Rydell & Mc | Study 4

(10) H/L - rel?
(11) attitude = avg (2), distance = single (1) pg 1004 plus..
(12) ~~att~~ (1), pg 998
(13) att (1), pg 998
(14) Not (0) for desired control, obs (1) for distance, pg 1004
(15) Sip Judge (4) for desired control, obs (5) for distance pg 1004
(16) Not rep (3), pg 999
(17) pics (1), pg 998
(18) One (1), pg 998
(19) 2nd (2) I, p1004 ~~E, I, B~~
(20) 1st (1) E, p1004
(21) 3rd (3) B, p1004
(22) bet (1)
(23) bet (1)
(24) Same (0)
(25) Same (0)
(26) sd-1 pg 1004, 998
(27) sd-1 pg 1004, 998
(28) desired dist - 9, Seat = 3 pg 1004
(29) spec = 3, pg 1004 & 998
(30) spec = 3, pg 1004 & 998
(31) comp - 2, pg 998, 1004
(32) Non-rec (0), pg 995
(33) dud (2)

error: put words (0) before

Understanding Implicit and Explicit Attitude Change: A Systems of Reasoning Analysis

Robert J. Rydell
University of California, Santa Barbara

Allen R. McConnell
Miami University

There is considerable controversy about how to conceptualize implicit and explicit attitudes, reflecting substantial speculation about the mechanisms involved in implicit and explicit attitude formation and change. To investigate this issue, the current work examines the processes by which new attitudes are formed and changed and how these attitudes predict behavior. Five experiments support a systems of reasoning approach to implicit and explicit attitude change. Specifically, explicit attitudes were shaped in a manner consistent with fast-changing processes, were affected by explicit processing goals, and uniquely predicted more deliberate behavioral intentions. Conversely, implicit attitudes reflected an associative system characterized by a slower process of repeated pairings between an attitude object and related evaluations, were unaffected by explicit processing goals, uniquely predicted spontaneous behaviors, and were exclusively affected by associative information about the attitude object that was not available for higher order cognition.

10
32

Keywords: implicit attitudes, explicit attitudes, attitude change

The study of attitudes—evaluations of the self, individuals, groups, and other objects—has a long and rich history in social psychology (Eagly & Chaiken, 1993). In recent years, the focus of attitude research has shifted from understanding explicit attitudes (i.e., attitudes that people can report and for which activation can be consciously controlled) to examining implicit attitudes (i.e., attitudes for which people do not initially have conscious access and for which activation cannot be controlled).[1] Past research has shown that relying on implicit rather than explicit measures of attitudes can circumvent self-presentational motives (e.g., Dunton & Fazio, 1997) and can often uniquely predict spontaneous behaviors (e.g., McConnell & Leibold, 2001); however, less is known about the processes underlying how implicit and explicit attitudes form and operate. The current work posited that there are important differences between them, especially in how they change. Specifically, we propose that explicit attitudes form and change through the use of fast-learning, rule-based reasoning, whereas implicit attitudes form and change through the use of slow-learning, associative reasoning (Sloman, 1996).

Heretofore, implicit attitude change and explicit attitude change have been studied in relative isolation. Indeed, research on explicit attitude change has been one of the most productive areas of study in social psychology (Eagly & Chaiken, 1993; Petty & Wegener, 1998). Although some researchers have found that implicit attitudes are relatively difficult to change with conventional attitude change manipulations (e.g., Gawronski & Strack, 2004; Gregg, Seibt, & Banaji, 2006; Petty, Tormala, Briñol, & Jarvis, 2006), other research has demonstrated that implicit attitudes can change relatively quickly in response to contextual stimuli or social roles (e.g., Barden, Maddux, Petty, & Brewer, 2004; Dasgupta & Greenwald, 2001; Wittenbrink, Judd, & Park, 2001). But despite these demonstrations, the theory underlying implicit attitude change is relatively underdeveloped (see Devine, 2001; Fazio & Olson, 2003; Wilson, Lindsey, & Schooler, 2000), and experimental paradigms that can systematically examine the concurrent formation and change of implicit and explicit attitudes

Robert J. Rydell, Department of Psychology, University of California, Santa Barbara; Allen R. McConnell, Department of Psychology, Miami University.

