A process dissociation model of implicit rapid revision in response to diagnostic revelations

Jeremy Cone* Jimmy Calanchini*

Williams College University of California Riverside

* These authors contributed equally to the research. Author order was determined by a coin flip.

Word Count: 9923 (not including figure captions)

Correspondence concerning this article should be addressed either to Jeremy Cone, Department of Psychology, Williams College, Williamstown, MA, 01267 or to Jimmy Calanchini, Department of Psychology, University of California Riverside, CA, 92521.

Contact: jdc2@williams.edu or jimmy.calanchini@ucr.edu

Abstract

Previous research has demonstrated that implicit evaluations can be reversed with exposure to a single impression-inconsistent behavior. But what exactly is changing when perceivers encounter diagnostic revelations about someone? One possibility is that rapid changes are occurring in the extent to which they view the person positively or negatively. Another possibility is that perceivers override the expression of initial evaluations through control-oriented processes. We conducted three studies (1 pre-registered) that utilized multinomial process trees to distinguish between these possibilities. We find consistent support across two different implicit measures that diagnostic behaviors result in rapid changes in evaluative processes. We obtained only inconsistent evidence for effects on more control-oriented processes. These findings thus help to reveal the cognitive processes underlying rapid implicit revision. Implications for theoretical perspectives on implicit attitudes are discussed.

Word count: 131

Key Words: implicit cognition, impression formation, multinomial process modeling

A process dissociation model of rapid implicit revision in response to diagnostic revelations

With the rise of the #metoo movement, many famous people have been revealed to have engaged in immoral behaviors, even some who were beloved by fans and had well-established, consistently positive reputations. When faced with these kinds of revelations that challenge our long-held beliefs, to what extent can we successfully update our impressions of someone? Though researchers generally agree that explicit impressions can be quickly and easily be updated, many contemporary theories argue that implicit impressions[1] may be relatively less easily updated when faced with impression-inconsistent information (e.g., Petty et al., 2006; Rydell & McConnell, 2006). Some evidence suggests that explicit evaluations can update very quickly when exposed to evidence that contradicts prior learning, but implicit evaluations update much more gradually, and only with repeated exposure to impression-inconsistent information (Rydell et al., 2007). Contemporary dual-process theories explain such dissociation by proposing that implicit evaluations rely upon associative mechanisms (e.g., Rydell & McConnell, 2006) that operate primarily on the basis of co-occurrence with positive or negative information (e.g., Gawronski & Bodenhausen, 2006, 2011).

Recent research, however, suggests that a single impression-inconsistent behavior can successfully overturn a well-rehearsed positive impression. In an initial demonstration (Cone & Ferguson, 2015), participants first learned about a stranger named Bob in a paradigm consisting of 100 instances of his behaviors indicating a consistently positive impression. Next, participants learned an additional behavior that was selected to be highly inconsistent with their previous

[1] In this manuscript, we use the terms *explicit* and *implicit* to refer to measurement conditions. Explicit measures assess participants' mental contents (e.g., evaluations) directly, e.g., through self-report. In contrast, implicit measures infer mental contents indirectly, based on the speed and/or accuracy of participants' responses. Thus, we use the term *implicit evaluations* to refer to evaluations that are measured indirectly, and we use the term *implicit revision* to refer to changes that occur in task performance on a measure of implicit evaluations (e.g., the Affect Misattribution Procedure; Payne et al., 2005; or, Implicit Association Test; Greenwald, McGhee, & Schwarz, 1998) in response to new evidence or some other intervention.

impression ("Bob recently mutilated a small, defenseless animal"). This single instance of extremely negative behavior caused a complete reversal of participants' implicit evaluations of Bob, shifting from significantly positive to significantly negative.

Since this initial demonstration, subsequent research has consistently replicated the basic effect of *rapid implicit revision* (Brannon & Gawronski, 2017; Cone, Flaharty, & Ferguson, 2019a, 2019b; Fourakis, Heggeseth, & Cone, 2019; Mann, Cone, Heggeseth, & Ferguson, 2019; for reviews, see Cone, Mann, & Ferguson, 2017; Ferguson, Mann, Cone, & Shen, 2019). This phenomenon generalizes across different implicit measures (Cone, Flaharty, & Ferguson, 2019a; Mann, Cone, Heggeseth, & Ferguson, 2019; Van Dessel et al., 2018), and occurs for both novel targets and for highly familiar, well-liked people (Cone, Flaharty & Ferguson, 2019a; Van Dessel et al., 2018).

With these considerations in mind, evidence of rapid implicit revision poses a challenge for theoretical perspectives that assume implicit attitudes reflect evaluative associations formed through repeated experience over time and, thus, are slow to form and slow to change (e.g., Cunningham, Zelazo, Packer, & van Bavel, 2007; Smith & DeCoster, 2000; Wilson, Lindsay, & Schooler, 2000). However, a possible counterargument is that rapid implicit revision does not reflect changes in evaluative associations, but instead reflects changes in other cognitive processes. A growing body of research utilizing multinomial processing trees (MPT: Batchelder & Riefer, 1999) indicates that multiple cognitive processes, in addition to evaluative associations, jointly contribute to responses on implicit measures (e.g., Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Meissner & Rothermund, 2013; Payne, 2001; Payne, Hall, Cameron, & Bishara, 2010; Stahl & Degner, 2007). Consequently, the purpose of the

present research is to use MPT modeling to examine the extent to which rapid implicit revision reflects changes in evaluative associations versus other cognitive processes.

Multinomial process tree modeling and implicit attitude change

Whereas implicit attitude measures were initially assumed to primarily reflect evaluative associations stored in memory (e.g., Fazio et al., 1995), MPT modeling is based on the assumption that multiple cognitive processes jointly contribute to responses on implicit measures. MPTs begin with a set of parameters that represent the hypothesized cognitive processes, and specify a system of equations that establish relationships among the parameters. Entering participants' actual responses as outcomes in the equations yields estimates of the extent of the processes hypothesized to produce those outcomes.

The MPTs most often used in social cognitive research are largely rooted in the dual-process tradition, which divides cognitive processes into two broad, qualitatively-distinct categories (e.g., Kahneman, 2003; Mischel, 1974; Shiffrin & Schneider, 1977). One category of cognitive processes is relatively fast, impulsive, and intuitive (e.g., "automatic", "hot", "system 1"), whereas the other category is relatively slow, deliberative, and intentional (e.g., "controlled", "cold", "system 2"). In the context of implicit social cognition, evaluative associations are generally assumed to belong to the first category, and control-oriented processes such as inhibition are generally assumed to belong to the second category.

All of the prominent social cognitive MPTs account for at least one type of evaluative process and at least one type of control-oriented process (e.g., Conrey et al., 2005; Meissner & Rothermund, 2013; Payne, 2001; Payne et al., 2010; Stahl & Degner, 2007). Thus, process modeling is well-positioned to examine the extent to which rapid implicit revision reflects changes in evaluative versus other processes. Previous MPT research has identified three

different patterns underlying implicit attitude change. For example, only evaluative associations are changed following exposure to positive Black and negative White exemplars (e.g., Michael Jordan and Adolph Hitler; Gonsalkorale, Allen, Sherman, & Klauer, 2010). Alternately, only control-oriented processes are changed when Black people are depicted in positive versus negative contexts (e.g., family barbecue versus prison; Allen, Sherman, & Klauer, 2010). Finally, both evaluative and control-oriented processes are changed after counter-prejudicial training (Calanchini, Gonsalkorale, Sherman, & Klauer, 2013).

Building upon previous MPT research into the processes underlying implicit attitude change, the present research examines the processes underlying rapid implicit revision. Does exposure to highly diagnostic revelations about someone manifest in changes to evaluative associations, control-oriented processes, or both? To answer this question, we adopted an exploratory/confirmatory approach. We sought to assess the generalizability of our findings across multiple experimental manipulations and paradigms, and across two of the most prominent implicit measures: the implicit association test (IAT; Greenwald, McGhee, & Schwarz, 1998) and the affect misattribution procedure (AMP; Payne et al., 2005). The experiment design, sample size, exclusion criteria, and analysis plan for Experiment 3 were pre-registered prior to conducting analyses, and are available on the Open Science Framework (OSF) at: https://osf.io/j2fp9/?view_only=8ca9c9fd6ee441ca9e48aea55a996d63[2]. Experiments 2 and 3 were re-analyses of previously-conducted studies. We describe the full procedure for these experiments in the Methods sections. However, we only re-analyzed the implicit measures in these studies. Consistent with the recommendations of Simmons, Nelson, & Simonsohn (2012),

[2] We wrote a pre-registration document for Experiment 1 prior to data collection. However, due to an oversight, we neglected to submit the registration on the Open Science Framework. The document was not altered from its initial draft, and is time-stamped and available on the OSF project page. We thus describe our design, sample size, exclusion criteria and analysis plan for Experiment 1 as *a priori* rather than as *pre-registered*.

the documents on OSF indicate our sample size and data exclusion policy, as well as all manipulations and measures in each study.

Experiment 1

Method

Participants. We recruited a total of N = 400 Mechanical Turk workers to participate in this experiment. Three participants failed to complete all of the components of the experiment and were excluded from analyses. An additional 47 participants were excluded because they had excessively fast reaction times (< 300 ms)[3] on greater than 10% of trials. The final sample was N = 350 (M_{age} = 35.9; % male = 52.9).

Procedure. Participants first completed a learning paradigm in which they were asked to learn about a new individual named Bob by reading behavioral statements about him while an image of his face appeared on screen (see Kerpelman & Himmelfarb, 1971; Rydell et al., 2006, 2007). Bob was depicted by a single image, selected from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015). On each of the 50 trials of the paradigm, participants saw either a positive behavioral statement (N=25; "Bob buys all of his friends a drink at the bar") or a negative one (N=25; "Bob made fun of his friend at a party") and were asked to indicate whether they thought the statement was characteristic (*c* key) or uncharacteristic (*u* key) of Bob. After responding, participants received immediate feedback on a 100% positive reinforcement schedule such that all positive behaviors were characteristic of Bob and all negative behaviors

[3] Our *a priori* exclusion criteria indicated that we would exclude participants for whom > 10% of trials had reaction times less than 400 ms. We deviated from these exclusion criteria because they turned out to be (a) much too conservative, resulting in a great deal of data loss (total exclusions: N = 149 [38%] participants), and (b) were inconsistent with best practices for IAT analyses (Greenwald et al., 2003). Instead, we followed Greenwald et al. (2003)'s D_2 algorithm exclusion criteria for all reported analyses.

were uncharacteristic. The order of the behavioral statements was randomized, and the statements were identical to those used in past work (e.g., Cone, Flaharty, & Ferguson, 2019a).

Next, we exposed participants to a 51st piece of information about Bob. In the experimental condition the behavior was selected to be extreme and highly impression-inconsistent: "Bob was recently convicted of child molestation" (Cone & Ferguson, 2015, Study 1a). In contrast, in the control condition, the behavior was innocuous: "Bob recently bought a soda." Participants were told that this information was more recent than what they had learned earlier and were advised to pay attention to it because it could be inconsistent with their previous exposure.

Next, participants completed measures of their implicit and explicit evaluations. For the implicit measure, participants completed a standard 7-block implicit association test (IAT; Greenwald, McGhee, & Schwartz, 1998), in which participants were tasked with categorizing four different types of targets into the categories *Bob*, *Not Bob*, *Good* or *Bad*. The positive target words were *nice*, *beautiful*, *friendly*, *good*, and *pleasant*. The negative target words were *annoying*, *awful*, *bad*, *rotten*, and *irritating*. For *Bob* targets, participants saw the same image from the learning paradigm. For *Not Bob* targets, participants saw five target images selected from the Chicago Face Database matched on age, race, and physical attractiveness to the image of Bob.

There were two sets of critical blocks. In one set, participants categorized *Bob+Bad* with one key and *Not Bob+Good* with the other. In the other set, participants categorized *Bob+Good* with one key and *Not Bob+Bad* with the other. All participants saw *Bob+Good* in critical blocks 3 and 4, and *Bob+Bad* in critical blocks 6 and 7. If participants provided an incorrect response, a red X appeared until the correct key was pressed. The reaction times in these critical blocks were

scored using the D_2 algorithm specified by Greenwald and colleagues (2003) such that positive D-scores indicated greater positivity towards Bob relative to neutral targets and negative D-scores indicated greater positivity towards neutral targets relative to Bob.

To assess explicit evaluations of Bob, participants first indicated how likeable they considered him to be on a Likert scale anchored with *1-very unlikeable* to *7-very likeable*, as well as five additional 7-point Likert scale items with anchors *very bad—very good*, *very mean—very pleasant*, *very disagreeable—very agreeable*, *very uncaring—very caring*, and *very cruel—very kind*. (We report analyses for explicit evaluations for this and all subsequent studies in Supplementary Materials [SM].)

Finally, participants completed a short demographic questionnaire that included measures of age, gender, and their familiarity and previous experience with studies on Mechanical Turk.

Results

Implicit evaluations. The D-scores of participants who learned that Bob bought a soda (M = 0.53, SD = 0.52) were significantly higher than the D-scores of participants who learned that Bob was convicted of a serious crime (M = 0.16, SD = 0.56), $t(348) = 6.34$, $p < .001$, d = 0.68, $CI_{95\%}$[0.25, 0.48] (Figure 1). However, inconsistent with previous work on rapid revision (e.g., Cone & Ferguson, 2015), there was no evidence of an implicit reversal in the experimental condition; the D-scores of participants in both conditions were significantly above zero, reflecting positive implicit evaluations of Bob: control, $t(180) = 13.751$, $p < .001$, d = 1.02, $CI_{95\%}$[0.45, 0.60], and experimental, $t(168) = 3.801$, $p < .001$, d = 0.29, $CI_{95\%}$[0.08, 0.25].

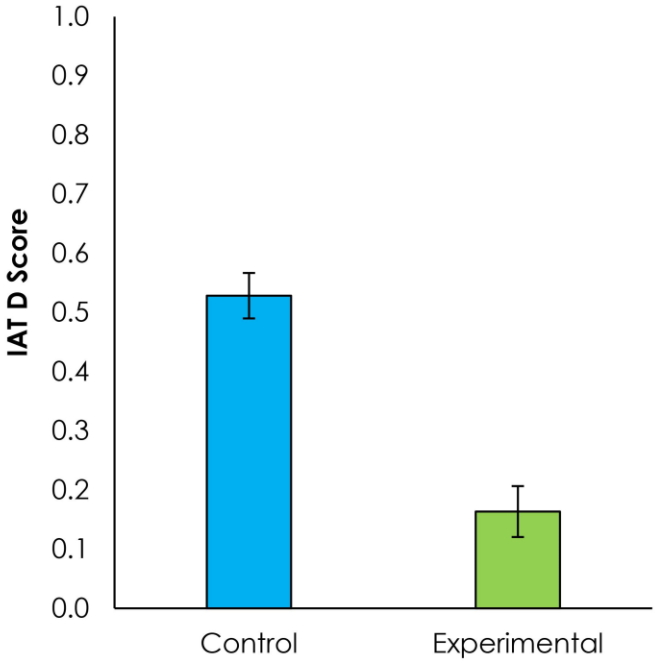


Figure 1. IAT D-scores in Study 1. Error bars represent standard errors.

Parameter estimation. The Quadruple process model (Conrey et al., 2005) is an MPT that specifies the contributions of four qualitatively distinct cognitive processes to responses on implicit measures. The Quad model has been applied to a variety of implicit measures, including sequential priming tasks (e.g., Allen et al., 2010) and the go/no-go association task (Gonsalkorale, von Hippel, Sherman, & Klauer, 2009), but has been most extensively applied to and validated on the IAT (for reviews, see Conrey et al., 2005; Sherman et al., 2008; Calanchini & Sherman, 2013).

