# Propositional Processes of Attribution (Sometimes) Influence Implicit Attitudes

Kathleen Schmidt

Charlottesville, Virginia, USA

Bachelors of Arts, Psychology, Swarthmore College, 2007
Masters of Arts, Psychology, University of Virginia, 2011

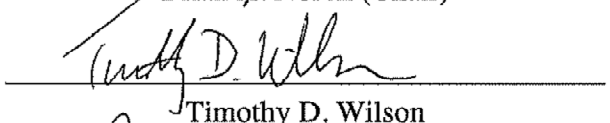
A Dissertation Presented to the
Graduate Faculty of the University of Virginia
in Candidacy for the Degree Of Doctor of Philosophy

Department of Psychology

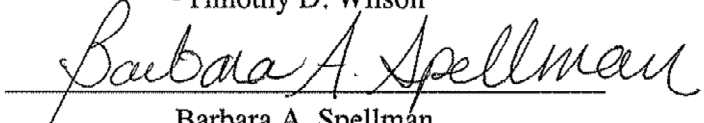
University of Virginia

July 24, 2014

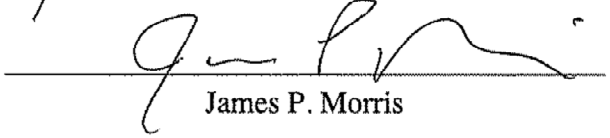
Brian A. Nosek (Chair)



Timothy D. Wilson



Barbara A. Spellman



James P. Morris

Abstract

Mainstream theoretical perspectives disagree on whether propositional processes influence implicit attitude formation and change. This dissertation addresses these opposing views by testing whether propositional knowledge in the form of attributions about the cause of negative or positive events impacts implicit attitudes. Overall, evidence supported the perspective that attributional processes influence implicit attitude formation. In Studies 1 through 4, participants were given the same information with different causal relationships to suggest either internal or external attributions of positive or negative events. In Studies 1, 2a, and 4, suggesting external attributions led to weaker explicit and implicit attitudes than suggesting internal attributions. In some studies, preconditions for testing the effect of attributions were not met: explicit attitudes were not influenced by the attribution manipulation (3 of 4 conditions in Study 4), or the implicit measure was insensitive to the manipulation (Study 3). I also observed mixed evidence regarding the impact of attributional processes on implicit attitude change. Studies 2a, 2b, and 5 measured attitudes before and after attribution manipulations for positive and negative events. Implicit attitudes varied by attribution condition after the manipulation in Study 2a and changed for negative events in Study 5 such that external attributions of negative events decreased implicit negativity. When no implicit preference was created initially (positive scenario conditions in Study 5) or initial preferences were not in the expected direction (Study 2b), external attributions did not change implicit attitudes. Across studies, the manipulations that impacted implicit attitude formation and change were emotionally evocative and conspicuous. Placed in a larger context of implicit attitude theory, this dissertation suggests that propositional processes do influence implicit attitudes under conditions where propositional knowledge drastically changes the meaning of the corresponding associations.

Acknowledgments

Two and a half years ago, the feats of writing a dissertation or completing my degree seemed impossible. Only with the patience and support of my friends, family, and colleagues have I been able to overcome personal and professional obstacles to realize these once out of reach goals.

I thank the current and past members of the Implicit Social Cognition Lab for their support and feedback along the way. Your growth as people and scientists has been an inspiration. To the other graduate students in the social area and other research labs of which I have been a member, thanks for sharing with me your scholarship and demonstrating what wide varieties of personality and study belong in my chosen field. To Carlee Hawkins, thank you for putting up with my frequent questions and requests for help over the last 6 years. And thanks to Calvin Lai for help with the original research idea for this dissertation.

To my dissertation committee – Brian Nosek, Tim Wilson, Bobbie Spellman, and Jamie Morris – thank you for your time, encouragement, critical feedback, and trust. This project has evolved from a simple question about sympathy to a theoretical exploration of cognitive processes. Thank you to Tim Wilson, particularly, for being an advisor and mentor over the years; you are truly brilliant. And to the entire faculty in the Social Area, thank you for your kindness, and the support and feedback you have given me on the many projects I have pursued in graduate school.

A special thanks to Brian Nosek, who has guided me as my advisor through the continual ups and downs of graduate school. All that you accomplish and give to the field will forever astonish me. You have taught me to value high-quality science and given me

the tools to conduct it. I am proud to be your student. To Project Implicit officers and staff, past and present, thanks for giving me the power (literally) to conduct the type of research I pursue. You have provided to me the skills and guidance that made this project achievable. And to the thousands of people who volunteer on Project Implicit every week, thank you for participating and making my research possible.

