reassess their preexisting attitudes toward trophy hunting.

Study 3

Given that Studies 1–2 indexed implicit evaluations using only one measure (the AMP), results may be specific to this measure and may not generalize to the more general construct of

implicit evaluation. In fact, previous studies suggest that the AMP may be especially responsive both to information prompting rapid revision of strong preexisting evaluations in general (e.g., Van Dessel et al., 2019) and to reinterpretation manipulations in particular (e.g., Mann & Ferguson, 2015). Such dissociations across implicit measures are unsurprising, given that they are each characterized by different operating conditions (e.g., De Houwer & Moors, 2011). Therefore, in the present study, we investigate whether the findings from Studies 1–2 replicate using a measure operating via response competition rather than the misattribution of affect: the implicit association test (IAT; Greenwald et al., 1998). In addition, this final experiment allows us to probe whether the findings from Studies 1–2 generalize to a more heterogeneous nonstudent sample recruited from different countries around the world. Finally, unlike Studies 1–2, Study 3 was administered fully online rather than in the lab.

Method

Participants and Design

Six hundred and two participants were recruited from the Project Implicit educational website (http://implicit.harvard.edu/). Participants who did not complete the IAT ($n = 4$) or responded excessively fast on more than 10% of IAT trials ($n = 3$) were excluded from analyses. The final sample size was 595. In deviation from the preregistered analysis plan, we did not exclude participants based on the manipulation check item given that results were identical with or without these participants retained.

The design of Study 3 was different from that of Studies 1–2. Specifically, participants were either assigned to a control condition ($n = 321$) in which they read only the initial negative information about Knowlton or an experimental condition ($n = 274$) in which they additionally also listened to the counterattitudinal podcast.

Materials

The materials were identical to the ones used in Studies 1–2, with the exception that five new facial images of Knowlton and five facial images of the four control targets each (see below) were obtained.

Procedure and Measures

Learning phase. The learning phase was similar to Studies 1–2, with a few exceptions. First, in addition to the vignette about Knowlton, participants also read a length-matched vignette about one of four control targets (Prince Andrew, Alec Baldwin, Bernie Madoff, and Martin Shkreli). The order of the two vignettes was counterbalanced. Participants in the control condition proceeded directly to the test phase, whereas participants in the experimental condition first listened to the experimental podcast from Studies 1–2.

Test phase. Implicit and explicit evaluations of Knowlton relative to the control target were measured in the same fixed order, followed by manipulation check and reinterpretation items in the experimental condition.

Implicit evaluations. Implicit evaluations were measured using a standard 5-block IAT. Category labels were "Corey Knowlton" and the full name of the control target and category stimuli were five facial images of each target. Attribute labels were "Good" and "Bad" and attribute stimuli were "good," "great," "fantastic," "pleasant," and "wonderful," as well as "awful," "bad," "horrible," "terrible," and "unpleasant." The order of the two combined blocks was counterbalanced. IAT scores were calculated in line with Greenwald et al. (2003), such that positive D scores indicate a preference for Knowlton over the control target.

Explicit evaluations. Explicit evaluations were measured by asking participants to what extent they agreed that each target was characterized by each attribute stimulus used on the IAT. Explicit evaluations were measured using 100-point sliding scales. The order of the two targets and the order of items within each target was randomized. To create an overall index of explicit evaluations, mean responses to negative items were first subtracted from mean responses to positive items within each target and then the difference score for the control target subtracted from the difference score for Knowlton. As such, similar to the IAT D score, higher scores on the explicit evaluation measure express a relative preference for Knowlton over the control target.

Manipulation check and reinterpretation items. Participants in the experimental condition completed the same manipulation check and reinterpretation items as in Studies 1–2. Agreement with the reinterpretation items was measured using 100-point sliding scales.