The structure of the Quad model is depicted as a processing tree in Figure 2. The Activation parameter refers to the degree to which evaluative information (e.g., an association between a stimulus and an evaluation) is activated by a stimulus on an IAT trial. The more accessible the information (e.g., the stronger the association between *Bob* and *pleasant*), the more likely that information is to be activated and produce a response tendency in a direction consistent with the evaluation. The Detection parameter reflects the likelihood that the

participant can discriminate between correct and incorrect responses, according to task requirements. Sometimes, activated evaluative information conflicts with the detected correct response. For example, on an IAT trial in which a picture of Bob appears, and *Bob* and *unpleasant* stimuli share a response key (i.e., a so-called *incompatible* trial), the response tendency activated by Bob (i.e., to press the button labeled *pleasant*) conflicts with the detected correct response (i.e., to press the button labeled *unpleasant*). In this case, the Quad model proposes that an inhibitory process resolves the conflict: Overcoming Bias. This parameter refers to a process that prevents activated evaluative information from influencing behavior when this information conflicts with correct responses. Finally, the Guessing parameter reflects any other processes that guide responses in the absence of influence from the other three parameters.

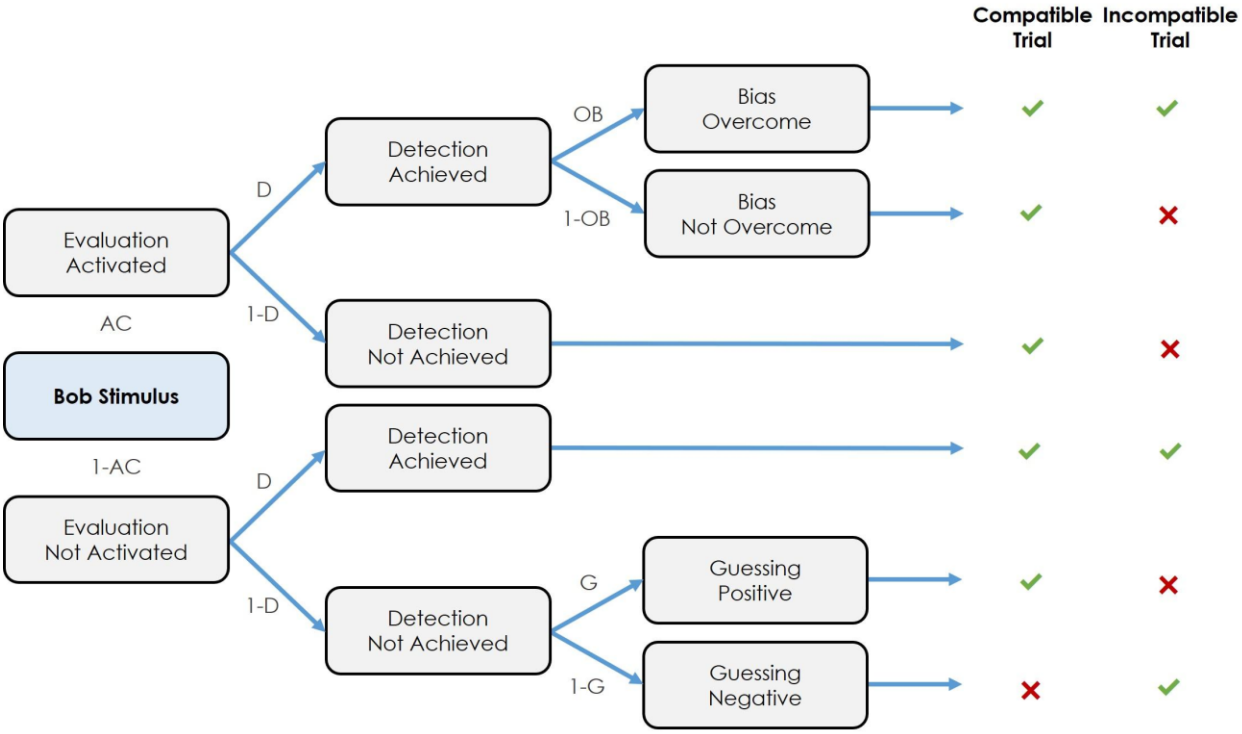


Figure 2. A portion of the Quad model. Each path represents a likelihood. Parameters with lines leading to them are conditional upon all preceding parameters. The table on the right side of the figure depicts correct (✓) and incorrect (✖) responses as a function of process pattern.

To estimate the parameters specified in the Quad model, we employed the Bayesian approach proposed by Klauer (2006, 2010) to fit a multilevel extension of the model that treats participants and items as random factors for each model parameter (Judd, Westfall, & Kenny, 2012), as implemented by the *TreeBUGS* R package (Heck, Arnold, & Arnold, 2018). In this Bayesian approach, the T_1 statistic summarizes how well the model accounts for the pattern of observed response frequencies (i.e., correct and incorrect responses to *Bob*, *Not Bob*, *pleasant*, and *unpleasant* stimuli on compatible and incompatible trials) aggregated across participants within each condition (Klauer, 2010). This statistic corresponds to the goodness-of-fit chi-square statistic used in traditional modeling approaches (Batchelder & Riefer, 1999). The T_2 statistic summarizes how well the model accounts for the variances and correlations of these frequencies computed across participants, which quantifies how well the model accounts for individual differences between participants in the individual response frequencies (Klauer, 2010).

For each participant, we estimated two Activation parameters, two Detection parameters[4], and one parameter each for Overcoming Bias and Guessing. One Activation parameter reflected the extent to which positive information is activated in response to *Bob*, and the other Activation parameter reflected the extent to which negative information is activated in response to the other, *Not Bob* target stimuli. One Detection parameter reflected accurate responding to attribute (i.e., *good*, *bad*) stimuli, and the other Detection parameter reflected accurate responding to target (i.e., *Bob*, *Not Bob*) stimuli. The Guessing parameter was coded so that values > .5 represent a

[4] The Quad model is typically specified to include one Detection parameter (e.g., Conrey et al., 2005). However, the Quad model has sometimes been specified to include two Detection parameters (e.g., Gonsalkorale, von Hippel, Sherman, & Klauer, 2009). In the present research, the two-Detection specification of the Quad model provides a better fit to these data (*AIC* = 18857.201, *BIC* = 18961.014, *FIA* = 9460.351) than does the one-Detection specification (*AIC* = 18971.232, *BIC* = 19057.744, *FIA* = 9512.396).

bias toward responding with the *good* key, values < .5 represent a bias towards responding with the *bad* key, and values = .5 represent no bias.

Across conditions, participants made 6.07% errors. At the individual level, the median *p*-value for T_1 was $p = .465$. At the group level, the observed versus predicted values for T_1 were 0.084 and 0.022, respectively, $p = .001$, and the observed versus predicted values for T_2 were 3.195 and 1.598, respectively, $p = .03$. The non-significant *p*-value for the individual-level statistic suggests that the Quad model provides good fit to these data, but the significant *p*-values for the group-level statistics suggest that the observed outcomes differed significantly from the predicted outcomes. Relative to the group-level tests, the individual-level test necessarily has less power to detect misfit. However, the group-level tests are based on an effective sample size of 350 participants x 120 IAT responses per participants = 42,000 observations, so the group-level tests are highly powered to detect even a small degree of misfit. Thus, the significant *p*-values indicate misfit but provide no insight into its degree. There is no agreed-upon method to quantify degree of model fit for the analyses used here in a way that controls for sample size. Consequently, we include graphs of the observed versus predicted frequencies and covariances for all experiments in SM. Visual inspection of these graphs indicates that differences between observed and predicted outcomes are minimal, which suggests that the Quad model provides good fit to these data.

Planned contrasts. We conducted a series of planned contrasts to examine the process-level effects underlying rapid implicit revision on each Quad parameter. We did so by subtracting the distributions for all posterior samples of a given parameter for the experimental condition from the distributions for all posterior samples of the same parameter for the control condition. In the resulting distribution of mean differences, the effect of negative information can

be interpreted as credibly different from the effect of neutral information if the 95% Bayesean Confidence Interval (BCI) does not contain zero. Below we summarize only credible differences, and report all parameter estimates in SM.

Bob-pleasant Activation parameters were higher in the neutral information condition (M = .014, $BCI_{95\%}$ [.001, .061]) than in the negative information condition (M = .001, $BCI_{95\%}$ [.00006, .004]), mean difference = .013, $BCI_{95\%}$ [.0002, .060]. Similarly, *Not Bob-unpleasant* Activation parameters were higher in the neutral information condition (M = .048, $BCI_{95\%}$ [.029, .071]) than in the negative information condition (M = .027, $BCI_{95\%}$ [.015, .041]), mean difference = .022, $BCI_{95\%}$ [.003, .045]. Detection for attributes was also higher in the neutral information condition (M = .910, $BCI_{95\%}$ [.892, .927]) than in the negative information condition (M=.882, $BCI_{95\%}$ [.859, .902]), mean difference=.028, $BCI_{95\%}$ [.003, .055]. Estimates of Detection for targets, Overcoming Bias, and Guessing did not differ between neutral and negative information conditions (Figure 3).

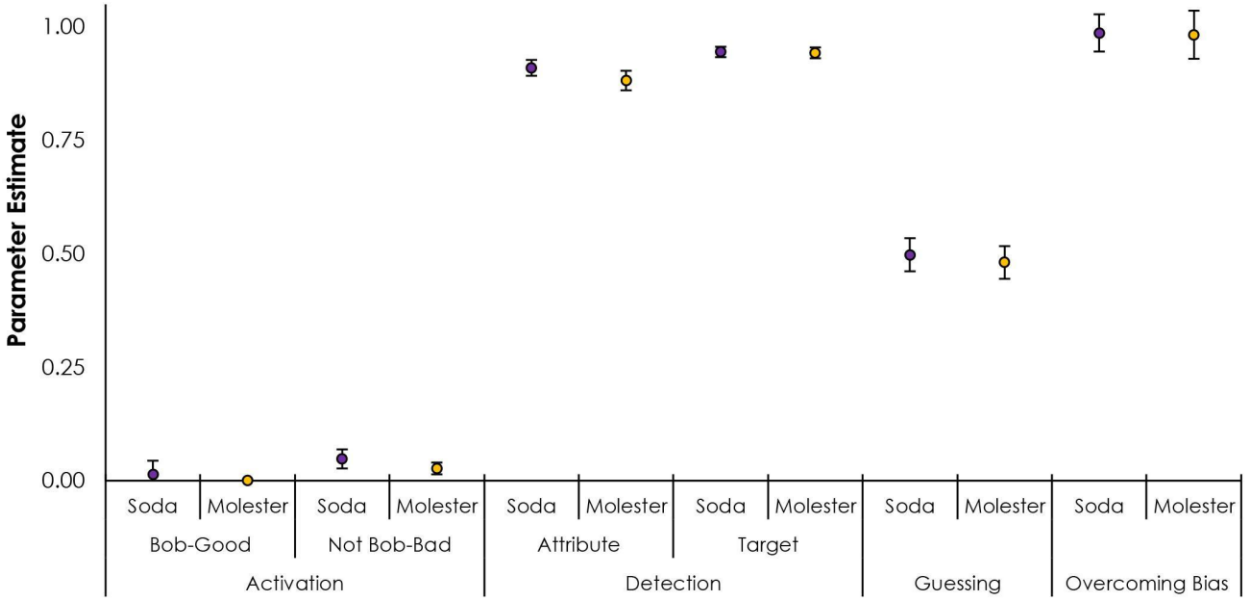


Figure 3. Quad model parameter estimates for Experiment 1. Error bars represent 95% BCIs.

Discussion

In Experiment 1, participants who learned highly diagnostic negative information about Bob had more negative implicit evaluations of him, as operationalized by the IAT D-score, than did participants who learned neutral information about him. More germane to the current investigation, Quad modeling revealed that the influence of diagnostic negative information was manifest on both of the Activation parameters, as well as on the Detection parameter. Thus, Experiment 1 provides initial evidence that rapid implicit revision changes both evaluative and control-oriented processes.

In the following experiments, we examine the extent to which the pattern of results observed in Experiment 1 generalizes to another implicit measure: the AMP (Payne et al., 2005). Task procedures necessarily determine which processes influence implicit responses (Bishara & Payne, 2009; Payne & Bishara, 2009; see also Ito et al., 2015), and the IAT and AMP differ in that the IAT depends on compatibility between stimuli and responses, whereas sequential priming-type tasks like the AMP depend on compatibility between target and prime stimuli (DeHouwer, 2003). Moreover, previous research suggests that implicit measures may diverge from one another in their sensitivity to deliberative processes (e.g., Gawronski & De Houwer, 2014). Indeed, research has sometimes shown divergence between the IAT and other implicit measures (e.g., Van Dessel et al., 2018). Thus, our goal in the next two studies was to explore whether the conclusions of the Quad model using an IAT generalized to another process model for the AMP.

Experiment 2

In Experiment 2, we re-analyzed an (exploratory) data set previously collected for Van Dessel, Cone, and Hughes (2019), in which participants first learned positive information about a stranger and later discovered a highly diagnostic revelation about him.

Method

Participants. The original study recruited a total of $N = 250$ participants on Prolific Academic (http://prolific.ac). Two participants failed to complete all components of the experiment and were excluded from analyses. An additional 2 participants were excluded because they self-reported speaking Chinese[5], and 21 participants were excluded because they pressed a single key on the implicit measure. One participant met both criteria. This left a final sample of N = 224 (M_{age} = 32.4, % male = 41.1%).

Procedure. The procedure was identical to Experiment 1 except for the following differences. The learning paradigm included 30 behaviors (15 *positive-characteristic* and 15 *negative-uncharacteristic*). To assess implicit evaluations of the target, participants completed an AMP (Payne et al., 2005). On each of the 72 trials of the measure, participants saw the following sequence of events: (a) a prime (an image of Bob, 36 trials; or an image of an age- and race-matched neutral stranger, 36 trials; 75 ms), (b) an inter-stimulus interval (ISI; 125 ms), (c) a Chinese pictograph (100 ms), and (d) white noise, which remained on the screen until participants responded. The images of Bob and the matched controls were selected from the Chicago Face Database. The primes consisted of each of these target images presented in either full-color, black-and-white, or sepia-tone (12 trials each). Participants' task was to evaluate the

[5] Experiment 2 relies on data initially reported in Van Dessel, Cone, & Hughes (2019). This experiment was split across two time points. However, the demographics survey containing the self-report items that assessed whether participants spoke Chinese occurred in the second session, so we could only exclude participants on the basis of this criterion who completed both time points. There were 68 participants who completed the first time point but who did not go on to complete the follow-up. These 68 participants were included in all analyses.

Chinese pictograph in terms of whether they considered them more pleasant than average (*d* key) or less pleasant than average (*k* key). Following prior procedures (Payne et al., 2005), participants were warned that the image presented before the pictograph could affect their evaluations and were asked to judge the pictograph while avoiding the influence of the prime. We calculated the proportion of times participants selected *more pleasant than average* for each prime as our measure of implicit evaluations of each target. Participants completed the AMP twice, immediately before (Time 1) and after (Time 2) learning the diagnostic information about Bob. This experiment did not include a control condition; instead, all participants were exposed to the same highly diagnostic behavior after completing the Time 1 AMP: "Bob recently mutilated a small, defenseless animal" (Cone & Ferguson, 2015, Study 1b).

Once they completed the Time 2 implicit and explicit evaluation measures, participants answered several exploratory items that assessed their perceptions of the diagnostic behavior and their memory for the information they learned. Finally, participants completed a demographics questionnaire that assessed their age, gender, race, education level, political orientation, and their familiarity with the procedures.