This research was supported by National Institute of Mental Health Grant MH068279 and National Science Foundation Grant BCS 0516931. Portions of this work were submitted by Robert J. Rydell in partial fulfillment of the requirements for a doctoral degree from Miami University. We thank Doris Bergen, Heather Claypool, Dave Hamilton, Kurt Hugenberg, Diane Mackie, Jeff Sherman, and Laura Strain for their extremely helpful comments and guidance in this research.

Correspondence concerning this article should be addressed to Robert J. Rydell, Department of Psychology, University of California, Santa Barbara, Santa Barbara, CA 93106-9660. E-mail: rydell@psych.ucsb.edu

[1] Although there is disagreement about the use of the terms *implicit attitudes* and *explicit attitudes* in the literature (e.g., Fazio & Olson, 2003), we agree with Strack and Deutsch (2004) who note that "explicit and implicit measures are defined by the cognitive operations that they capture. In this sense, explicit measures tap into people's knowledge or beliefs, implicit measures tap into their associative structures" (p. 239; see also, Wilson et al., 2000). Because we contrast and compare implicit and explicit measures, we use the terms *implicit attitudes* and *explicit attitudes* throughout this article.

Journal of Personality and Social Psychology, 2006, Vol. 91, No. 6, 995–1008
 DOI: 10.1037/0022-3514.91.6.995

toward him. Second, the introduction of the counterattitudinal information provided a window in which explicit (which relies on the fast-learning system) but not implicit (which relies on the slow-learning system) attitudes should change in the face of new target-relevant information. Thus, we have an opportunity to assess and understand how this new information affects implicit attitudes and explicit attitudes differently, shedding light on the processes involved in their change.

We expected to observe that people would quickly change their explicit attitudes in the face of counterattitudinal information, especially when the initial learning was very consistent (Kerpelman & Himmelfarb, 1971) and when the counterattitudinal information was negative (Fiske, 1980). However, we did not expect implicit attitudes to change as quickly in response to a modest amount of counterattitudinal information nor did we expect to observe a valence asymmetry for implicit attitudes.

Method

Participants. A sample of 170 undergraduates at Miami University participated in return for research credit in their introductory psychology courses. They were randomly assigned to a 2 (valence of learned attitude: positive vs. negative) × 2 (level of reinforcement: 100%, 75%) × 2 (counterattitudinal condition: control vs. counterattitudinal conditioning) between-subjects factorial.

Learning task. The current work used a modified version of the attitude learning paradigm developed by Kerpelman and Himmelfarb (1971). Specifically, participants were presented with a target person's behaviors that were either relatively positive or negative in valence, and participants judged whether each behavior was characteristic or uncharacteristic of him. As part of a between-subjects manipulation, participants were given different levels of reinforcement in their responses, leading them to form different attitudes toward him.

First, participants completed the learning task on a computer, in which they were told that they would be receiving information about a person named *Bob*. In the initial learning trials, participants read 100 behaviors performed by *Bob* while a picture of *Bob* was presented on the computer monitor directly above each behavior.[3] After reading each behavior, participants indicated whether they believed that the behavior was characteristic or uncharacteristic of *Bob* by pressing the *C* key (characteristic) or the *U* key (uncharacteristic). After they responded, participants were given feedback about whether the behavior was characteristic of *Bob* for 5 s. Specifically, feedback consisted of the word *correct* (in blue text) or *incorrect* (in red text) positioned in the center of the computer monitor and, at the same time, the behavior was restated "correctly," on the basis of the assigned reinforcement condition, at the bottom of the computer monitor (e.g., "Helping the neighborhood children is characteristic of Bob." or "Helping the neighborhood children is uncharacteristic of Bob."). In the initial 100 learning trials, the feedback given portrayed *Bob* as positive or as negative in 100% or in 75% of the behaviors (with 25 of the trials in the 75% reinforcement condition being counterattitudinal). The ordering of the behaviors and feedback were randomly determined (in accordance with the experimental condition) for each participant.