Results and Discussion

Explicit and implicit evaluations of Knowlton relative to the control target shifted in a positive direction from the control to the experimental condition, reflecting the effect of the counterattitudinal manipulation (see Figure 5). In the control condition, Knowlton was evaluated equivalently to the control targets both on the explicit measure, $\beta_0 = -0.01$, $t(586) = -0.23$, $p = .820$, and on the implicit measure, $\beta_0 = -0.04$, $t(593) = -0.73$, $p = .468$. A significant difference between both conditions emerged on both the explicit, $\beta_1 = 0.87$, $t(586) = 13.00$, $p < .001$, and on the implicit measure, $\beta_1 = 0.51$, $t(593) = 6.51$, $p < .001$. Inferentially identical results emerged in Bayesian mixed-effects models explicitly accounting for control target identity. Finally, similar to Studies 1–2, implicit evaluations in the experimental condition were significantly correlated with scores on the reinterpretation scale, $r = .27$, $t(268) = 4.54$, $p < .001$. Taken together, these results underscore the generalizability of the findings from

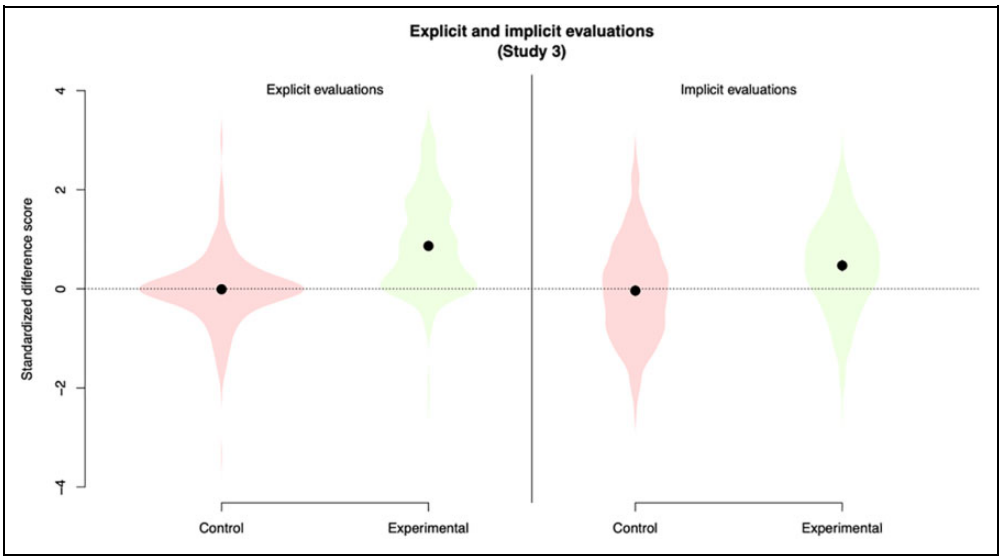

Figure 5. Distribution of explicit and implicit evaluations by condition (Study 3). The dashed horizontal line marks neutrality, and the solid dots show condition means. Error bars represent 95% confidence intervals derived from the relevant regression model. Positive scores indicate positive evaluations of Knowlton over the control target.

Studies 1–2 to a different implicit measure, a different setting, and a different, and more diverse, sample.

General Discussion

Across three studies, we found consistent evidence that implicit attitudes toward a target, subject to strongly unfavorable evaluations due to his involvement in a negatively evaluated activity (trophy hunting), can be substantially revised in response to an 8-min podcast presenting persuasive counterattitudinal information. As such, the present experiments corroborate existing results suggesting that implicit evaluations can be updated via reinterpretation of previously encountered evidence (Mann & Ferguson, 2015, 2017; Wyer, 2010).

Crucially, the present work demonstrates the robustness of reinterpretation as a modality of updating in several ways. Most importantly, unlike past work on implicit attitude change via reinterpretation, and most work on implicit attitude change more generally, the present studies used ecologically realistic materials that were created outside the experimental context. These materials contained ample negative information and even conveyed a measure of ambiguity on the focal issue. As such, the studies reported here found an effect of reinterpretation on implicit evaluations using an arguably more subtle manipulation than previous relevant work involving full reversal of the moral character attributed to a target from abhorrent to heroic. Moreover, unlike previous work involving fully novel targets performing unambiguously positive or unambiguously negative behaviors, the present work introduced a target in the context of a highly contentious real-world issue. Finally, Study 3 provided evidence for the generalizability of reinterpretation effects across implicit measures, samples, and study settings.

In addition, across all three studies, the revision of implicit evaluations was predicted by participants' self-reported willingness to revise their views on trophy hunting, thus providing evidence for the importance of propositional processes in implicit evaluation (Cone et al., 2017; De Houwer et al., 2020; Kurdi & Dunham, 2020). Moreover, in Study 2, a manipulation that prompted participants to retrieve and reflect on their preexisting attitudes had no impact on the revision of implicit evaluations. Although this result underscores the robustness of the main finding of this article, its theoretical interpretation is ambiguous. It is conceivable that the manipulation of memory retrieval was not sufficiently strong to influence subsequent updating. Alternatively, two countervailing processes, one inhibiting learning (inoculation; McGuire, 1961; McGuire & Papageorgis, 1961; Tannenbaum et al., 1966) and the other one conducive to learning (temporary memory instability upon retrieval; Schiller et al., 2010) may have canceled each other out. These competing explanations may be directly tested in the future.