Results

Implicit revision. A 2 (Time: 1, 2) × 2 (Target: Bob, neutral) mixed ANOVA revealed main effects of both Time, $F(1, 223) = 17.710$, $p < .001$, $\eta_p^2 = 0.074$, and Target, $F(1, 223) = 8.674$, $p = .004$, $\eta_p^2 = 0.037$, qualified by the predicted two-way interaction, $F(1, 223) = 63.231$, $p < .001$, $\eta_p^2 = 0.221$ (Figure 4). At Time 1, participants exhibited an implicit preference for Bob ($M = 0.59$, $SD = 0.22$) relative to neutral strangers ($M = 0.52$, $SD = 0.23$), $t(223) = 3.504$, $p < .01$, $d = 0.23$, $CI_{95\%}[0.03, 0.12]$. At Time 2, however, they exhibited the opposite pattern, expressing

greater implicit negativity towards Bob (M = 0.42, SD = 0.27) relative to neutral strangers (M = 0.61, SD = 0.25), $t(223) = -6.732$, $p < .001$, $d = 0.45$, $CI_{95\%}$[-0.24, -0.13].

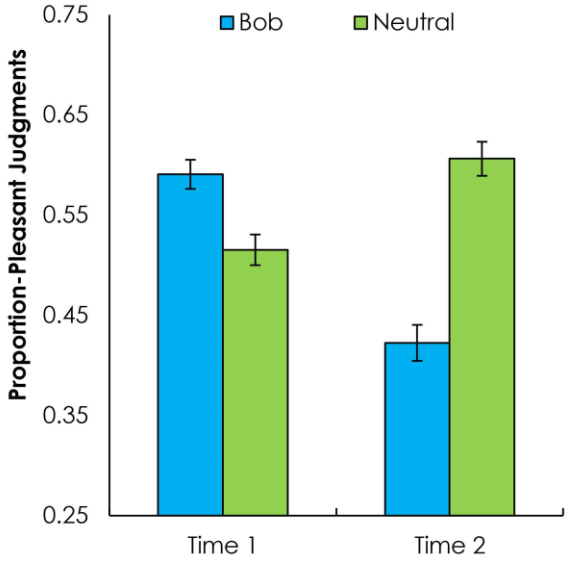


Figure 4. Implicit evaluations in Experiment 2. Error bars represent standard errors.

Parameter estimation. The AMP model (Payne et al., 2010) is an MPT that was developed and validated specifically for the AMP, though it has been applied to other sequential priming tasks (e.g., Rivers, Sherman, Rees, Reichardt, & Klauer, 2019). The AMP model specifies the contributions of three qualitatively distinct cognitive processes to responses on the AMP, and is depicted as a processing tree in Figure 5. The AMP model assumes that responses can be driven by two cognitive processes: evaluations of the target (i.e., Chinese pictograph) stimuli, which is reflected in the Pictograph parameter, or affective reactions to the prime (i.e., *Bob*, *Not Bob*) stimuli, which is reflected in the Affect parameter. The Misattribution parameter reflects the likelihood that responses are driven by reactions to the pictographs versus the primes. Because participants are instructed to judge the pictographs while avoiding the influence of the primes, the Misattribution parameter is conceptualized as a control-oriented process, in that misattribution reflects a failure to adhere to task instructions. When Misattribution occurs (i.e.,

control fails), affective reactions to the target stimuli drives responses; when Misattribution fails (i.e., control succeeds), evaluative reactions to the prime stimuli drives responses.

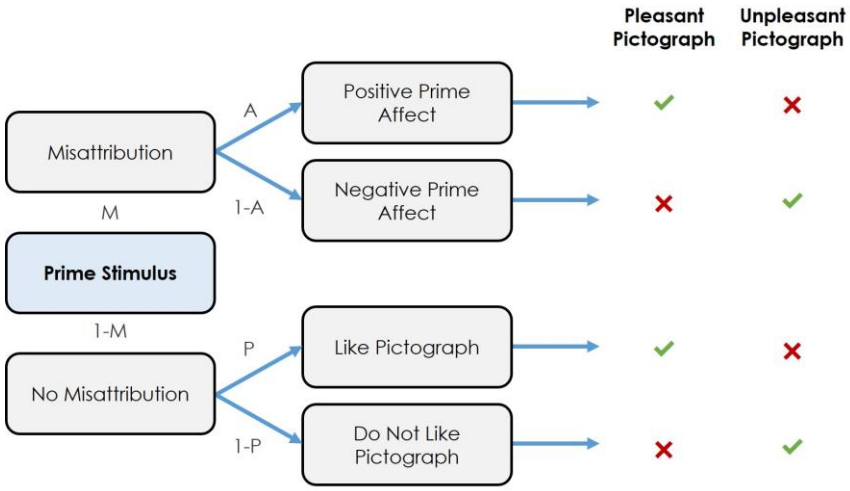


Figure 5. The AMP model. Each path represents a likelihood. Parameters with lines leading to them are conditional upon all preceding parameters. The table on the right side of the figure depicts normatively correct (✓) and incorrect (✖) responses as a function of process pattern.

To estimate the parameters specified in the AMP model, we employed the same Bayesian approach (Klauer, 2006, 2010) as in Experiment 1. For each participant we estimated two Affect parameters, two Pictograph parameters, and one Misattribution parameter for each measurement time. One Affect parameter reflected affective reactions to *Bob*, and the other Affect parameter reflected affective reactions to *Not Bob*. The Affect parameters are coded so that values greater than .5 represent positive evaluations, values less than .5 represent negative evaluations, and .5 represents a neutral evaluation. We relied on the pretest data reported in Payne et al. (2010, Footnote 1) to categorize each pictograph as either normatively pleasant or unpleasant. However, not all of the pictographs used in our version of the AMP were normed by Payne and colleagues. Thus, we excluded any trials for which we did not have normative ratings (Experiment 2: 20 trials excluded for each participant; Experiment 3: 14 trials excluded). One Pictograph parameter

reflected evaluations of the normatively pleasant pictographs, and the other Pictograph parameter reflected evaluations of the normatively unpleasant pictographs. The full AMP model includes more parameters than degrees of freedom, so we imposed constraints on the model based on Payne et al. (2010) to make the model identifiable. Specifically, we constrained the *Pictograph-pleasant* parameters to be equal across Time 1 and Time 2, as well as equal across participants. Similarly, we constrained the *Pictograph-unpleasant* parameters to be equal across Time 1 and time 2, as well as equal across participants.

Across measurement times, 46.53% of participants' responses conflicted with the normative responses of Payne and colleagues (2010). At the individual level, the median *p*-value for T_1 was $p = .454$. At the group level, the observed versus predicted values for T_1 were 0.160 and 0.026, respectively, $p < .001$, and the observed versus predicted values for T_2 were 17.420 and 4.500, respectively, $p < .001$. Visual inspection of graphs of the observed versus predicted frequencies and covariances (see SM) indicates that differences between observed and predicted outcomes are minimal, which suggests that the AMP model provides good fit to these data.

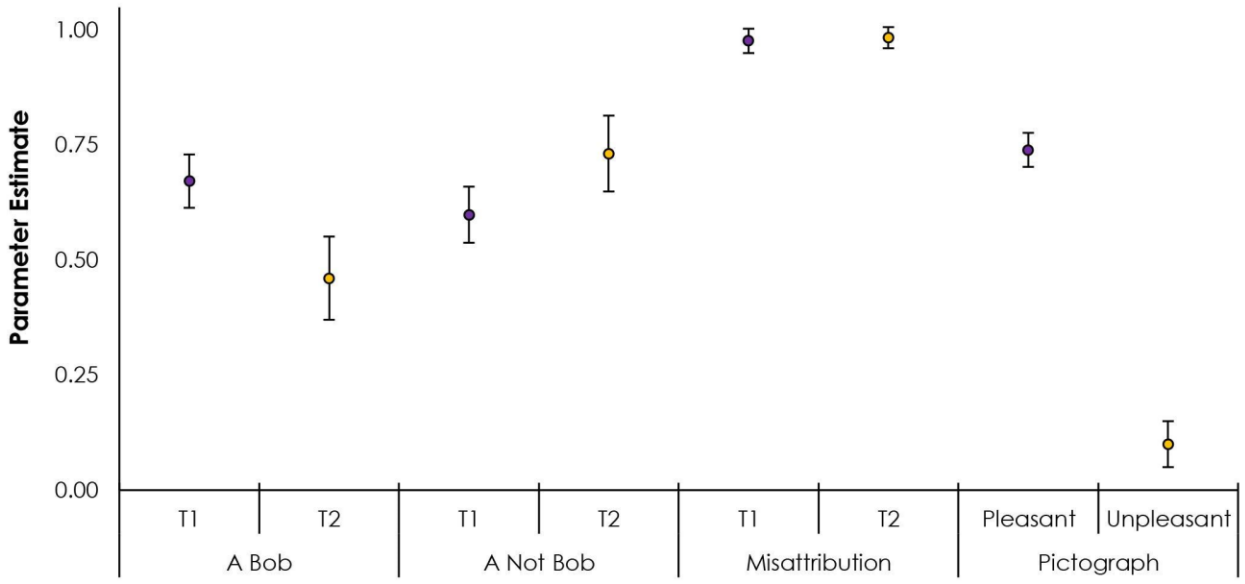


Figure 6. AMP model parameter estimates for Experiment 2. Error bars represent 95% BCIs.

Planned contrasts. The AMP parameters are plotted in Figure 6. We conducted a series of planned contrasts for each AMP parameter, using the same analytic method as in Experiment 1. Below we summarize only credible differences, and report all parameter estimates in SM.

Bob Affect parameters were higher at Time 1 (M = .672, $BCI_{95\%}$ [.613, .729]) than at Time 2 (M = .461, $BCI_{95\%}$ [.371, .552]), mean difference = .211, $BCI_{95\%}$ [.118, .300]. Thus, Affect towards Bob became credibly more unpleasant between time points.

Not Bob Affect parameters were lower at Time 1 (M = .599, $BCI_{95\%}$ [.540, .662]) than at Time 2 (M = .732, $BCI_{95\%}$ [.648, .813]), mean difference = -.132, $BCI_{95\%}$ [-.217, -.045]. Thus, participants' Affect towards Not Bob became credibly more pleasant between time points.

We tested the Time × Target interaction by comparing the effects of Time on the *Bob* Affect parameters to the effects of Time on the *Not Bob* Affect parameters. We found that there was a significant interaction, mean difference of differences = .343, $BCI_{95\%}$ [.197, .485].

Misattribution was not different between Time 1 (M = .977, $BCI_{95\%}$ [.943, .996) and Time 2 (M = .984, $BCI_{95\%}$ [.953, .999]), mean difference = -.001, $BCI_{95\%}$ [-.041, .025].

Discussion

In Experiment 2, participants had more positive implicit evaluations of Bob before versus after learning highly diagnostic negative information about him. AMP modeling revealed that the influence of diagnostic negative information was manifest exclusively on the Affect parameters. This finding conceptually replicates Experiment 1, to the extent that the Affect parameter of the AMP model and the Activation parameter of the Quad model both reflect evaluative reactions to Bob. However, AMP modeling in Experiment 2 found no influence of diagnostic information on

the Misattribution parameter. This finding diverges from Experiment 1, to the extent that the Misattribution parameter of the AMP model and the Detection parameter of the Quad model both reflect control-oriented processes (though there are important qualitative differences between Misattribution and Detection, a point we return to in the General Discussion.) Taken together, these findings suggest that the effects of rapid implicit revision on evaluative processes generalize across implicit measures, whereas the effects on control-oriented processes may be measure-specific. Given these inconsistencies, our primary goal in our final experiment was to replicate the findings of Experiment 2 in a confirmatory pre-registered design.

Experiment 3

In Experiment 3, we conducted a pre-registered conceptual replication of Experiment 2. Like Experiment 2, this experiment relied on previously-collected data (Cone, Flaharty, & Ferguson, 2019b, Study 1). The original experiment was split into two sessions that occurred 3 days apart from one another[6]. We only modeled data collected in the first session in which all participants first learned positive information about a stranger and later discovered a highly diagnostic revelation about him. We did not deviate from our pre-registered analysis plan.

Method

Participants. The original experiment recruited a total of 500 Prolific Academic participants. Following the original pre-registered exclusion criteria, we excluded (a) participants who did not complete all components of the experiment (N = 2), (b) participants who self-reported speaking Mandarin or Cantonese (N = 14), and (c) participants who pressed a single key

[6] The original study was pre-registered prior to data collection. Its original pre-registration document can be found here: https://osf.io/sg3vd/. The original study protocol did not deviate from the pre-registration. Note that, due to a transcription error, the pre-registration document for the current study mistakenly reports the sample size as N=499. However, there are N=498 valid participants; five hundred were recruited and 2 failed to complete all components of the study.

on the AMP (N = 25). This left a final sample of N = 459 participants (M_{age} = 29.4, % Male = 54.7).

Procedure. The procedure for Experiment 3 was similar to the procedure for Experiment 2 with the following differences. The target of the learning paradigm was named Kevin rather than Bob, and was identified by one of 6 counterbalanced images drawn from the Chicago Face Database. The learning paradigm included 50 behaviors (25 *positive-characteristic* and 25 *negative-uncharacteristic*).

The AMP consisted of 60 trials: 30 Kevin primes and 30 neutral stranger primes. Unlike the Experiment 2, only one full color image of Kevin was used as a prime, while the neutral primes consisted of 5 full-color images of unfamiliar targets. The explicit evaluation measure was identical to the previous studies.

The diagnostic information that participants learned between the Time 1 and Time 2 AMPs was similar to the previous experiments, except that participants were given more context about the information. Specifically, they were told that they should imagine that they learned the information from a friend of theirs and should assume it was true. The information they learned was that "Kevin was arrested a few years back for child molestation of his young niece." Importantly, unlike in Experiments 1 and 2, participants in Experiment 3 were *not* told that this information was more recent, nor were they warned that it could be inconsistent with their previous exposure. They were also not advised to update their impressions if the information was inconsistent, as was the case in the previous two experiments.

After the Time 2 implicit and explicit evaluation measures, participants completed single-item measures that assessed how believable, how diagnostic of his character, and how good or

bad they considered the information they learned about Kevin to be. Finally, participants completed a demographics questionnaire that was similar to the previous studies.

Results

Implicit revision. A 2 (Time: 1, 2) × 2 (Target: Kevin, neutral) mixed ANOVA revealed a main effect of Time, $F(1, 458) = 40.101$, $p < .001$, $\eta_p^2 = 0.081$, that was qualified by the predicted two-way interaction, $F(1, 458) = 140.209$, $p < .001$, $\eta_p^2 = 0.234$ (Figure 7). At Time 1, participants exhibited an implicit preference for Kevin ($M = 0.62$, $SD = 0.27$) relative to neutral strangers ($M = 0.47$, $SD = 0.22$), $t(458) = 9.287$, $p < .001$, $d = 0.43$, $CI_{95\%}[0.12, 0.18]$. At Time 2, however, they exhibited the opposite pattern, exhibiting greater implicit negativity towards Kevin ($M = 0.43$, $SD = 0.31$) relative to neutral strangers ($M = 0.56$, $SD = 0.25$), $t(458) = -6.859$, $p < .001$, $d = 0.32$, $CI_{95\%}[-0.17, -0.09]$[7].

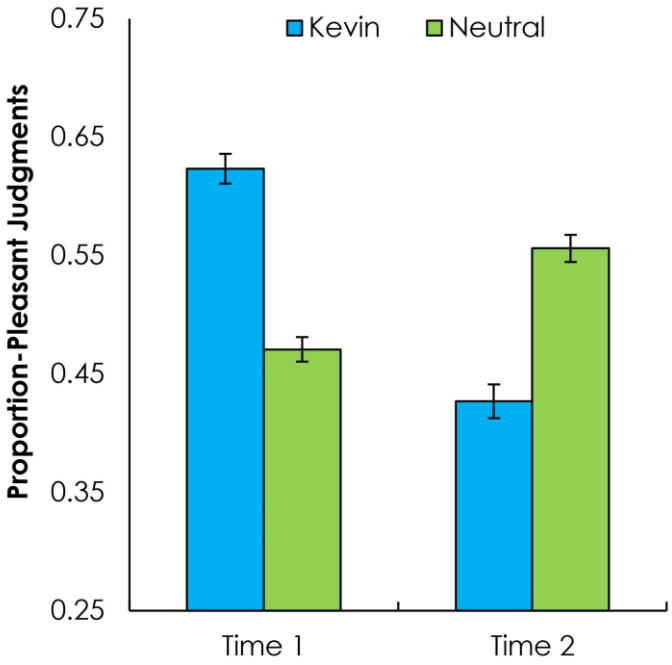


Figure 7. Implicit evaluations in Experiment 3. Error bars represent standard errors.