Following these 100 trials, participants in the control condition received 20 neutral trials (i.e., the behavior performed by *Bob* was neither positive nor negative; e.g., "Bob waited at the street corner."). However, participants in the counterattitudinal condition (20 CA) received counterattitudinal feedback about *Bob* on 20 trials (i.e., the behaviors that were described as characteristic or uncharacteristic of *Bob* were opposite of the valence presented during the initial learning trials). Finally, participants completed implicit and explicit attitude measures.[4]

Explicit attitude measure. To assess explicit attitudes, participants judged how likable *Bob* was on a scale ranging from 1 (*very unlikable*) to 9 (*very likable*). In addition, they completed five semantic differential scales, each using a 9-point scale to describe *Bob*: good–bad, pleasant–mean, agreeable–disagreeable, caring–uncaring, and kind–cruel. Further, participants provided their evaluation of *Bob* on a feeling thermometer that ranged in temperature from 0° to 100°. The response for each explicit measure was standardized and an overall mean was computed (in all experiments to be reported, $\alpha s > .90$). Then the standardized scores in the negative valance condition were reverse scored so that greater scores on this measure indicated that explicit attitudes were more extreme in the direction of initial learning.

Implicit attitude measure. The Implicit Associations Test (IAT; Greenwald, McGhee, & Schwartz, 1998) was used to assess implicit attitudes toward *Bob*. The IAT had 26 stimuli: 1 picture of *Bob*, 5 different pictures of White men who were not *Bob*, 10 positive adjectives (e.g., *wonderful*), and 10 negative adjectives (e.g., *disgusting*). All stimuli were presented in the center of the monitor, and the adjectives were always presented in lowercase letters.

This IAT task was a modified version of the task used by McConnell and Leibold (2001), featuring seven blocks with 20 trials per block. Participants were informed that the task involved making category judgments for a variety of stimuli (photos or words) presented on a computer monitor by using one of two responses (the *D* or *K* keys on the keyboard). During each block, category label reminders were displayed on the left and right sides of the display (assignment of particular labels to the *D* and *K* keys was counterbalanced across participants and produced no effects). Participants were instructed to complete that task quickly while also minimizing errors, and they were told to keep their index fingers on the *D* and *K* keys throughout the experiment to minimize delays in responding. There was a 250-ms intertrial interval.

In Block 1, participants judged photos of *Bob* or *not Bob* and in Block 2 they judged whether the adjectives were "negative" or "positive." In Blocks 3 and 4 (Combination 1), participants judged whether the stimuli were "Bob or negative" or "not Bob or positive." In Block 5, participants performed the same judgment task as Block 2 except the assignment of response keys assigned to the two valence categories was reversed. Finally, in Blocks 6 and 7 (Combination 2), participants judged whether the stimuli were "Bob or positive" or "not Bob or negative." As in past IAT research, half of the participants performed Combination 1 in Blocks 3–4 and Combination 2 in Blocks 6–7, whereas the rest performed Combination 2 in Blocks 3–4 and Combination 1 in Blocks 6–7 (this counterbalancing manipulation produced no effects).

In order to assess implicit attitudes toward *Bob*, we subtracted the mean response latencies of Combination 2 from the mean response latencies of Combination 1 (regardless of the order they were completed).[5] Again, the

---

[3] Photographs of one of 5 different White males were randomly presented as *Bob*. These 5 White males were judged as equal in attractiveness and the target used did not affect the results in any of the experiments. The positive and the negative behaviors used in the current work were borrowed from those developed by McConnell, Sherman, and Hamilton (1994a).

[4] In all of the experiments, half of the participants completed the implicit measure first and the other half completed the explicit measure first. This order variable produced no effects in any of the studies and thus is not discussed further.

[5] Following Greenwald et al. (1998), all trials in the critical blocks were retained, responses faster than 300 ms were recoded as 300 ms, and trials slower than 3,000 ms were recoded as 3,000 ms. After any such adjustments were made, each latency was then log transformed to reduce positive skew inherent in response latency data (Fazio, 1990). Alternative scoring techniques for the IAT (e.g., Greenwald, Nosek, & Banaji, 2003) produced the same results in all studies reported. Analyses were performed on the log-transformed values, but means are reported as standardized scores.

standardized scores in the negative valance condition were reverse scored so that greater scores on this measure indicated that implicit attitudes were more extreme in the direction of initial learning.