The present results make contact with recent findings suggesting that implicit evaluations of even very well-known individuals can be subject to rapid revision. For example, in a study by Van Dessel et al. (2019), implicit evaluations of Gandhi shifted toward negativity when participants learned that he withheld lifesaving treatment from his wife but later accepted the same treatment for himself. Cone et al. (2019) created a full reversal of implicit attitudes toward Jack Black, a well-known and universally liked public figure, via doctored information about him being a perpetrator of domestic abuse. However, the current results differ from such previous work in two crucial ways. First, they demonstrate revision of initially negative implicit evaluations toward positivity, which is a direction of updating arguably more challenging to achieve given the dominance of negative information in impression formation

(e.g., Reeder & Brewer, 1979; Trafimow & Schneider, 1994). Second, in the present work, updating was produced via reinterpretation rather than via the introduction of entirely new information.

The latter distinction may be seen as particularly pertinent given differences in the memory processes assumed to play a role in updating via addition of new information versus reinterpretation of old information. Specifically, new information may be less effective in shifting implicit evaluations than reinterpretation of old information given persistence of the original memory trace with competing evaluative implications in the former, but not in the latter, case (Bouton, 1994; Gawronski et al., 2018; Gershman et al., 2013). In line with this idea, Mann and Ferguson (2015) have found that implicit evaluations of burglar-turned-savior Francis West were updated more strongly in response to reinterpretation of the previous evidence than in response to unrelated, but equally positive, information (but see Mann et al., 2020). In addition, initially negative implicit evaluations of Francis West were found to be amenable to updating via reinterpretation even 48 hrs after exposure to the original information (Mann & Ferguson, 2017). Whether these results, obtained using carefully crafted experimental materials, would generalize to more ecologically realistic settings, such as the one investigated here, await empirical testing.

Authors' Note

Benedek Kurdi is a member of the Scientific Advisory Board of Project Implicit, a 501(c)(3) nonprofit organization and international collaborative of researchers who are interested in implicit social cognition. Thomas C. Mann is now at Humana.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Benedek Kurdi https://orcid.org/0000-0001-5000-0584

References

Bouton, M. E. (1994). Context, ambiguity, and classical conditioning. *Current Directions in Psychological Science*, *3*(2), 49–53. http://doi.org/10.1111/1467-8721.ep10769943

Cone, J., Flaharty, K., & Ferguson, M. J. (2019). Believability of evidence matters for correcting social impressions. *Proceedings of the National Academy of Sciences*, *116*(20), 9802–9807. http://doi.org/10.1073/pnas.1903222116

Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. *Advances in Experimental Social Psychology*, *57*, 131–199. http://doi.org/10.1016/bs.aesp.2017.03.001

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

De Houwer, J., & Moors, A. (2011). Implicit measures: Similarities and differences. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 176–196). Guilford Press.

De Houwer, J., Van Dessel, P., & Moran, T. (2020). Attitudes beyond associations: On the role of propositional representations in stimulus evaluation. *Advances in Experimental Social Psychology*, *61*, 127–183. http://doi.org/10.1016/bs.aesp.2019.09.004

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. http://doi.org/10.1037//0022-3514.56.1.5

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*(2), 229–238. http://doi.org/10.1037/0022-3514.50.2.229

Gawronski, B., Rydell, R. J., De Houwer, J., Brannon, S. M., Ye, Y., Vervliet, B., & Hu, X. (2018). Contextualized attitude change. *Advances in Experimental Social Psychology*, *57*, 1–52. http://doi.org/10.1016/bs.aesp.2017.06.001

Gershman, S. J., Jones, C. E., Norman, K. A., Monfils, M.-H., & Niv, Y. (2013). Gradual extinction prevents the return of fear: Implications for the discovery of state. *Frontiers in Behavioral Neuroscience*, *7*. http://doi.org/10.3389/fnbeh.2013.00164

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4–27. http://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*(6), 1464–1480. https://doi.org/10.1037//0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037//0022-3514.85.2.197

Katz, J., Mann, T. C., Ferguson, M. J., Shen, X., & Goncalo, J. A. (2020). Implicit impressions of creative people. Manuscript submitted for publication. University of Illinois Urbana-Champaign.