To my students, past, present, and future, thanks for giving me the determination to move forward in my career. To my research assistants over the years, thanks for putting up with me and helping me achieve my goals. Much gratitude to Nicholas Howald for his work on this and other projects. Nick, your intelligence and humor never fail to surprise me. I hope all your dreams come true.

To Sam Tornello, Jen Maiorano, and Lian LaRussa, I cannot express how meaningful your friendships have been to me. Thank you for filling my time in Charlottesville with love and joy. To Laura Shea and my “friends from around town”, I am so grateful for your support. Thanks for showing me another way.

To my family, I would be lost without you. To my parents, you have been the best models of the balance of scholarship and life one can have as academics. Thank you for talking about psychology at the dinner table my whole life. To my dad, thanks for being a badass psychologist, athlete, and person. Our jocular yet genuine conversations about teaching and research have made this process more enjoyable. To my mom, I could not have done any of this without your unlimited support. Thanks for the frequent pep talks, assistance, and invaluable guidance in work and life. To my sister, you are such an amazing and talented woman. Thank you for being a friend, confidant, and exemplar of a

young professional. And to all three of you, thanks for never giving up on me. You have made me who I am and rescued me from hopeless depths.

Finally, to my dog Anya, who stares up at me from her place at my feet. You may be an illiterate a tornado of fur and filth, but you are my family. Thanks for keeping me healthy, happy, and warm.

## Table of Contents

# Propositional Processes of Attribution (Sometimes) Influence Implicit Attitudes

Imagine you meet a woman named Janet for the first time. In meeting Janet, you learned two pieces of information about her: 1) Janet is a prostitute, and 2) Janet is unable to feed her children. Both of these pieces of information are negative. However, the causal relations between these pieces of negative information may affect your evaluation of Janet. If you learn that Janet is unable to feed her children because she is a prostitute, you may evaluate her very negatively. But, if you learn that Janet is a prostitute because she is unable to feed her kids, you may evaluate her more positively because the causal order suggests a different interpretation of the negative events that could create sympathy for Janet.

Attribution is the process of explaining the causes of events, specifically human behavior. Social psychology is not lacking theories describing this process (e.g., Bem, 1972; Jones & Davis, 1965; Kelley, 1967; Trope, 1986; Weiner et al., 1972). Common among most of these theories is the distinction between attributing behavior to the person and attributing behavior to the situation. External attributions are causal explanations for events that focus on the situation or the circumstances as the source. In contrast, Internal attributions are causal explanations that focus on the person and their individual characteristics, holding him or her responsible for the events.

Returning to the fact that the bits of information about Janet are all negative, it is only their causal relationship and the attribution resulting from it that differ between the two scenarios. Learning that Janet is a prostitute because she cannot feed her kids supports an external attribution—her choice of job comes from her poor circumstances; Learning that Janet cannot feed her kids because she is a prostitute supports an internal

attribution—her poor circumstances have resulted from her choice of job. Attribution factors into our appraisals of people, informing conscious evaluation to influence our explicit attitudes– that is how we consciously feel and report attitudes about others. Given the causal order that suggests an external attribution, we will feel sympathy for her and like her more—at least explicitly. However, explicit consideration is not the only mode of evaluation.

Measurement and theoretical innovation has provided evidence and new understanding that attitudes can operate outside of awareness and without endorsement (Greenwald & Banaji, 1995; Banaji, Nosek, & Greenwald, 2004). Such advances have led to a distinction between implicit attitudes, automatic or unconscious evaluations, and explicit attitudes, evaluations that individuals are able and willing to report. The current understanding of explicit and implicit attitudes is that they are distinct but related constructs that uniquely predict behavior (Nosek & Smyth, 2007; Greenwald et al., 2009).

The attribution of events falls into the category of propositional knowledge—defined here as the meaning of a statement beyond the individual words that adds or subtracts truth value (Gawronski and Bodenhausen, 2006). The causal relationship between the two statements about Janet is propositional knowledge. The purpose of this dissertation is to examine whether propositional knowledge in the form of causal attribution can influence the formation of implicit and explicit attitudes or change existing attitudes. The predictions for explicit evaluation are clear and follow directly from established literature reviewed next. The predictions for implicit evaluation are less clear, with two theoretical perspectives seeming to make opposing predictions.