[7] There was an unpredicted marginal Target × Counterbalanced Image interaction, $F(1, 454) = 2.268$, $p = .061$, $\eta_p^2 = 0.020$. Otherwise, no effects of Counterbalanced Image emerged.

Parameter estimation. The AMP model was specified in the same way as Experiment 2. Across measurement times, 46.30% of participants' responses conflicted with the normative responses of Payne and colleagues (2010). At the individual level, the median *p*-value for T_1 was $p = .471$. At the group level, the observed versus predicted values for T_1 were 0.078 and 0.012, respectively, $p < .001$, and the observed versus predicted values for T_2 were 9.242 and 1.924, respectively, $p < .001$. Visual inspection of graphs of the observed versus predicted frequencies and covariances (see SM) indicates that differences between observed and predicted outcomes are minimal, which suggests that the AMP model provides good fit to these data.

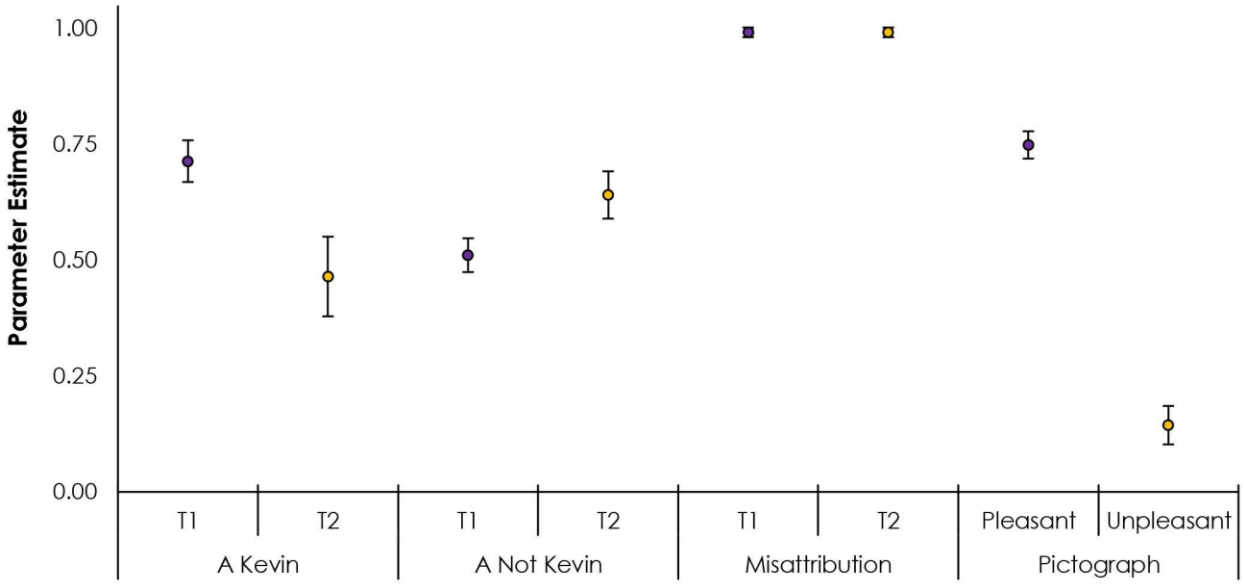


Figure 8. AMP model parameter estimates for Experiment 3. Error bars represent 95% BCIs.

Planned contrasts. The AMP parameters are plotted in Figure 8. We conducted a series of planned contrasts for each AMP parameter, using the same analytic method as Experiment 2. Below we summarize only credible differences, and report all parameter estimates in SM.

Kevin Affect parameters were higher at Time 1 (M = .714, $BCI_{95\%}$ [.669, .759]) than at Time 2 (M = .465, $BCI_{95\%}$ [.379, .551]), mean difference = .250, $BCI_{95\%}$ [.158, .337]. This pattern of results replicates the findings of Experiment 2, and indicates that Affect towards Kevin became credibly more unpleasant between measurement points.

Not Kevin Affect parameters were lower at Time 1 (M = .511, $BCI_{95\%}$ [.475, .548]) than at Time 2 (M = .641, $BCI_{95\%}$ [.589, .691]), mean difference = -.130, $BCI_{95\%}$ [-.177, -.081]. Thus, participants' Affect towards *Not Kevin* became credibly more pleasant between measurement points.

We tested the Time × Target interaction by comparing the effects of Time on the *Kevin* Affect parameters to the effects of Time on the *Not Kevin* Affect parameters. We found that there was a significant interaction, mean difference of differences = .380, $BCI_{95\%}$ [.270, .485].

Misattribution was not different between Time 1 (M = .992, $BCI_{95\%}$ [.978, .999]) and Time 2 (M = .992, $BCI_{95\%}$ [.978, .999]), mean difference = -.001, $BCI_{95\%}$ [-.015, .013].

Discussion

The results of Experiment 3 mirror those of Experiment 2: Participants had more positive implicit evaluations of Bob before versus after learning highly diagnostic negative information about him, and AMP modeling revealed that the influence of diagnostic information was manifest on the Affect parameters but not the Misattribution parameter. This pattern of results suggests that the effects of rapid implicit revision on evaluative processes reliably generalize across implicit measures, whereas the effects on control-oriented processes may be measure-specific.

General Discussion

When perceivers encounter new evidence about someone that is highly inconsistent with their impressions, previous research indicates that they exhibit rapid reversals of their implicit evaluations (Brannon & Gawronski, 2018; Cone & Ferguson, 2015; Cone et al., 2019a, 2019b; Fourakis & Cone, 2019; Mann et al., 2019; Van Dessel et al., 2018; for review, see Cone et al., 2017; Ferguson et al., 2019). The present research investigated the underlying cognitive processes that give rise to this rapid updating. One possibility is that rapid implicit revision could have left evaluative processes intact, and manifested only in changes to control-oriented processes that constrain the expression of evaluative processes (e.g., Allen et al., 2010). Another possibility is that rapid implicit revision could have reflected changes to both evaluative and control-oriented processes (e.g., Calanchini et al., 2013). However, a different pattern of results emerged in the present research: rapid implicit revision consistently manifested on parameters representing evaluative processes.

The findings reported here are notable for the extent to which they generalize across variations in experimental procedures and paradigms: sample population (MTurk versus Prolific Academic), implicit measure (IAT versus AMP), experiment design (within- versus between-participants), extent of exposure to positive information (50 versus 100 trials), images used to represent Bob (from past impression formation work versus the Chicago Face Database), and whether these images were counterbalanced versus fixed across participants. Taken together, the effects of rapid implicit revision on evaluative processes appear to be robust, and not an artifact of any particular feature of any specific experimental paradigm (e.g., Cole, Howard, & Maxwell, 1981).

The pattern of results observed in the present research complements recent, related work examining the process-level effects of exposure to evaluative information. Smith and colleagues

(Smith, Calanchini, Hughes, Van Dessel, & DeHouwer, 2019) had participants in one condition complete an evaluative learning task (e.g., approach-avoidance training) intended to create attitudes towards novel product brands, then measured participants' implicit attitudes towards the brands using an IAT. In a separate condition, participants received instructions describing the evaluative learning task (e.g., "When you see this brand, you will use your joystick to approach it.")—but never actually experienced the learning task—then completed an IAT. To the extent that implicit attitudes reflect associations that are formed through repeated experience, experienced evaluative learning should create stronger associations than instructed evaluative learning. However, Quad modeling revealed that both experienced and instructed evaluative learning influenced Activation parameters, and the effects of experienced versus instructed learning were indistinguishable at the process level.

When Smith and colleagues' (2019) findings are considered in conjunction with the present research, they are difficult to reconcile with traditional dual-process perspectives on implicit social cognition. If responses on implicit measures reflect associations formed through repeated experience and, thus, are slow to form and slow to change (e.g., Rydell & McConnell, 2006; Smith & DeCoster, 2000), then implicit attitudes should not form in the absence of experience, nor should they change in response to a single piece of information. But they can.

Alternatives explanations and alternative perspectives

An alternative interpretation of the present research is that the parameters conceptualized to represent evaluative processes in the Quad and AMP models do not, in fact, reflect evaluative associations. Drawing on traditional dual-process theory, previous Quad model research has typically described Activation as reflecting associations (e.g., Conrey et al., 2005), though sometimes Activation has been described with language that is relatively more agnostic about

representational structure (e.g., “the activation of an impulsive response tendency”; Sherman et al., 2008, p. 316). Previous AMP model research has been more consistent in conservatively describing Affect without reference to a specific representational structure (e.g., “positive affect towards the prime”; Payne et al., 2010, p. 1399). Though both the Quad and AMP models are validated to indicate that the Activation and Affect parameters reflect evaluative processes, none of this evidence favors one representational structure (e.g., associations) over another (e.g., propositions). Consequently, the fact that rapid implicit revision is manifest on the Activation and Affect parameters does not necessarily indicate that *associations* are changing. The representational structure of implicit attitudes is a matter of some debate, and our findings cannot necessarily adjudicate between these alternatives. However, they provide an important data point in this discussion by demonstrating the rapidity with which evaluative processes can be updated while simultaneously casting doubt on control-oriented processes as the (sole) mechanism underlying evaluative updating.

Traditional dual-process perspectives would be more easily reconciled with the current work if the effects of diagnostic information were manifest solely on control-oriented processes. Dual-process models largely agree that control-oriented processes depend, in part, on top-down processes such as motivation (e.g., Kahneman, 2003; Mischel, 1974; Shiffrin & Schneider, 1977), such that they can be updated on the fly to reflect current goals, intentions, or task demands. In contrast, the present research revealed only inconsistent (and perhaps measure-specific) effects of diagnostic information on control-oriented processes, but consistently revealed effects on evaluative processes. Thus, the learning characteristics of updates to the evaluative processes underlying implicit attitudes are much more rapid than many theoretical

perspectives currently grant, and these changes cannot be fully explained by rapid changes to processes that override previous learning.

Other theoretical perspectives are more easily reconciled with the current findings. For example, Amodio (2019) proposes that there is not a single, monolithic process or system that governs implicit cognition; rather, there are multiple, interacting implicit memory systems that each have their own properties and learning characteristics. For instance, semantic learning may be slow and incremental, such as when we learn to associate *doctor* with *nurse* through frequency of co-occurrence, but other types of implicit learning, such as fear learning, can happen very quickly and take a great deal more counter-learning to undo. On this view, diagnostic revelations may give rise to rapid revision through the operation of one or more fast-learning implicit processes. Alternatively, single-process propositional models (De Houwer, 2014) would propose that the effects observed here are not the result of associative learning, but rather a propositional process that exhibits fast-learning characteristics. Whereas associations are simple links between concepts, propositions are associations that imply a truth value and describe a state of affairs that may or may not be accurate or endorsed (e.g., Strack & Deutsch, 2004). With these perspectives in mind, the pattern of results observed in the present research could be interpreted to reflect changes in processes that are evaluative but may or may not be associative in the traditional dual-process sense. Successfully adjudicating between theoretical perspectives in the context of rapid revision in response to diagnostic revelations will be a task for future research.

Synthesizing the Quad and AMP models

In the present research, we observed consistent effects of diagnostic information on parameters conceptualized to reflect evaluative process in both the Quad and AMP models. This

pattern of results is noteworthy because these evaluative processes are operationalized in different ways in the two models. Evaluations in the AMP model (i.e., Affect) can only influence responses when control fails (i.e., Misattribution succeeds). In contrast, evaluations in the Quad model (i.e., Activation) compete with a control-oriented process (i.e., Detection) to influence responses, and this conflict is resolved by a different control-oriented process (i.e., Overcoming Bias). Thus, our findings suggest that the effects of diagnostic information are robust across nuances in the specification of evaluative processes.

As a complement to the parallel effects observed on evaluative processes, we observed inconsistent effects on control-oriented processes. Diagnostic information influenced one control-oriented process in the Quad model (i.e., Detection), but not the (only) control-oriented process in the AMP model (i.e., Misattribution). This outcome is perhaps unsurprising because Detection and Misattribution are conceptualized to reflect qualitatively different forms of control. Detection is an accuracy-oriented process that would always drive a correct response, but whether Detection drives responses depends on another control-oriented process, Overcoming Bias (and only when Activation and Detection would produce conflicting responses). In contrast, Misattribution determines the extent to which other processes can influence responses. In this way, Misattribution and Overcoming Bias are qualitatively similar processes. Neither Misattribution nor Overcoming Bias drive responses, per se, but instead determine which of two other processes drive responses: in the AMP model, Misattribution determines whether Affect versus Pleasantness drive responses, and in the Quad model Overcoming Bias determines whether Activation versus Detection drives responses. The fact that diagnostic information did not affect either of these parallel processes increases our confidence in interpreting these null effects. That said, we cannot rule out ceiling effects as an alternate

explanation for these findings: across all conditions in all three experiments, both Overcoming Bias and Misattribution parameter estimates were >.97 on a likelihood scale ranging [0,1]. Consequently, future research should continue to investigate the influence of diagnostic information on control-oriented inhibitory processes like Overcoming Bias and Misattribution. Nevertheless, there is no parallel to Detection in the AMP model, so the extent to which the effects of diagnostic information generalize to other accuracy-oriented processes remains an open question.

Detection is not the only process unique to one of the models used in the present research. The AMP and Quad models differ in other ways. For example, the Guessing parameter of the Quad model serves as a catch-all, conceptualized to reflect any process that guides responses besides Activation, Detection, and Overcoming Bias. Because the Quad model accounts for the influence of all other processes in this way, its other three parameters can be interpreted as reflecting relatively more pure estimates of their intended cognitive processes than can the parameters of the AMP model (though, to be clear, neither model claims that any of their parameters is a "pure" estimate of any cognitive process). The AMP model also includes a parameter that does not have a parallel in the Quad model. Whereas Affect in the AMP model and Activation in the Quad model both reflect evaluations of the construct of interest (e.g., attitudes towards Bob), the Pictograph parameter in the AMP model reflects evaluations unrelated to the construct of interest (i.e., Chinese pictographs). In this way, the AMP model distinguishes between different sources of affect, though it makes no claim that the affect activated by pictographs is qualitatively different from the affect activated by the primes. Consequently, it is not accurate to say that the Pictograph parameter reflects a cognitive process

not reflected in the Quad model, *per se*. Instead, construct-unrelated affect of this type would not seem to influence responses on the IAT.

One potential limitation of the modeling strategy employed in the present research is that it did not test whether the Pictograph parameters are affected by diagnostic information. Because the AMP model is saturated (i.e., containing more parameters than response categories), we imposed constraints in order for the estimation to proceed. In doing so, we specified the model to test hypotheses that most closely correspond to traditional dual-process social cognitive theory. The Pictograph parameters reflect a construct-unrelated evaluative process, so by constraining them to be equal across measurement times, we were able to test whether the effects of rapid implicit revision are manifest on a control-oriented process (i.e., Misattribution) or a construct-related evaluative processes (i.e., Affect). However, an alternate model specified to test for Pictograph effects (i.e., by constraining Affect parameters to be equal across measurement times) would seem to have merit: Pictograph in the AMP model and Detection in the Quad model are similar, in that both would always produce a correct response if allowed to drive responses (by Overcoming Bias and Misattribution, respectively). From this perspective, given that diagnostic information affected Detection, it seems worthwhile to examine whether diagnostic information also affected Pictograph. We report such analyses fully in the SM. In short, diagnostic information generally decreased Pictograph estimates on the AMP, which mirrors the effect of negative information on Detection on the IAT. However, this decreased influence of processes that would always produce correct responses would not seem to account for rapid implicit revision effects: that is, rather than rapid implicit revision reflecting the increased influence of these processes, implicit revision occurs despite their decreased influence. When considered in tandem with the analyses reported in the main text, highly diagnostic information would seem to

influence multiple cognitive processes, but rapid implicit revision as an outcome appears to be best explained in terms of changes to evaluative processes.