Results

Explicit attitudes. To examine whether explicit attitudes changed in response to small amounts of counterattitudinal information and were more likely to show attitude change with greater initial reinforcement, a 2 (valence of learned attitude) × 2 (level of reinforcement) × 2 (counterattitudinal condition) analysis of variance (ANOVA) was conducted on explicit attitude extremity (see Figure 1). First, there were significant main effects of level of reinforcement, $F(1, 162) = 39.22$, $p < .001$, and of counterattitudinal condition, $F(1, 162) = 89.90$, $p < .001$. As one would expect, the main effect of level of reinforcement showed that explicit attitudes were more extreme in the direction of initial learning in the 100% reinforcement condition ($M = 0.94$, $SD = 0.62$) than in the 75% reinforcement condition ($M = 0.54$, $SD = 0.45$). Similarly, the main effect of counterattitudinal condition revealed that explicit attitudes were more extreme in the direction of initial learning in the control condition ($M = 1.04$, $SD = 0.52$) than in the 20 CA condition ($M = 0.44$, $SD = 0.47$). More important, the anticipated two-way interaction between level of reinforcement and counterattitudinal condition was significant, $F(1, 162) = 19.06$, $p < .001$. To examine this interaction, the simple effect of counterattitudinal condition was examined for each level of reinforcement. In the 75% reinforcement condition, there was a simple effect of counterattitudinal condition, $F(1, 162) = 13.54$, $p < .001$, showing that participants in the control condition had more extreme explicit attitudes toward *Bob* ($M = 0.71$, $SD = 0.42$); than participants in the 20 CA condition ($M = 0.38$, $SD = 0.41$). In the 100% reinforcement condition, there was an even stronger effect of counterattitudinal condition, $F(1, 162) = 97.03$, $p < .001$, indicating that although participants in the control condition had especially extreme explicit attitudes (in the direction of initial conditioning) toward *Bob* ($M = 1.37$, $SD = 0.37$), counterattitudinal information led to far less extreme attitudes toward *Bob* ($M = 0.51$, $SD = 0.52$). Thus, the interaction reflects the much larger effect of counterattitudinal condition on explicit attitude extremity in the 100% reinforcement condition than in the 75% reinforcement condition (replicating Kerpelman & Himmelfarb, 1971). Consistent with negative asymmetries, the two-way interaction between counterattitudinal condition and valence of learned attitude was also significant, $F(1, 162) = 16.64$, $p < .001$. In the positive learned attitudes condition, those in the control condition had far more extreme explicit attitudes ($M = 1.17$, $SD = 0.59$) than those in the 20 CA condition ($M = 0.28$, $SD = 0.47$), $F(1, 162) = 76.70$, $p < .001$. In the negative learned attitudes condition, this effect was significant but weaker, with those in the control condition having more extreme explicit attitudes ($M = 0.91$, $SD = 0.39$) than those in the 20 CA condition ($M = 0.61$, $SD = 0.42$), $F(1, 162) = 8.12$, $p < .005$. In other words, negative counterattitudinal information had a greater impact on attitude extremity than did positive counterattitudinal information (e.g., Fiske, 1980; Skowronski & Carlston, 1987). No other effects were significant.

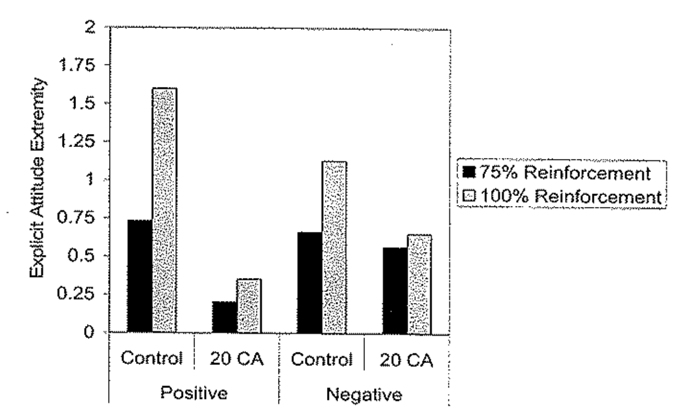

Figure 1. Explicit attitude extremity as a function of reinforcement and counterattitudinal condition (20 CA) in Experiment 1. Values for the negative initial learning condition have been reverse scored to reflect attitude extremity.