The present research establishes a confrontational test of these theoretical perspectives: 1) Implicit attitudes are shaped by associative learning and are not influenced by propositional knowledge such as attribution; or 2) Implicit attitudes are sensitive to the meaning of causal relations between associations – external attributions affect the formation or expression of implicit attitudes, attenuating the strength of the attitude. After reviewing the literature on attribution and propositional knowledge as related to explicit evaluation, I present the competing perspectives for implicit evaluation.

**Explicit Attitudes and Attribution**

The processes related to attribution directly impact explicit attitudes. These processes include determining the cause of an event, ascribing relevant traits to the person behaving, and emotional reactions to the causal interpretation. Attribution theories provide dissimilar accounts for how observers make attributions. These theories have focused on how behavior varies over time (Kelley, 1967), the actor's intentions and the observer's reactions (Jones and Davis, 1967), or lay theories of psychology (Malle, 2011). Regardless of what information people use in making attributions, the resulting judgments of causality have distinct implications for evaluation.

Attributions for a given event may elicit dispositional judgments that in turn influence evaluation. Internal attributions can result from inferential processes that lead perceivers to ignore the situation and hold the actor responsible for his or her behavior. This bias, known as the Fundamental Attribution Error or correspondence bias, often leads the observer to ascribe traits to an actor based on their behavior (Ross, 1977; Gilbert and Malone, 1995; see Gilbert, 1998, for review). The traits ascribed to an actor based on his or her behavior are likely valenced and therefore influence attitudes.

**Sympathy**. The accounts that people construct to explain behaviors also have emotional consequences. Weiner (2001) discusses the attributional processes that follow the causal inference and influence emotion and evaluation. Attributions of achievement or moral responsibility may influence emotional responses that in turn lead to more positive or negative evaluations. Responsibility for a positive event evokes positive evaluations and admiration while blame for a negative event evokes negative evaluations and antipathy.

In contrast, external attributions of negative events cause sympathy for the misfortunate person (Weiner, 2001). Sympathy is feeling compassion in response to another's hardship (Batson et al., 1995). The negativity of the situation may be attached to the target of sympathy leading to negative evaluations of the target; however, feelings of caring for that person could also yield positive evaluations. Sympathy leads to valuing the welfare of a person, especially if that person is not responsible for their misfortune (Weiner, Perry, & Magnusson, 1988), and negative affect attributed to the suffering of others has been shown to increase positive attitudes towards those individuals (Batson, et al. 1995).

Both external attributions and sympathy for others can increase positive explicit evaluation of those people. Models of stereotypes predict correlations between sympathy and liking in high warmth and lower competence individuals (Fiske et al., 2002; Fiske, Cuddy, and Glick, 2007). Perspective taking interventions allow for individuals to make more external attributions and create sympathy for an individual or a group, leading to an increase in explicit attitude positivity (Dovidio et al., 2010; Vescio et al., 2003; Clore and

Jeffrey, 1972). Further, external explanations for the outgroup characteristics foster pro-social responses (Gill & Andreychik, 2007; 2009).

**Explicit Attitudes and Propositional Knowledge**

Whether this discussion of attribution focuses on causal attribution processes, dispositional inferences, or emotional consequences of attribution, one conclusion can be made: all of these processes fall into the category of propositional processes that influence the truth value about the associations with the behaviors being attributed. Explicit attitudes are those evaluations formed from propositional processing and based on inferences made from propositional knowledge relevant to the judgment (Gawronski and Bodenhausen, 2006). Given sufficient processing, explicit attitudes will be influenced by propositional knowledge such as attribution.

Consider Janet from the opening. When you learn she cannot feed her kids because she is a prostitute her circumstances are easily attributed as having an internal source. When you learn that Janet is a prostitute because she cannot feed her kids her circumstances are more easily attributed as having an external source. Positive or negative evaluations of Janet could be based on the actual attributions of causality (she was led to prostitution by her bad situation vs. chooses to be a prostitute), trait inferences made from the causal order (she is caring vs. negligent), or emotional responses (sympathy vs. antipathy) to the different causal orders presented. The mechanism by which the resulting explicit attitudes are more positive may imply different processes, but those are propositional processes. Propositional cognitive processing incorporates the relationship between the negative associations to impact explicit attitude formation and change (Gawronski and Bodenhausen, 2006).