When multiple process models have been applied in previous research, the goal has usually been to determine which model provides better fit to data from a specific implicit measure (e.g., Bishara & Payne, 2009; Conrey et al., 2005; Meissner & Rothermund, 2015; Rivers et al., 2019). In the present research, we used an alternative approach, and applied different process models to different implicit measures in search of converging evidence. This approach helps to establish the reliability of effects (i.e., which ones emerge consistently) as well as the generalizability of effects (i.e., ones that emerge across different operationalizations of the same construct). This would seem to be a fruitful template for future research investigating the processes underlying social cognitive phenomena.

Inconsistent effects of revision versus reversal

Though we found consistent effects of diagnostic revelations on evaluative parameters in the present research, we found inconsistent effects on responses (e.g., *pleasant*, *unpleasant*) to the implicit measures themselves. Specifically, whereas perceivers exhibited a complete reversal of their implicit evaluations on the AMP (shifting from significantly positive to significantly negative over time)—a finding consistent with past work on the AMP (Brannon & Gawronski, 2018; Cone & Ferguson, 2015; Cone, Flaharty, & Ferguson, 2019a, 2019b; Mann et al., 2019) and the evaluative priming task (Cone, Flaharty, & Ferguson, 2019a)—their implicit evaluations on the IAT were attenuated, remaining significantly positive at Time 2 even after participants learned that Bob was responsible for an egregious crime. This pattern of results dovetails with other recent work that has found diverging effects on the IAT relative to other widely-used implicit measures (Van Dessel et al., 2018; see Gawronski & De Houwer, 2014, for a review).

Consequently, future research should continue to investigate the reliability of, and mechanisms underlying, such divergent response to diagnostic revelations.

Other routes to rapid implicit revision

The present work focused on understanding the process-level effects of diagnostic revelations, which complements previous research that has identified two additional routes to rapid revision (see Cone, Mann, & Ferguson, 2017): (a) reinterpretation, in which perceivers learn new details that cast previous learning in a different light and change its meaning, and (b) believability, in which perceivers' impressions exhibit greater updating to the extent that they see new impression-inconsistent information as more subjectively believable and credible. A task for future research will be to apply process modeling to these additional routes, which would shed light not only on the underlying cognitive processes that give rise to each of these alternative routes, but also on the extent to which each route is unique and empirically separable (see Cone, Mann, & Ferguson, 2017, for a more detailed discussion). Believability, for example, may be a route to rapid revision that could implicate greater controlled processing than the current work. If so, it would not only shed light on how perceivers respond to questionable evidence, but would also suggest it is distinct from diagnosticity as a mechanism for implicit evaluative reversals. Thus, process modeling can help to shed light not just on the underlying cognitive processes of rapid revision, but also on the relationship between different routes to evaluative change.

Conclusion

The present research indicates that the effects of a single piece of diagnostic information on implicit evaluations consistently manifest on evaluative processes. Our findings suggest that the evaluative processes that contribute to responses on implicit measures are capable of much more rapid change than previously suspected and that such changes are not (solely) occurring as

a result of control-oriented processes that inhibit previous learning. The present research thus provides additional evidence of *how* such changes are occurring, and helps to advance our understanding of the cognitive processes that contribute to implicit attitudes.

References

Allen, T. J., Sherman, J. W., & Klauer, K. C. (2010). Social context and the self-regulation of implicit bias. *Group Processes & Intergroup Relations, 13*(2), 137-149.

Amodio, D. M. (2018). Social Cognition 2.0: An interactive memory systems account. *Trends in cognitive sciences*.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review, 6*(1), 57-86.

Bishara, A. J., & Payne, B. K. (2009). Multinomial process tree models of control and automaticity in weapon misidentification. *Journal of Experimental Social Psychology, 45*(3), 524-534.

Brannon, S. M., & Gawronski, B. (2017). A second chance for first impressions? Exploring the context-(in) dependent updating of implicit evaluations. *Social Psychological and Personality Science, 8*(3), 275-283.

Calanchini, J., Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2013). Counter-prejudicial training reduces activation of biased associations and enhances response monitoring. *European Journal of Social Psychology*, *43*(5), 321-325.

Calanchini, J., & Sherman, J. W. (2013). Implicit attitudes reflect associative, non-associative, and non-attitudinal processes. *Social and Personality Psychology Compass, 7*(9), 654-667.

Cole, D. A., Howard, G. S., & Maxwell, S. E. (1981). Effects of mono-versus multiple-operationalization in construct validation efforts. *Journal of Consulting and Clinical Psychology*, *49*(3), 395-405.

Cone, J., & Ferguson, M. J. (2015). He did what? The role of diagnosticity in revising implicit evaluations. *Journal of personality and social psychology, 108*(1), 37.

Cone, J., Flaharty, K., & Ferguson, M. J. (2019a). Believability of evidence matters for correcting social impressions. *Proceedings of the National Academy of Sciences*, 201903222.

Cone, J., Flaharty, K., & Ferguson, M. J. (2019a). *The long-term effects of exposure to questionable information on social impressions.* Manuscript submitted for publication.

Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. In *Advances in Experimental Social Psychology* (Vol. 56, pp. 131-199): Elsevier.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of personality and social psychology, 83*(6), 1314.

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition, 25*(5), 736-760.

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

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of personality and social psychology, 69*(6), 1013.

Ferguson, M. J., Mann, T. C., Cone, J., & Shen, X. (2019). When and How Implicit First Impressions Can Be Updated. *Current Directions in Psychological Science*, 0963721419835206.

Fourakis, E., Heggeseth, B., & Cone, J. (2019). *Explaining Why: The role of causal attribution on impression impression formation and revision.* Manuscript submitted for publication.

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

Gawronski, B., & Bodenhausen, G. V. (2011). The associative–propositional evaluation model: Theory, evidence, and open questions. In *Advances in experimental social psychology* (Vol. 44, pp. 59-127): Elsevier.

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. *Handbook of research methods in social and personality psychology, 2*, 283-310.

Gonsalkorale, K., Allen, T. J., Sherman, J. W., & Klauer, K. C. (2010). Mechanisms of group membership and exemplar exposure effects on implicit attitudes. *Social Psychology*.

Gonsalkorale, K., von Hippel, W., Sherman, J. W., & Klauer, K. C. (2009). Bias and regulation of bias in intergroup interactions: Implicit attitudes toward Muslims and interaction quality. *Journal of Experimental Social Psychology*, *45*(1), 161-166.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, *74*(6), 1464.

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.

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior research methods*, *50*(1), 264-284.

Ito, T. A., Friedman, N. P., Bartholow, B. D., Correll, J., Loersch, C., Altamirano, L. J., & Miyake, A. (2015). Toward a comprehensive understanding of executive cognitive function in implicit racial bias. *Journal of personality and social psychology*, *108*(2), 187.

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*(1), 54.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, *58*(9), 697.

Kerpelman, J. P., & Himmelfarb, S. (1971). Partial reinforcement effects in attitude acquisition and counterconditioning. *Journal of Personality and Social Psychology*, *19*(3), 301.

Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika*, *71*(1), 7-31.

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*(1), 70-98.

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.

Mann, T., Cone, J., Heggeseth, B., & Ferguson, M. (2019). Updating implicit impressions: New evidence on intentionality and the Affect Misattribution Procedure. *Journal of personality and social psychology*.

McConnell, A. R., & Rydell, R. J. (2014). The systems of evaluation model. *Dual-process theories of the social mind*, 204-217.

Meissner, F., & Rothermund, K. (2013). Estimating the contributions of associations and recoding in the Implicit Association Test: The ReAL model for the IAT. *Journal of personality and social psychology*, *104*(1), 45.

Mischel, W. (1974). Processes in delay of gratification. In *Advances in experimental social psychology* (Vol. 7, pp. 249-292): Elsevier.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of personality and social psychology*, *81*(2), 181.

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.

Payne, B. K., Hall, D. L., Cameron, C. D., & Bishara, A. J. (2010). A process model of affect misattribution. *Personality and Social Psychology Bulletin*, *36*(10), 1397-1408.

Petty, R. E., Briñol, P., & DeMarree, K. G. (2007). The Meta–Cognitive Model (MCM) of attitudes: Implications for attitude measurement, change, and strength. *Social Cognition*, *25*(5), 657-686.

Rivers, A. M., Sherman, J. W., Rees, H. R., Reichardt, R., & Klauer, K. C. (2019). On the roles of stereotype activation and application in diminishing implicit bias. *Personality and Social Psychology Bulletin*, 0146167219853842.

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.

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*(5), 867-878.

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

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Available at SSRN 2160588.*

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.

Stahl, C., & Degner, J. (2007). Assessing automatic activation of valence: A multinomial model of EAST performance. *Experimental Psychology, 54*(2), 99-112.

Van Dessel, P., Cone, J., & Hughes, S. (2019). *Factors affecting the durability of implicit rapid revision.* Manuscript in progress.

Van Dessel, P., Ye, Y., & De Houwer, J. (2018). Changing Deep-Rooted Implicit Evaluation in the Blink of an Eye: Negative Verbal Information Shifts Automatic Liking of Gandhi. *Social Psychological and Personality Science*, 1948550617752064.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological review, 107*(1), 101.

A process dissociation model of implicit rapid revision in response to diagnostic revelations

Jeremy Cone* | Jimmy Calanchini*
Williams College | University of California Riverside

Supplementary Information

Explicit Evaluations

Experiment 1

Participants in the control condition had significantly more favorable explicit evaluations of Bob (M = 6.46, SD = 0.74) than those in the experimental condition (M = 2.94, SD = 1.7), $t(395) = 26.893$, $p < .001$, d = 2.68, 95% CI[3.26, 3.78].

Experiment 2

Participants were more favorable towards Bob at Time 1 before they learned the diagnostic information (M = 6.25, SD = 0.96) than at Time 2 after they learned that he had mutilated an animal (M = 2.79, SD = 1.55), $t(223) = 28.648$, $p < .001$, d = 1.91, 95% CI[3.22, 3.7].

Experiment 3

Participants were more favorable towards Kevin at Time 1 before they learned the diagnostic information (M = 6.31, SD = 0.85) than at Time 2 after they learned that he had been arrested for child molestation (M = 3.21, SD = 1.77), $t(458) = 34.343$, $p < .001$, d = 1.60, 95% CI[2.92, 3.28].

Summary Tables of Modeling Parameters Across All Studies

The following tables provide the estimates of all modeling parameters across all three studies.

	Soda condition		**Molester condition**	
	M	95% BCI	*M*	95% BCI
Activation				
Not Bob-Bad	0.048	[0.029, 0.071]	0.027	[0.015, 0.041]
Bob-Good	0.014	[0.001, 0.061]	0.001	[0.00006, 0.004]
Detection				
Attribute	0.91	[0.892, 0.927]	0.882	[0.859, 0.902]
Target	0.945	[0.933, 0.956]	0.943	[0.93, 0.954]
Guessing	0.498	[0.462, 0.535]	0.481	[0.445, 0.517]
Overcoming Bias	0.987	[0.918, 1]	0.983	[0.894, 1]

Table S1. Model parameter estimates in Study 1.

	Time 1		Time 2	
	M	95% BCI	*M*	95% BCI
Activation				
Bob	0.672	[0.613, 0.729]	0.461	[0.371, 0.552]
Not Bob	0.599	[0.54, 0.662]	0.732	[0.648, 0.813]
Misattribution	0.977	[0.943, 0.996]	0.984	[0.953, 0.999]
Pictograph				
Pleasant*	0.723	[0.687, 0.761]		
Unpleasant*	0.167	[0.116, 0.216]		

* *Note:* Each of these parameters is fixed to be equal across measurement times as well as equal across participants.

Table S2. Model parameter estimates in Study 2.

	Time 1		Time 2	
	M	95% BCI	*M*	95% BCI
Activation				
Kevin	0.714	[0.669, 0.759]	0.465	[0.379, 0.551]
Not Kevin	0.511	[0.475, 0.548]	0.641	[0.589, 0.691]
Misattribution	0.992	[0.978, 0.999]	0.992	[0.978, 0.999]
Pictograph				
Pleasant*	0.749	[0.721, 0.78]		
Unpleasant*	0.144	[0.103, 0.186]		

* *Note:* Each of these parameters is fixed to be equal across measurement times as well as equal across participants.

Table S3. Model parameter estimates in Study 3.

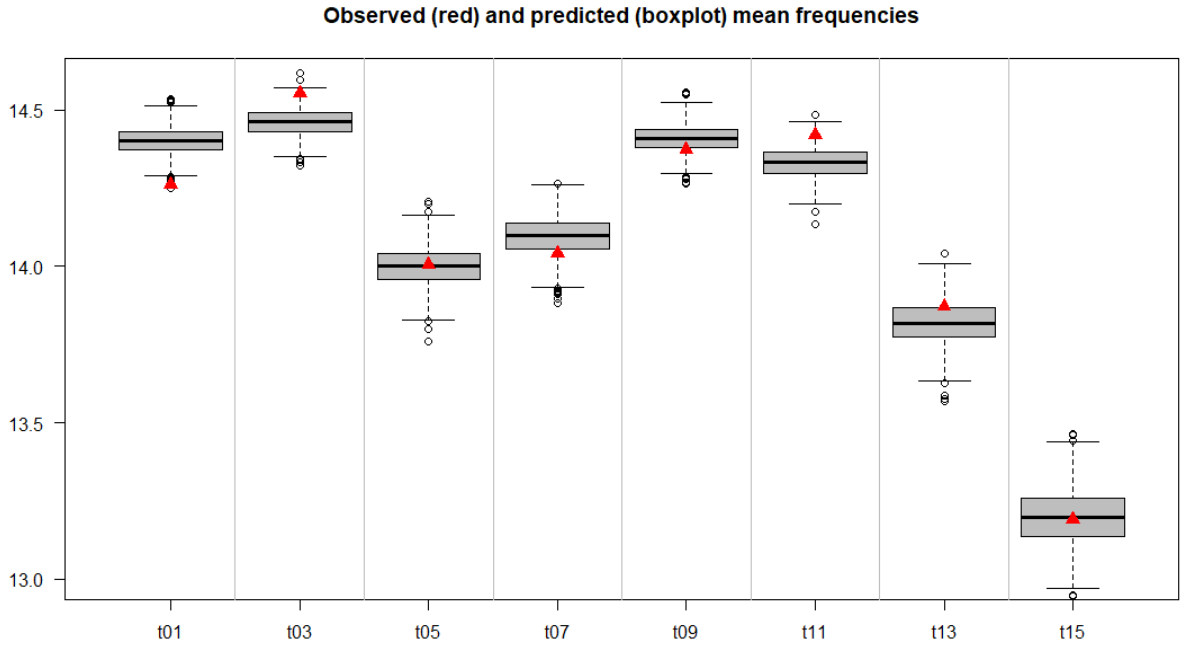
Graphs of observed versus expected means and covariances