Implicit attitudes. As with the explicit attitude data, a 2 (valence of learned attitude) × 2 (level of reinforcement) × 2 (counterattitudinal condition) ANOVA was conducted on implicit attitude extremity (see Figure 2). In stark contrast to the explicit attitudes, the interaction of reinforcement and counterattitudinal condition and the interaction of valence of learned attitude and counterattitudinal condition were not significant for implicit attitudes ($Fs < 1$). In fact, the only effect to obtain for implicit attitudes was an effect showing the that grand mean was significantly different than zero, $F(1, 166) = 55.12$, $p < .001$ ($M = 0.50$, $SD = 0.87$). This shows that participants formed implicit attitudes about *Bob* in accordance with the valence of their initial learning but that subsequent counterattitudinal information had no impact on them. It is important that this effect was not statistically moderated by any of the experimental manipulations, showing no evidence of changes in attitude extremity or negative asymmetries for implicit attitudes.[6]

16

Discussion

A systems of reasoning conceptualization of attitude change was supported in this experiment because explicit attitudes were changed dramatically by the introduction of counterattitudinal information, whereas implicit attitudes were unaltered by this same information. This suggests that explicit attitudes are the product of a fast-learning system, whereas implicit attitudes reflect a slow-learning system. In this study, participants did form implicit attitudes about *Bob*, but,

[6] When implicit and explicit attitude measures were simply standardized (i.e., the standardized attitudes in the negative valence of learned attitude condition were not reverse scored) and submitted to a 2 (valence of learned attitude) × 2 (level of reinforcement) × 2 (counterattitudinal condition) × 2 (standardized attitude measure: implicit vs. explicit, a repeated measure) mixed-model factorial ANOVA, the expected four-way interaction was significant, $F(1, 156) = 3.98$, $p < .05$, reflecting differential responses to counterattitudinal feedback for explicit attitudes and implicit attitudes. In all subsequent experiments, similar omnibus analyses were conducted by using the standardized attitude measure as a within-subjects factor, and the highest order interaction obtained in each experiment ($Fs > 3.88$, $ps < .03$). These analyses reveal that examining implicit and explicit attitudes separately throughout the article is justified inferentially. In the current work, we present the data as examining attitude extremity by reverse scoring the negative learning condition attitude measures in order to simplify the presentation of how implicit and explicit attitudes are differentially affected by our manipulations.

tudes predict more deliberative, intentional behavior (e.g., Jellison et al., 2004; McConnell & Leibold, 2001). We were interested in whether the attitudes toward *Bob* created in the current experiments could predict behavior in the same manner as past research. Specifically, would explicit attitudes toward *Bob* only predict deliberate judgments about him but not predict more subtle forms of behavior (i.e., seating distance)? Similarly, would implicit attitudes toward *Bob* only predict subtle behaviors but not explicit judgments about him? Experiment 4 tests these predictions, anticipating unique predictive value for implicit and explicit attitudes.

These findings could be important for at least three additional reasons. First, past research has shown such double dissociations on the basis of measures of group prejudice (e.g., Dovidio et al., 2002; Jellison et al., 2004); however, this would be the first time that such effects have been shown for a different type of attitude object (i.e., a target person). Second, this previous work has shown these outcomes for groups with preexisting attitudes, whereas this would be the first study to demonstrate such dissociation effects on the basis of attitudes engineered in a controlled laboratory setting. For example, it is possible that cultural prescriptions might shape both implicit prejudice and subtle forms of social behavior toward social group members, providing the appearance of an attitude–behavior relation when, in fact, other factors may produce both. By engineering attitudes in the laboratory without any other target-relevant knowledge, it is far more likely that behavior reflects the influence of attitudes directly. Finally, if we show that implicit attitudes have unique predictive utility for subtle behavior in this study, then the findings would argue against concerns that our implicit measure has poor sensitivity. One might argue that slow changes on our implicit measure may reflect a relatively weak measure (i.e., it is simply less responsive to change than our explicit measures) rather than a slow-learning system. By establishing that our implicit (but not explicit) attitude measure can uniquely predict theoretically derived types of behavior, we could provide evidence inconsistent with a position that our implicit attitude measure is simply a poor measure.

Method

19-25

Participants. A sample of 29 undergraduates at Miami University participated in return for research credit in their introductory psychology courses. Participants were randomly assigned to receive no counterattitudinal information about *Bob* (control) or to receive 20 counterattitudinal pieces of information about *Bob* (20 CA).