**Implicit Attitudes and Propositional Knowledge**

In contrast to the more reflective evaluations that determine explicit attitudes, experiences can also influence judgments and behaviors without awareness on the part of individuals (Greenwald & Banaji, 1995). Implicit attitudes consist of evaluative associations that individuals are unable or unwilling to report. Within a cognitive network of representations, attitude objects, such as Janet, are linked to evaluative concepts, such as "bad". These connections that are built from experience may exist simultaneously with conflicting connections between concepts (Janet and "good"). These associations can be activated automatically in response to relevant stimuli, perhaps without the awareness of their activation, and shape subsequence perception, judgment, or action (see Nosek, Hawkins, & Frazier, 2012, for review).

Consider again the two causal statements from the opening – Janet is a prostitute because she is unable to feed her kids or Janet is Janet is unable to feed her children because she is a prostitute. The associative information for these two statements is the same: Janet is a prostitute and Janet cannot feed her kids. The only way that they differ is in the causal relation between the associative information. Several models of attitudes assert that implicit attitudes consist of purely associative information (e.g., Olson & Fazio, 2009; Shanks, 2007). Propositional knowledge, the causal relations, would not influence implicit attitudes in this case. Another possibility is that propositional knowledge may alter implicit attitude formation and change (e.g., Gawronski & Bodenhausen, 2006; De Houwer, 2009). So, from the association models of implicit attitude formation – these two conditions ought to produce the same negative implicit evaluation of Janet. But from models that allow propositional influences on implicit

attitudes, the causal explanation that Janet is a prostitute because she cannot feed her children may create less implicit negativity than its alternative.

In line with the former perspective that implicit attitudes are purely associative, Shanks (2007) argues that association formation processes typically exist without awareness and are unaffected by propositional information. Contingency judgment research supports the view that implicit attitudes are formed without the influence of deliberative processes (see Shanks, 2007, for review). Awareness of the relationship between the two concepts is unnecessary in attitude formation (e.g., Baeyens, et al., 1992, Baeyens, Eelen, and Van de Bergh, 1990). Other researchers suggest that evaluative conditioning creates implicit attitudes without regard to how the pairing is evaluated explicitly (Olson & Fazio, 2001) and that implicit learning does not occur along non-evaluative dimensions (Olson, Kendrick, & Fazio, 2009). Gregg, Seibt, and Banaji (2006) found that informing people that an attitude induction provided incorrect information led to changes in explicit but not implicit attitudes. Participants were not able to use their propositional knowledge that the induction was false to alter their already formed implicit attitudes. Similarly, Gawronski and Strack (2004) found that writing a counterattitudinal essay under conditions cognitive dissonance changed explicit but not implicit attitudes. They interpreted this as indicating that inferential processes in dissonance reduction were ineffective at altering implicit attitudes.

The perspective that propositional processing can affect implicit attitudes suggests that associative learning may sometimes depend on the more deliberate formation and evaluation of propositions. For instance, some research suggests that propositions about the accuracy of associations appear to influence the effect of evaluative conditioning on

implicit attitude formation (De Houwer, 2006; 2007; 2009). Propositional information regarding existing attitudes may affect positive associations or suppress negative associations (Gawronski and Bodenhausen, 2007). For example, Petty et al. (2006) induced a positive or negative attitude towards a target through evaluative conditioning and then told participants the information about the target was false. They found that disconfirming the information about the target, changed implicit attitudes but not as much as it changed explicit attitudes. Smith, De Houwer, and Nosek (2013) tested the effect of source credibility on persuasion finding evidence that implicit (and explicit) attitudes were malleable to information from a highly credible source but not one low in credibility. Accuracy information, falsification, and source credibility can all be classified as propositional knowledge that influences the truth-value of the information as propositional knowledge. Similarly, interventions creating situational attributions have been shown to reduce automatic stereotyping (Stewart et al., 2010). These studies provide further evidence that propositional processes influence implicit attitudes. Finally, perhaps the implicit/explicit distinction is irrelevant to the influence of propositional processes on attitudes; nearly any process that affects explicit attitudes can affect implicit attitudes (Hassin, 2013).

**Sympathy.** The two causal orders in the opening may also created different emotional reactions because of the attributional process. If external attributions for negative events create sympathy for the target, then the influence of sympathy on implicit negativity is relevant to this debate.