Kellert, S. R. (1993). Attitudes, knowledge, and behavior toward wildlife among the industrial superpowers: United States, Japan, and Germany. *Journal of Social Issues*, *49*(1), 53–69. http://doi.org/10.1111/j.1540-4560.1993.tb00908.x

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, *63*(3), 190–194. https://doi.org/10.1093/analys/63.3.190

Kurdi, B., & Dunham, Y. (2020). Propositional accounts of implicit evaluation: Taking stock and looking ahead. *Social Cognition*, *38*(Supplement), s42–s67. http://doi.org/10.1521/soco.2020.38.supp.s42

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. http://doi.org/10.3758/s13428-014-0532-5

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*(6), 823–849. http://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*(C), 122–127. http://doi.org/10.1016/j.jesp.2016.06.004

Mann, T. C., Kurdi, B., & Banaji, M. R. (2020). How effectively can implicit evaluations be updated? Using evaluative statements after aversive repeated evaluative pairings. *Journal of Experimental Psychology: General*, *149*(6), 1169–1192. http://doi.org/10.1037/xge0000701

McGuire, W. J. (1961). Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterarguments. *Journal of Abnormal and Social Psychology*, *63*(2), 326–332. http://doi.org/10.1037/h0048344

McGuire, W. J., & Papageorgis, D. (1961). The relative efficacy of various types of prior belief-defense in producing immunity against persuasion. *Journal of Abnormal and Social Psychology*, *62*(2), 327–337. http://doi.org/10.1037/h0042026

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277–293. http://doi.org/10.1037/0022-3514.89.3.277

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, *86*(1), 61–79. http://doi.org/10.1037/0033-295X.86.1.61

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*(6), 995–1008. https://doi.org/10.1037/0022-3514.91.6.995

Schiller, D., Monfils, M.-H., Raio, C. M., Johnson, D. C., LeDoux, J. E., & Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, *463*(7277), 49–53. http://doi.org/10.1038/nature08637

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*(2), 108–131. https://doi.org/10.1207/s15327957pspr0402_01

Tannenbaum, P. H., Macauley, J. R., & Norris, E. L. (1966). Principle of congruity and reduction of persuasion. *Journal of Personality and Social Psychology*, *3*(2), 233–238. http://doi.org/10.1037/h0022893

Trafimow, D., & Schneider, D. J. (1994). The effects of behavioral, situational, and person information on different attribution judgments. *Journal of Experimental Social Psychology*, *30*(4), 351–369. http://doi.org/10.1006/jesp.1994.1017

Van Dessel, P., Ye, Y., & De Houwer, J. (2019). Changing deep-rooted implicit evaluation in the blink of an eye: Negative verbal information shifts automatic liking of Gandhi. *Social Psychological and Personality Science*, *10*(2), 266–273. http://doi.org/10.1177/1948550617752064

Wyer, N. A. (2010). You never get a second chance to make a first (implicit) impression: The role of elaboration in the formation and revision of implicit impressions. *Social Cognition*, *28*(1), 1–19. http://doi.org/10.1521/soco.2010.28.1.1

Wyer, N. A. (2016). Easier done than undone . . . by some of the people, some of the time: The role of elaboration in explicit and implicit group preferences. *Journal of Experimental Social Psychology*, *63*(C), 77–85. http://doi.org/10.1016/j.jesp.2015.12.006

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, *104*(20), 8235–8240. https://doi.org/10.1073/pnas.0701408104

Author Biographies

Benedek Kurdi is an experimental psychologist. His research examines learning and memory in the context of basic social processes, specifically evaluations of and beliefs about other people. He received his PhD at Harvard University and is currently a postdoctoral associate in the Department of Psychology at Yale University.

Thomas C. Mann received his PhD in social psychology at Cornell University. He was most recently a National Science Foundation postdoctoral research fellow at Harvard University and investigates implicit attitude change.

Melissa J. Ferguson is an experimental social psychologist. She received her doctorate in social psychology from New York University and is currently a professor of psychology at Yale University. Her research focuses on the implicit cognitive processes that enable evaluation, goal-pursuit, self-control, and social behavior.

Handling Editor: Eva Walther