Experiment 1

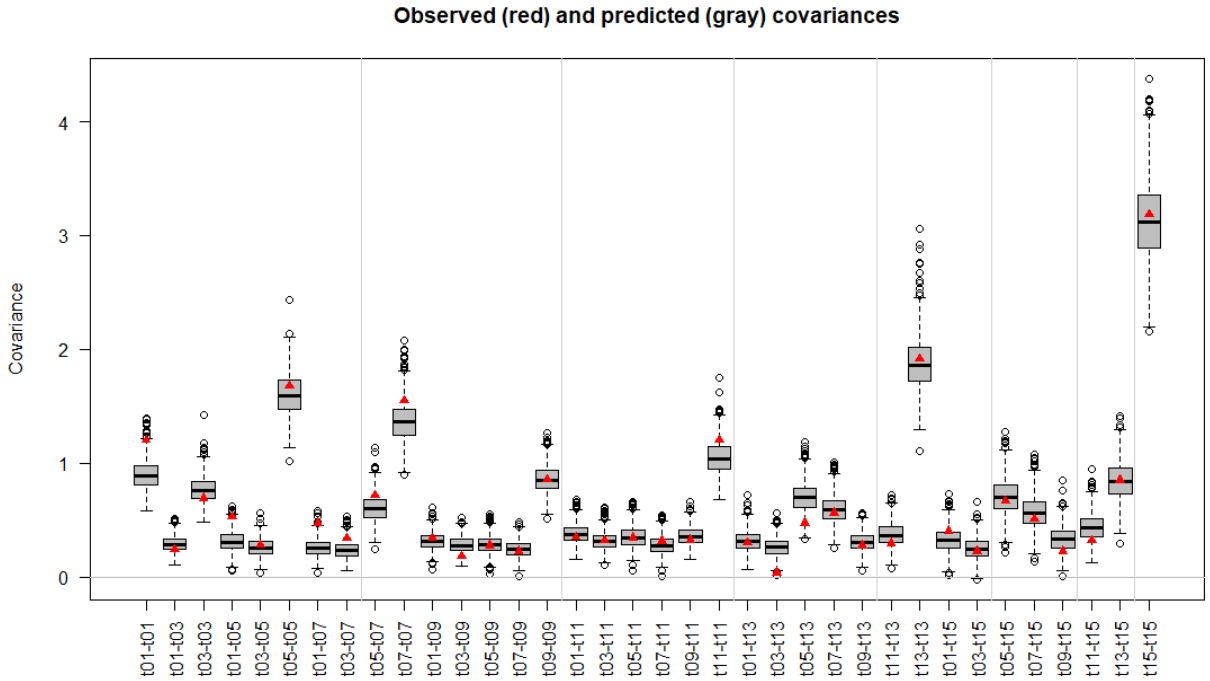
Legend:

t01: correct Bob responses when Bob / pleasant share response key
t03: correct Not Bob responses when Not Bob / unpleasant share response key
t05: correct pleasant responses when Bob / pleasant share response key
t07: correct unpleasant responses when Not Bob / unpleasant share response key
t09: correct Bob responses when Bob / unpleasant share response key
t11: correct Not Bob responses when Not Bob / pleasant share response key
t13: correct pleasant responses when Bob / unpleasant share response key
t15: correct unpleasant responses when Not Bob / pleasant share response key

T_1 graph:



T_2 graph:

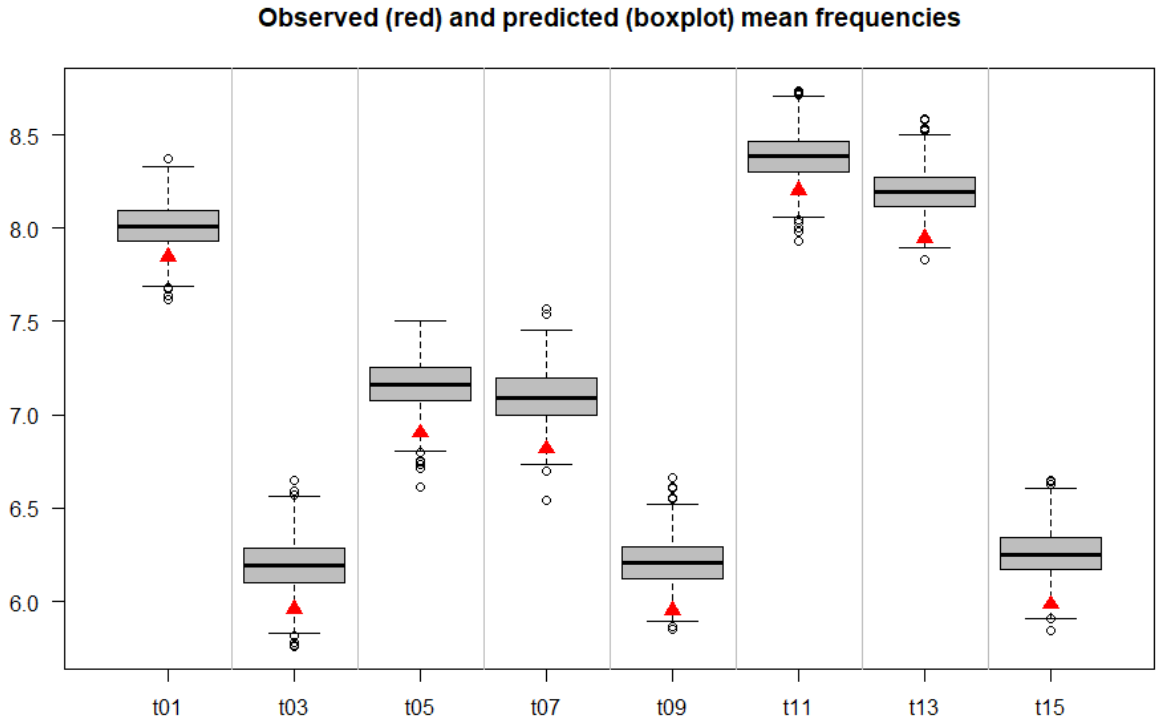


Experiment 2

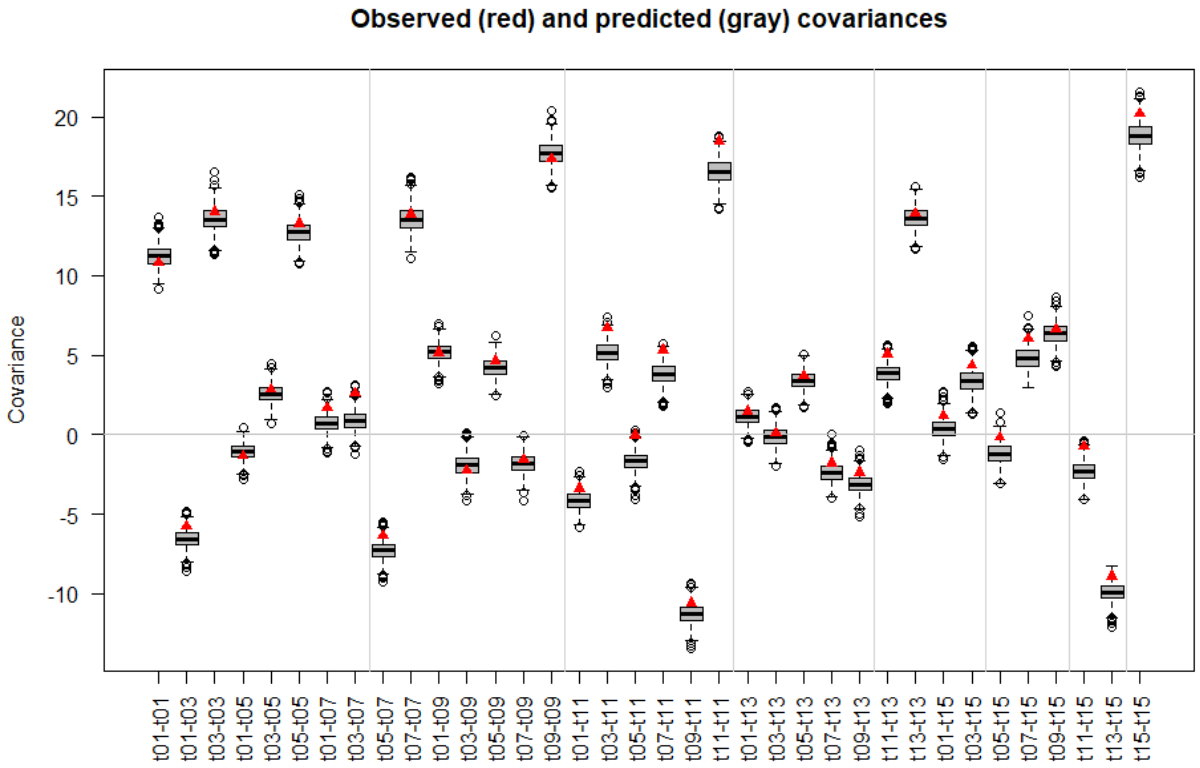
Legend:

t01: correct responses to pleasant pictographs following Bob primes at Time 1
t03: correct responses to unpleasant pictographs following Bob primes at Time 1
t05: correct responses to pleasant pictographs following Not Bob primes at Time 1
t07: correct responses to unpleasant pictographs following Not Bob primes at Time 1
t09: correct responses to pleasant pictographs following Bob primes at Time 2
t11: correct responses to unpleasant pictographs following Bob primes at Time 2
t13: correct responses to pleasant pictographs following Not Bob primes at Time 2
t15: correct responses to unpleasant pictographs following Not Bob primes at Time 2

T_1 graph:



T_2 graph:

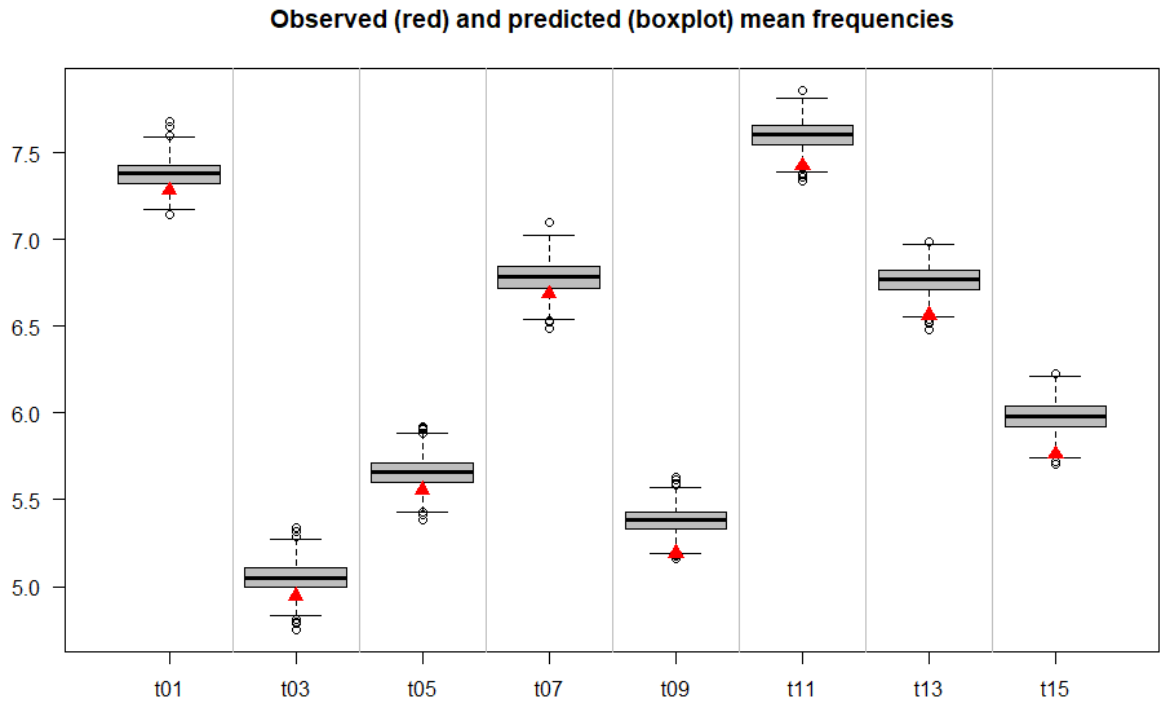


Experiment 3

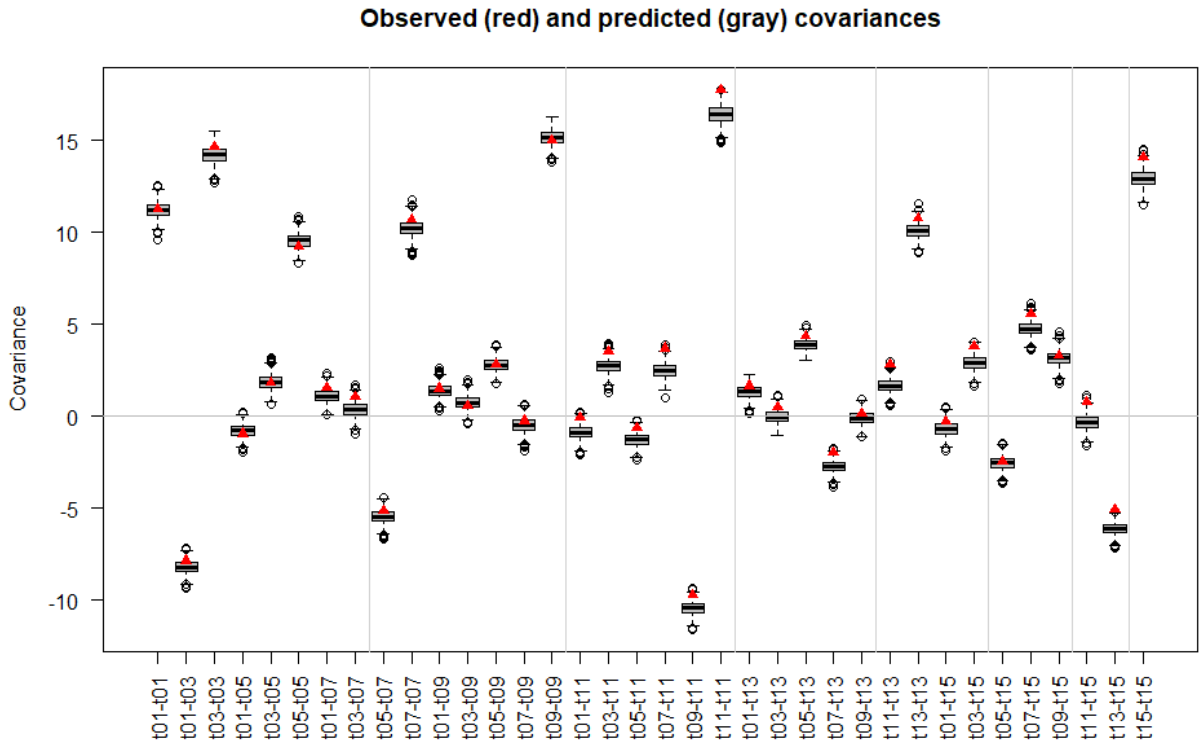
Legend:

t01: correct responses to pleasant pictographs following Kevin primes at Time 1
t03: correct responses to unpleasant pictographs following Kevin primes at Time 1
t05: correct responses to pleasant pictographs following Not Kevin primes at Time 1
t07: correct responses to unpleasant pictographs following Not Kevin primes at Time 1
t09: correct responses to pleasant pictographs following Kevin primes at Time 2
t11: correct responses to unpleasant pictographs following Kevin primes at Time 2
t13: correct responses to pleasant pictographs following Not Kevin primes at Time 2
t15: correct responses to unpleasant pictographs following Not Kevin primes at Time 2

T_1 graph:



T_2 graph:



Alternative AMP model specification

Here we report analyses on data from Experiments 2 and 3 in which the AMP model was specified to test for differences on the Pictograph parameter. Specifically, we constrained the *Affect-Bob* (*Kevin*) parameters to be equal across Time 1 and Time 2, and as well as equal across participants. Similarly, we constrained the *Affect-Not Bob (Not Kevin)* parameters to be equal across Time 1 and Time 2, as well as equal across participants. Model fit and analyses are reported below, followed by parameter estimates and model fit graphs.