Some research suggests that sympathy can co-occur with or even cause negative implicit attitudes (e.g., Frantz, et al., 2004; Uhlmann, Brescoll, & Paluck, 2006;

Andreychik & Gill 2012). The negative emotional underpinnings of implicit attitudes towards suffering people could be empathy or anxiety. For instance, one attitude-conditioning finding implies that associations between groups and oppression may facilitate negative associative responding on implicit tasks (Uhlmann, Brescoll, & Paluck, 2006). Other studies suggest that as implicit bias against low status groups increases, positive explicit attitudes towards that group also increases for those individuals who cite external causes for the low social status because of sympathy (Andreychik & Gill, 2012). Empathic responding to the unfortunate could just build up more negative associations that affect implicit measures differently than explicit measures. Because sympathy is only creating more negative associations in this case and not changing the meaning of the associations, these studies support the perspective that propositional processes do not impact implicit attitudes.

In contrast, some research suggests the affective nature of experiences is central to the formation and change of implicit attitudes (Rudman, 2004; Castelli, et al., 2010). The emotions elicited by attributional processing (e.g., sympathy) could alter an experience by shifting how negative affect influences association formation. Allowing for external attributions and increasing sympathy towards outgroups with perspective taking interventions may also decrease implicit bias (Galinksy & Moskowitz, 2000). Sympathy created from attributional processes may also cause a mood interference with expression of negativity on implicit measures (Huntsinger, Sinclair & Clore, 2009). Seen from the perspective that propositional knowledge can influence implicit attitudes, propositional processes likely drove the emotional responses that impacted implicit attitudes in theses studies.

## The Present Research

This dissertation will address whether propositional knowledge in the form of attribution can influence implicit attitude formation or change. Using novel attitude targets and scenarios like Janet the prostitute who cannot feed her kids, I compared the influence of negative information on attitude formation when suggesting external as opposed to internal attributions (Study 1). Then, in two studies, I examined how the same attribution manipulations influence negative explicit and implicit attitudes that have already been established (Study 2a and 2b). Next, I attempted to measure the bad and good associations involved in implicit attitude formation separately to determine if external attributions creating sympathy add good associations or suppress bad associations to influence evaluations (Study 3). An additional study examined the influence of causal order on the formation of positive and negative attitudes (Study 4). Finally, I gave external attributions during and after positive and negative attitude formation in a final study to further examine the influence of propositional knowledge in the form of attribution on implicit attitude formation and change (Study 5).

# Study 1

## Study Overview

The purpose of Study 1 was to establish whether propositional knowledge in the form of attribution decreases the impact of negative information on the formation of implicit and explicit attitudes. This study presented information about two unknown individuals, either Henry and Nathan or Janet and Alice. Participants read information about both individuals, one of whom, the target person, had negative characteristics. Two

conditions presented a single piece of negative information about the target (e.g., Janet is a prostitute or Janet cannot feed her kids). In the other two conditions, both pieces of information were presented in a causal relationship. One form of the causal relationship implies an internal attribution that evokes antipathy – Janet cannot feed her kids because she is a prostitute. The other form of the relationship implies an external attribution that evokes sympathy – Janet is a prostitute because she cannot feed her kids. By comparing participant reactions to the information differences by condition, I hope to uncover if attributions that create sympathy specifically, and propositional knowledge generally, can decrease implicit negativity.

The Implicit Association Test (IAT; Greenwald et al, 1998) and several explicit measures of attitudes, sympathy, and liking were used to assess participants' impressions of the two individuals. The IAT is a standard measure of implicit attitudes that has been used to measure conditioned associations. In previous studies on novel attitudes, the IAT appears sensitive to induction manipulations such similar the one employed in the present study (e.g., Ranganath and Nosek, 2008).

I predicted that explicit attitudes toward the negative target would be more positive in the causal order condition that suggests internal attributions than in the condition that suggests external attributions and the individual negative information conditions. The negativity of the information presented in the internal attribution condition will be decreased by feelings of sympathy for the target brought on by the causal relation between the negative characteristics.

Based on existing theory, competing predictions for implicit attitude formation exist. From one point of view (Shanks, 2007; Olsen & Fazio, 2009), implicit attitudes

should be sensitive only to the associations formed by each piece of negative information. If so, then the both causal order conditions should have the same effect as they are composed of the same negative information. From another point of view (Gawronski & Bodenhausen, 2006; De Houwer, 2009) implicit attitudes may be sensitive to the causal relations between the associative information. If so, the internal attribution condition should create less negative implicit attitudes than the external attribution condition. In parallel to explicit attitudes, the presence of more negative information in the external attribution condition will be reflected with decreased positivity in the that condition in comparison to the negative information conditions.

**Method**

**Participants.** Participants were visitors to the Project Implicit Website (https://implicit.harvard.edu). They were randomly assigned to participate in the study, one of approximately 10 running in the research pool at the same time. For more information about the virtual laboratory run by Project implicit, see Nosek (2005) and Nosek et al. (2007). Of the 3030 participants provided informed consent, 2004 (66%) completed the study fully. Analyses include all participants who completed at least one of the measures (N = 2362).