33, 26, 27, 29, 30, 31

Procedure. All materials, methods, and measures paralleled Experiment 1, with these exceptions. ~~First, only the positive-valence condition was used,~~ and only the 100% reinforcement condition was used. The two experimental conditions (control and 20 CA) were selected to maximize the discrepancy between implicit and explicit attitudes. In Experiment 1 there was a drastic change in explicit attitudes between the control and the 20 CA conditions, however there was no difference in implicit attitudes between them. Additionally, as in Experiment 3, because there was no negative initial learning condition to reverse score, greater standardized measures of attitudes reflected more positive attitudes toward *Bob*.

11

In addition to the attitude measures, participants completed explicit judgments of desire for social contact with *Bob*. Specifically, participants rated the extent to which they would want to have *Bob* as a neighbor, friend, classmate, roommate, and family member, each on 100-point scales ($\alpha = .92$). Greater scores on this measure indicated that they wanted more social contact with *Bob*.

14 15 29

29 30 28 14 11 14 15

After completing the attitude measures and the explicit social contact judgments, participants were told that they would "have a 2-min get acquainted session with Bob." They were escorted to a different room in which two chairs were set 221 cm apart. One chair had a book bag and a book next to it (where *Bob* was supposedly sitting), the other chair (for the participant) was on wheels and set against the wall of the room. The experimenter told each participant, "It looks like *Bob* has stepped out for a moment. Take that seat against the wall and move it so that you can have a face-to-face conversation with Bob." Participants took the seat and moved it into a position to converse with *Bob*. Afterward, they were told that they were not going to meet *Bob* and were then debriefed. The seating distance between the participant's chair and the chair where *Bob* had supposedly been sitting served as our measure of subtle, spontaneous behavior.

Results

The attitude measures were examined with one-way ANOVAs of counterattitudinal condition. The only effect to obtain was the predicted effect of counterattitudinal condition for explicit attitudes, $F(1, 27) = 12.86$, $p < .005$. Replicating the findings of Experiment 1, explicit attitudes were more positive in the control condition ($M = 0.48$, $SD = 0.80$) than in the 20 CA condition ($M = -0.51$, $SD = 0.77$), $F(1, 27) = 11.57$, $p < .005$. In contrast, implicit attitude data did not show an effect of counterattitudinal condition ($F < 1$).

The effect of counterattitudinal condition for social contact judgments was also examined with a one-way ANOVA. This analysis showed, as expected, that people reported wanting more social contact when they were in the control condition ($M = 74.53$, $SD = 15.83$) than when they were in the 20 CA condition ($M = 61.21$, $SD = 17.71$), $F(1, 27) = 4.57$, $p < .05$. Also, there was no effect of counterattitudinal condition on seating distance ($F < 1$). Thus, the counterattitudinal condition manipulation affected deliberate behavior (i.e., desire for social contact) but not the subtle behavior (i.e., seating distance).

To examine the main hypotheses, the correlation between explicit attitudes, implicit attitudes, deliberate behavior (i.e., desire for social contact), and subtle behavior (i.e., seating distance) were calculated. As expected, more positive explicit attitudes were related to greater desire for social contact ($r = .71$, $p < .001$) but were unrelated to seating distance ($r = .04$, *ns*). It is important that more positive implicit attitudes were unrelated to desire for social contact ($r = -.03$, *ns*) but were significantly related to closer seating distance ($r = -.41$, $p < .03$). Moreover, two multiple regressions analyses were conducted in which explicit and implicit attitudes served to predict desire for social contact (first analysis) and seating distance (second analysis). As predicted, explicit attitudes ($\beta = 0.70$, $p < .001$) but not implicit attitudes ($\beta = -0.01$, *ns*) predicted desire for social contact. On the other hand, implicit attitudes ($\beta = -0.41$, $p < .04$) but not explicit attitudes ($\beta = 0.02$, *ns*) predicted seating distance. Thus, explicit attitudes uniquely predicted deliberate judgments and implicit attitudes uniquely predicted subtle, spontaneous behaviors.

Discussion

Experiment 4 showed that the differential formation and change of implicit and explicit attitudes demonstrated in Experiments 1–3 have important implications for predicting behavior toward an attitude object, which in turn, reflect different systems of reason-