Experiment 2

Model fit. At the individual level, the median *p*-value for T_1 was $p = .451$. At the group level, the observed versus predicted values for T_1 were 0.619 and 0.028, respectively, $p < .001$, and the observed versus predicted values for T_2 were 88.843 and 4.942, respectively, $p < .001$. Visual inspection of graphs of the observed versus predicted frequencies and covariances indicates that this alternative model specification did not provide as good fit to data as did the specification reported in the manuscript. Namely, for the alternative model specification three of the observed category frequencies and seven of the observed covariances fell outside of the 95% BCIs for the predicted category frequencies and covariances, whereas for the original

specification all observed frequencies and covariances were within the 95% BCI for the predicted frequencies and covariances.

Additionally, on three model selection indices, the alternative model specification did not provide as good fit to data as did the specification reported in the manuscript. For the alternative model: AIC=31666.88, BIC=31731.32, FIA=15817.94. For the original model: AIC=31607.54, BIC=31671.99, FIA=15788.23.

Planned contrasts. We conducted a series of planned contrasts for each AMP parameter, using the same analytic method as reported in the manuscript.

Pleasant Pictograph parameters were not different between Time 1 (M = 0.592, $BCI_{95\%}$ [0.561, 0.662]) and Time 2 (M = 0.552, $BCI_{95\%}$ [0.503, 0.599]), mean difference = 0.040, $BCI_{95\%}$ [-0.001, 0.082]. Both estimates are reliably > .5, so the *Pleasant Pictograph* parameters reflect positive evaluations of the normatively pleasant pictographs, as should be expected.

Unpleasant Pictograph parameters were higher at Time 1 (M = 0.528, $BCI_{95\%}$ [0.491, 0.566]) than at Time 2 (M = 0.438, $BCI_{95\%}$ [0.380, 0.498]), mean difference = 0.090, $BCI_{95\%}$ [0.040, 0.139]. The Unpleasant Pictograph parameter at Time 2 is reliably < .5, which reflects negative evaluations of the normatively unpleasant pictographs, as should be expected. However, the *Unpleasant Pictograph* parameter at Time 1 is not reliably different from .5, so it reflects neutral evaluations of the normatively unpleasant pictographs, which is theoretically inconsistent.

Misattribution parameters were lower at Time 1 (M = 0.000, $BCI_{95\%}$ [0.000, 0.000]) than at Time 2 (M = 0.016, $BCI_{95\%}$ [0.002, 0.048]), mean difference = -0.016, $BCI_{95\%}$ [-0.048, -0.002].

Experiment 3

Model fit. At the individual level, the median p-value for T_1 was p = .339. At the group level, the observed versus predicted values for T_1 were 1.691 and 0.013, respectively, p < .001, and the observed versus predicted values for T_2 were 103.939 and 2.063, respectively, p < .001. Visual inspection of graphs of the observed versus predicted frequencies and covariances indicates that this alternative model specification did not provide as good a fit to data as did the specification reported in the manuscript. Namely, for the alternative model specification four of the observed category frequencies and 10 of the observed covariances fell outside of the 95% BCIs for the predicted category frequencies and covariances, whereas for the original specification all observed frequencies and covariances were within the 95% BCI for the predicted frequencies and covariances.

Additionally, on three model selection indices, the alternative model specification did not provide as good fit to data as did the specification reported in the manuscript. For the alternative model: AIC=58075.79, BIC=58144.99, FIA=29024.78. For the original model: AIC=57260.81, BIC=57330.02, FIA =28617.25.

Planned contrasts. We conducted a series of planned contrasts for each AMP parameter, using the same analytic method as reported in the manuscript.

Pleasant Pictograph parameters were higher at Time 1 (M = 0.652, $BCI_{95\%}$ [0.622, 0.683]) and Time 2 (M = 0.554, $BCI_{95\%}$ [0.524, 0.584]), mean difference = 0.098, $BCI_{95\%}$ [0.064, 0.133]. Both estimates are reliably > .5, so the *Pleasant Pictograph* parameters reflect positive evaluations of the normatively pleasant pictographs, as should be expected.

Unpleasant Pictograph parameters were higher at Time 1 (M = 0.571, $BCI_{95\%}$ [0.537, 0.606]) than at Time 2 (M = 0.457, $BCI_{95\%}$ [0.426, 0.489]), mean difference = 0.114, $BCI_{95\%}$ [0.079, 0.149]. The Unpleasant Pictograph parameter at Time 2 is reliably < .5, which reflects negative evaluations of the normatively unpleasant pictographs, as should be expected. However, the *Unpleasant Pictograph* parameter at Time 1 is reliably >.5, so it reflects positive evaluations of the normatively unpleasant pictographs, which is theoretically inconsistent.

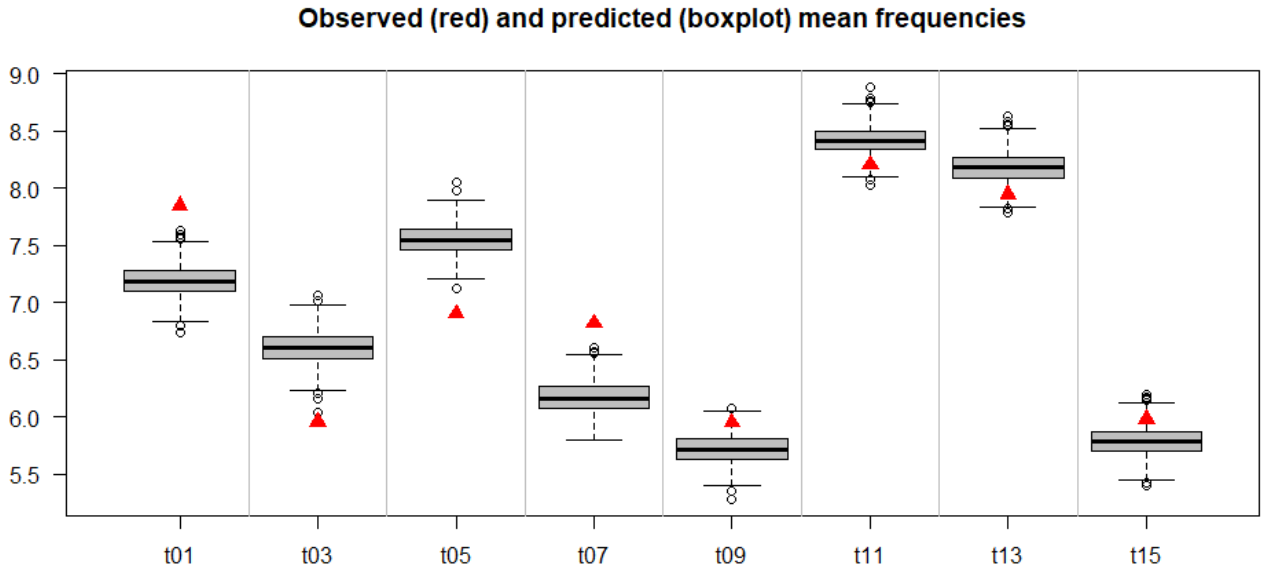
Misattribution parameters were higher at Time 1 (M = 0.036, $BCI_{95\%}$ [0.017, 0.062]) than at Time 2 (M = 0.000, $BCI_{95\%}$ [0.000, 0.000]), mean difference = 0.036, $BCI_{95\%}$ [0.017, 0.062]. Of note, this is the opposite pattern of results as observed in Experiment 2.

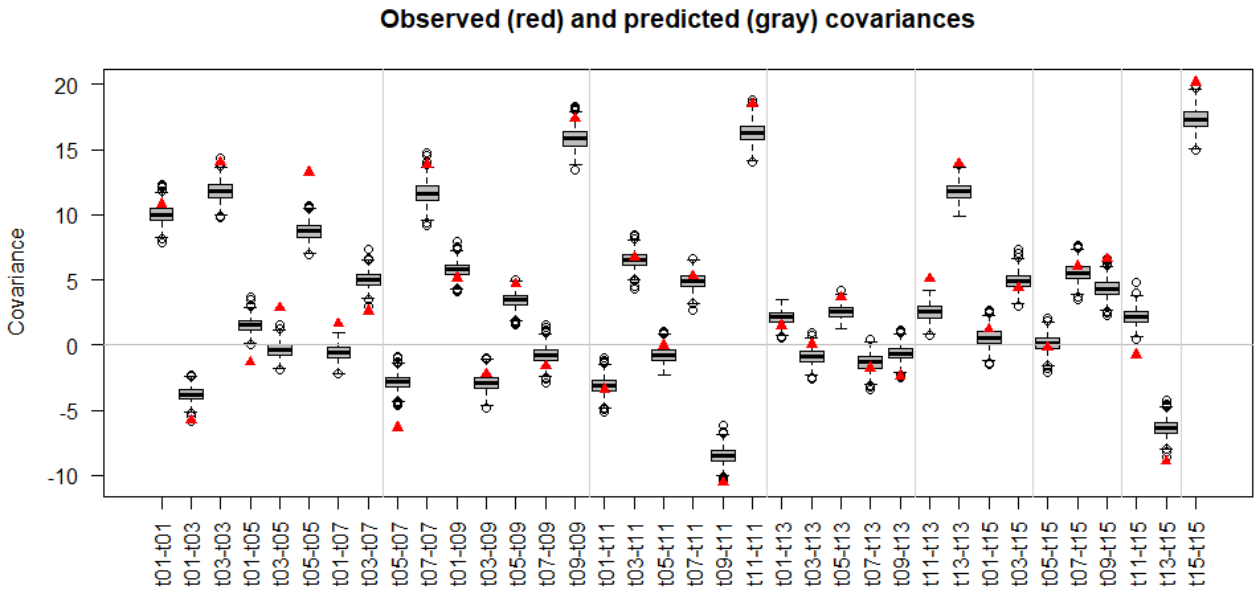
Experiment 2

	Time 1		Time 2	
	M	95% BCI	*M*	95% BCI
Pictograph				
Pleasant	0.592	[0.561, 0.662]	0.552	[0.503, 0.599]
Unpleasant	0.528	[0.491, 0.566]	0.438	[0.380, 0.498]
Misattribution	0.000	[0.000, 0.000]	0.016	[0.002, 0.048]
Activation				
Bob*	0.015	[0.004, 0.030]		
Not Bob*	0.983	[0.965, 0.998]		

* *Note:* Each of these parameters is constrained to be equal across measurement times as well as equal across participants.

Table S4. Model parameter estimates in Study 2 with an alternative model.



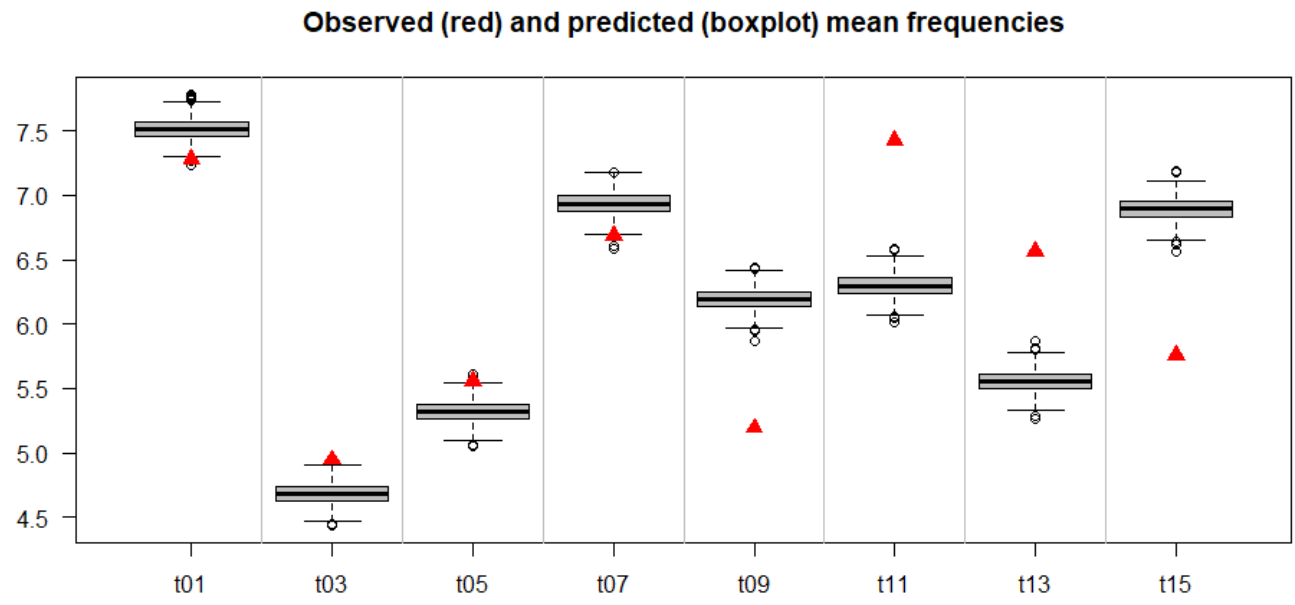


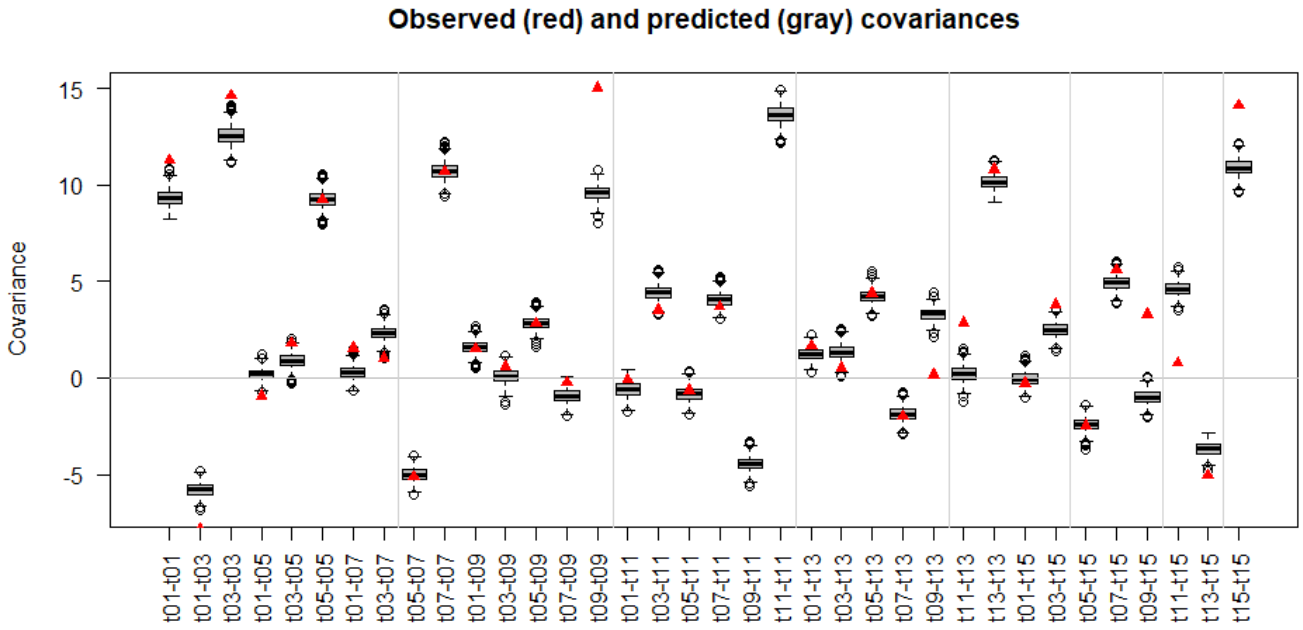
Experiment 3

	Time 1		Time 2	
	M	95% BCI	*M*	95% BCI
Pictograph				
Pleasant	0.652	[0.622, 0.683]	0.554	[0.524, 0.584]
Unpleasant	0.571	[0.537, 0.606]	0.457	[0.426, 0.489]
Misattribution	0.036	[0.017, 0.062]	0.000	[0.000, 0.000]
Activation				
Kevin*	0.975	[0.960, 0.990]		
Not Kevin*	0.025	[0.002, 0.061]		

* *Note:* Each of these parameters is constrained to be equal across measurement times as well as equal across participants.

Table S3. Model parameter estimates in Study 3.


Observed (red) and predicted (boxplot) mean frequencies
7.5
7.0
6.5
6.0
5.5
5.0
4.5
t01
t03
t05
t07
t09
t11
t13
t15


Observed (red) and predicted (gray) covariances
Covariance
15
10
5
0
-5
t01-t01
t01-t03
t03-t03
t01-t05
t03-t05
t05-t05
t01-t07
t03-t07
t05-t07
t07-t07
t01-t09
t03-t09
t05-t09
t07-t09
t09-t09
t01-t11
t03-t11
t05-t11
t07-t11
t09-t11
t11-t11
t01-t13
t03-t13
t05-t13
t07-t13
t09-t13
t11-t13
t13-t13
t01-t15
t03-t15
t05-t15
t07-t15
t09-t15
t11-t15
t13-t15
t15-t15

A process dissociation model of implicit rapid revision in response to diagnostic revelations

[Authors removed for double-blind review]

Stimulus Materials

Study 1

Learning Paradigm Behavior List

Bob bought a friend a potted plant on Valentine's Day
Bob bought plane tickets for his parents on their 25th anniversary for a trip to Hawaii
Bob built a bookcase for his new apartment
Bob built a stained glass lamp shade for his sister's wedding present
Bob buys all of his friends a drink at the bar
Bob called the SPCA upon coming across a hurt animal
Bob does not forget to feed the birds every morning
Bob donates his time at the soup kitchen
Bob drove to Colorado to visit his parents
Bob fixed a squeaking rocking chair for his friend in the country
Bob found an expensive briefcase on the street and placed an ad in the newspaper to locate its owner
Bob gave $20 to the United Appeal
Bob gave a hitchiker a ride to a shelter
Bob gave up his seat on the bus to an older person
Bob gives directions to a driver who is lost
Bob gives the best gifts on Christmas
Bob had a surprise party for his father's birthday
Bob helped a foreign student locate a place to live
Bob helped a friend review for a test
Bob helped a lost child find his way home
Bob helped friends move into a new house
Bob helped his friend fix up his house so his friend could afford a better house for his family
Bob helped his friend plant a large vegetable garden in the backyard
Bob helped some neighboring campers set up their tent
Bob is a reading specialist who volunteered to teach reading in a free school
Bob absentmindedly ran through a red light at an intersection and smashed into the side of another car
Bob became very angry and threatened two children who were collecting insects near his home
Bob bought items at a discount store and returned them for a refund at a different store with higher prices
Bob butted into the front of a long line at a movie theater
Bob cheated during a poker game
Bob cheated on a take-home exam from the university
Bob convinced his boss to fire an employee for arriving late for work
Bob defaced a large rock with spray paint in a public park
Bob deliberately smoked even though there was no smoking allowed in the restaurant
Bob made fun of his friend at a party
Bob did not show up for a prearranged tennis game
Bob didn't show up as a material witness for a drug case that involved a friend
Bob flunked out of college after one term
Bob found a good watch in a park and pawned it
Bob irritated his friend by asking distracting and irrelevant questions
Bob got drunk and insulted everybody at a cocktail party

Bob had his driver's license suspended for drunken driving
Bob had someone else take a math final for him
Bob ignored a hit-and-run accident in which a person was thrown from a car
Bob ignored someone's cries for help on the ski slope by skiing by too fast to stop
Bob insulted a man who asked for directions to the nearest post office
Bob kicked his dog for eating some cheese left near the edge of the table
Bob left a burner on when leaving his house
Bob left his campfire burning at night causing a small forest fire
Bob loudly made racist statements to his friends while in a tavern

Time 2 Diagnostic Behavior

You are now going to learn one more piece of information about Bob. This information is more recent than the information you learned about earlier. Also note that this piece of information is CHARACTERISTIC of Bob.

Please pay careful attention to this piece of information because it may or may not be at odds with what you learned about Bob earlier. If it is inconsistent, please update your impression of him based on this new piece of characteristic information.

Click ">>" when you are ready to continue.

Bob was recently convicted of molesting children.

(Remember, this information is characteristic of Bob, and you will be tested on it later in the experiment.)

Explicit evaluations

We'd like you to answer some questions about your perceptions of Bob.

1. How likeable is Bob? (1-very unlikeable – 7-very likeable)

Please rate Bob on the following scales.

1. Bob is … 1-very bad – 7- very good
2. Bob is … 1-very unpleasant – 7-very pleasant
3. Bob is … 1-very disagreeable – 7-very disagreeable
4. Bob is … 1-very uncaring – 7-very caring
5. Bob is … 1-very cruel – 7-very kind

Demographics Questionnaire

Please provide the following demographics about yourself.

Gender:

(a) male
(b) female

Age: (free response)

Please specify your race (Choose one or more categories)

[ ] White/caucasian (Anglo/Euro) American
[ ] Black or African American
[ ] Asian or Asian American
[ ] American Indian or Pacific Islander
[ ] Hispanic/Latino
[ ] Multicultural
[ ] Prefer not to say

Highest level of education completed:

( ) Less than a high school
( ) High school diploma
( ) Vocational Training
( ) Attended College
( ) Bachelor's Degree
( ) Graduate Degree
( ) Prefer not to say

With which US political party do you identify more strongly? 1-strongly Republican – 7-Strongly Democrat

Politically, how liberal/conservative are you on social issues? 1-Very liberal – 7-Very conservative

Politically, how liberal/conservative are you on fiscal issues? 1-Very liberal – 7-Very conservative

To what extent have you previously participated in surveys like this one? 1-Nothing like this scenario – 5-Exactly this scenario

Study 2

Learning Paradigm Behavior List

Bob bought a friend a potted plant on Valentine's Day
Bob bought plane tickets for his parents on their 25th anniversary for a trip to Hawaii
Bob built a bookcase for his new apartment
Bob built a stained glass lamp shade for his sister's wedding present
Bob buys all of his friends a drink at the bar
Bob called the SPCA upon coming across a hurt animal
Bob does not forget to feed the birds every morning
Bob donates his time at the soup kitchen
Bob drove to Colorado to visit his parents

Bob fixed a squeaking rocking chair for his friend in the country
Bob found an expensive briefcase on the street and placed an ad in the newspaper to locate its owner
Bob gave $20 to the United Appeal
Bob gave a hitchiker a ride to a shelter
Bob gave up his seat on the bus to an older person
Bob gives directions to a driver who is lost
Bob absentmindedly ran through a red light at an intersection and smashed into the side of another car
Bob became very angry and threatened two children who were collecting insects near his home
Bob bought items at a discount store and returned them for a refund at a different store with higher prices
Bob butted into the front of a long line at a movie theater
Bob cheated during a poker game
Bob cheated on a take-home exam from the university
Bob convinced his boss to fire an employee for arriving late for work
Bob defaced a large rock with spray paint in a public park
Bob deliberately smoked even though there was no smoking allowed in the restaurant
Bob made fun of his friend at a party
Bob did not show up for a prearranged tennis game
Bob didn't show up as a material witness for a drug case that involved a friend
Bob flunked out of college after one term
Bob found a good watch in a park and pawned it
Bob irritated his friend by asking distracting and irrelevant questions

Time 2 Diagnostic Behavior

Part 2

This is Part 2 of the experiment. You are now going to learn one more piece of information about Bob. This information is more recent than the information you learned about earlier. Also note that this piece of information is CHARACTERISTIC of Bob.

Please pay careful attention to this piece of information because it may or may not be at odds with what you learned about Bob earlier. If it is inconsistent, please update your impression of him based on this new piece of characteristic information.

Click ">>" when you are ready to continue.

Bob recently mutilated a small, defenseless animal.

(Remember, this information is characteristic of Bob, and you will be tested on it later in the experiment.)

Explicit evaluations

We'd like you to answer some questions about your perceptions of Bob.

2. How likeable is Bob? (1-very unlikeable – 7-very likeable)

Please rate Bob on the following scales.

6. Bob is … 1-very bad – 7- very good
7. Bob is … 1-very unpleasant – 7-very pleasant
8. Bob is … 1-very disagreeable – 7-very disagreeable
9. Bob is … 1-very uncaring – 7-very caring
10. Bob is … 1-very cruel – 7-very kind

Exploratory Items

1. How much did you pay attention to the information that was presented about Bob?
2. How motivated were you to form an accurate impression of Bob based on what you were told in Parts 1 and 2 of the experiment?
3. How confident are you in the impression you've formed of Bob based on what you were told in Parts 1 and 2 of the experiment?
4. In Part 2 of the experiment, we gave you one last piece of information about Bob. What was that last piece of information? Please enter this behavior as quickly as possible and then press >> to continue.

The last piece of information we gave in Part 2 was that Bob mutilated a small, defenseless animal.

1. We'd now like you to reflect only on this piece of information. How likely do you think it is that Bob actually engaged in this behavior?
2. How much do you think learning this last piece of information tells you about Bob's true character. In other words, do you think this last piece of information is likely to predict how he will act in the future?
3. To what extent do you think that Bob engaged in this action consciously and intentionally vs. unintentionally or due to situational circumstances?
4. How positive or negative do you consider this information to be?
5. To what extent was this final piece of information on your mind when you were completing the cognitive task in which you were asked to judge the pleasantness of Chinese pictographs?
6. To what extent do you agree with the following: "After reading the last piece of information about Bob, it was easy to see that he is a bad person."
7. To what extent do you agree with the following: "Bob has both positive and negative traits."
8. How motivated were you to remember the information you learned about Bob after reading it?
9. To what extent did you focus on the negative consequences of Bob's actions when you were forming your impression of him?
10. To what extent do you personally care about harm to animals?
11. To what extent did you form a mental image of Bob actually mutilating an animal?
12. How extensively did you think about the information you learned about Bob?

Demographics Questionnaire

Please provide the following demographics about yourself.

Gender:
(a) male
(b) female

Age: (free response)

Please specify your race (Choose one or more categories)
[ ] White/caucasian (Anglo/Euro) American
[ ] Black or African American
[ ] Asian or Asian American
[ ] American Indian or Pacific Islander
[ ] Hispanic/Latino
[ ] Multicultural
[ ] Prefer not to say

Highest level of education completed:
( ) Less than a high school
( ) High school diploma
( ) Vocational Training
( ) Attended College
( ) Bachelor's Degree
( ) Graduate Degree
( ) Prefer not to say

With which US political party do you identify more strongly? 1-strongly Republican – 7-Strongly Democrat

Politically, how liberal/conservative are you on social issues? 1-Very liberal – 7-Very conservative

Politically, how liberal/conservative are you on fiscal issues? 1-Very liberal – 7-Very conservative

To what extent have you previously participated in surveys like this one? 1-Nothing like this scenario – 5-Exactly this scenario

Study 3

Learning Paradigm Behavior List

Kevin bought a friend a potted plant on Valentine's Day
Kevin bought plane tickets for his parents on their 25th anniversary for a trip to Hawaii
Kevin built a bookcase for his new apartment

Kevin built a stained glass lamp shade for his sister's wedding present
Kevin buys all of his friends a drink at the bar
Kevin called the SPCA upon coming across a hurt animal
Kevin does not forget to feed the birds every morning
Kevin donates his time at the soup kitchen
Kevin drove to Colorado to visit his parents
Kevin fixed a squeaking rocking chair for his friend in the country
Kevin found an expensive briefcase on the street and placed an ad in the newspaper to locate its owner
Kevin gave $20 to the United Appeal
Kevin gave a hitchhiker a ride to a shelter
Kevin gave up his seat on the bus to an older person
Kevin gives directions to a driver who is lost
Kevin gives the best gifts on Christmas
Kevin had a surprise party for his father's birthday
Kevin helped a foreign student locate a place to live
Kevin helped a friend review for a test
Kevin helped a lost child find his way home
Kevin helped friends move into a new house
Kevin helped his friend fix up his house so his friend could afford a better house for his family
Kevin helped his friend plant a large vegetable garden in the backyard
Kevin helped some neighboring campers set up their tent
Kevin is a reading specialist who volunteered to teach reading in a free school
Kevin absentmindedly ran through a red light at an intersection and smashed into the side of another car
Kevin became very angry and threatened two children who were collecting insects near his home
Kevin bought items at a discount store and returned them for a refund at a different store with higher prices
Kevin butted into the front of a long line at a movie theater
Kevin cheated during a poker game
Kevin cheated on a take-home exam from the university
Kevin convinced his boss to fire an employee for arriving late for work
Kevin defaced a large rock with spray paint in a public park
Kevin deliberately smoked even though there was no smoking allowed in the restaurant
Kevin made fun of his friend at a party
Kevin did not show up for a prearranged tennis game
Kevin didn't show up as a material witness for a drug case that involved a friend
Kevin flunked out of college after one term
Kevin found a good watch in a park and pawned it
Kevin irritated his friend by asking distracting and irrelevant questions
Kevin got drunk and insulted everybody at a cocktail party
Kevin had his driver's license suspended for drunken driving
Kevin had someone else take a math final for him
Kevin ignored a hit-and-run accident in which a person was thrown from a car
Kevin ignored someone's cries for help on the ski slope by skiing by too fast to stop
Kevin insulted a man who asked for directions to the nearest post office
Kevin kicked his dog for eating some cheese left near the edge of the table
Kevin left a burner on when leaving his house
Kevin left his campfire burning at night causing a small forest fire
Kevin loudly made racist statements to his friends while in a tavern

Time 2 Diagnostic Behavior

You will now learn some additional information about Kevin.

Imagine your boss asks you to do some research online, categorizing arrest records from your local Police Department website. While you are going through hundreds of entries, you stumble across more than one arrest record for the new guy, Kevin, who was just hired at the office. The most recent court case is laid out in front of you, including an appendix with pictures. The record shows that **Kevin was arrested a few years back for child molestation of his young niece**.

Explicit evaluations

We'd like you to answer some questions about your perceptions of Kevin.

3. How likeable is Kevin? (1-very unlikeable – 7-very likeable)

Please rate Kevin on the following scales.

11. Kevin is … 1-very bad – 7- very good
12. Kevin is … 1-very unpleasant – 7-very pleasant
13. Kevin is … 1-very disagreeable – 7-very disagreeable
14. Kevin is … 1-very uncaring – 7-very caring
15. Kevin is … 1-very cruel – 7-very kind

Additional Items

1. We'd now like you to reflect on the later information you learned about child molestation. How likely do you think it is that Kevin actually engaged in this behavior?
2. How positive or negative do you consider the information you learned to be?
3. For this question, we'd like you to think about how much learning this information would shed light on Kevin's character. In other words, does learning this information about Kevin tell us something about who Kevin is and how he is likely to behave in the future? Or is the behavior more likely to be a product of extenuating situational circumstances and unlikely to predict his future actions?

Demographics Questionnaire

Please provide the following demographics about yourself.

Gender:
(a) male
(b) female

Age: (free response)

Do you speak Mandarin?
( ) Yes
( ) No

Do you speak Cantonese?

( ) Yes
( ) No

Please specify your race (Choose one or more categories)

[ ] White/caucasian (Anglo/Euro) American
[ ] Black or African American
[ ] Asian or Asian American
[ ] American Indian or Pacific Islander
[ ] Hispanic/Latino
[ ] Multicultural
[ ] Prefer not to say

Highest level of education completed:

( ) Less than a high school
( ) High school diploma
( ) Vocational Training
( ) Attended College
( ) Bachelor's Degree
( ) Graduate Degree
( ) Prefer not to say

With which US political party do you identify more strongly? 1-strongly Republican – 7-Strongly Democrat

Politically, how liberal/conservative are you on social issues? 1-Very liberal – 7-Very conservative

Politically, how liberal/conservative are you on fiscal issues? 1-Very liberal – 7-Very conservative

To what extent have you previously participated in surveys like this one? 1-Nothing like this scenario – 5-Exactly this scenario