Women comprised 65% of the participants who consented. The ethnic composition was 9% Hispanic and the racial composition was 75% White, 6% Black, and 7% Asian. Participant age ranged from 12 to 91 ($M$=36.2, $SD$=13.9). Demographic information did not predict participant dropout except for race, $F(8, 2654) = 2.52$, $p = 0.010$, $\eta^2 = 0.008$. Simple effects indicate that Asian participants were less likely to finish the study than White participants.

**Design.** The study employed a 2 (scenario set) x 4 (negative information condition) between-subjects design. Target identity, implicit and explicit measure order, the order of the explicit measures was randomized.

**Materials.** I created two sets of scenarios, one describing two women and the other describing two men. These scenario sets contained separate descriptions of the individuals who were each represented by a cartoon drawing (see Appendix A). The descriptions were adapted from scenarios that Ratliff and Nosek (2008) pre-tested and used as evalutively neutral filler information:

This is Nathan [Alice]. Nathan [Alice] was born in Gary, Indiana. Nathan [Janet] attended Lake Ridge Middle School. Nathan enjoys cooking, hiking, and listening to music. Nathan [Janet] is warm and considerate, but tends to be slightly dishonest at times. Nathan [Alice] is a social worker.

This is Henry [Janet]. Henry [Janet] was born in Dayton, Ohio. Henry [Janet] attended Oakwood Junior High School. Henry [Janet] enjoys gardening, biking, and playing card games. Henry [Janet] is kind and thoughtful, but tends to be slightly greedy at times. Henry [Janet] is a social worker.

The target description included additional negative information instead of the final sentence. The Negative 1 and Negative 2 conditions had single pieces of negative information. Presenting these two pieces of negative information together with different causal relations created the internal and external attribution order conditions. For the male scenarios, the following sentences defined the four conditions: Negative 1: “One fact about Nathan is of particular concern to others: Nathan has cancer.”, Negative 2: “One fact about Nathan is of particular concern to others: Nathan uses illegal drugs.”, Internal: “One fact about Nathan is of particular concern to others: Nathan has cancer

because Nathan uses illegal drugs", and External: "One fact about Nathan is of particular concern to others: Nathan uses illegal drugs because Nathan has cancer."

For the female scenarios, one of these statements was provided to correspond with each condition: Negative 1: "One fact about Janet is of particular concern to others: Janet is unable to feed her kids."; Negative 2: "One fact about Janet is of particular concern to others: Janet is a prostitute."; Internal: "One fact about Janet is of particular concern to others: Janet is unable to feed her kids because Janet is a prostitute."; or External: "One fact about Janet is of particular concern to others: Janet is a prostitute because Janet is unable to feed her kids.". The four conditions were pre-tested and the causal order conditions produced similar negativity ratings, but the condition designed to suggest external attributions was rated as more sympathetic than the other three.

**Measures**

***IAT.*** The Implicit Association Test (IAT; Greenwald et al., 1998) measures the strength of associations between two concepts (e.g., Janet and Alice) and two attributes (e.g., good and bad). This computer task requires participants to categorize the pictures or words that appear in the middle screen as belonging to categories represented on the left or right side of the screen using the "I" (for right) or "E" (for left) keys. The IAT is an indirect measure that includes multiple trials with different categorization instructions. In one condition, participants press the "I" key when words from one category (in this case a person like Janet) and "good" or pleasant words appear on the screen and the "E" key when the other category (e.g., Alice) or "bad" or unpleasant words appear on the screen. In a second condition, participants press the "I" key for pleasant words and the alternative category (e.g., Alice) and the "E" key for unpleasant words and the first category (e.g.,

Janet). If Janet is more strongly associated with good than Alice is, then categorizing good with Janet on one key and bad with Alice on the other should be faster than categorizing bad with Janet and good with Alice. This would be interpreted as an implicit preference for Janet over Alice.

The IATs consisted of seven blocks of trials, following the recommendation of Nosek, Greenwald, and Banaji (2005). I analyzed the response latencies using the *D* algorithm, timing participants from the trial start to the correct response (Greenwald, Nosek, & Banaji, 2003). Latencies less than 400 ms and greater than 10000 ms were removed from the calculations. Participants with greater than 30% error rates overall or 40% in one block (3% of sample) were excluded. Positive *D* scores indicate implicit preference for the target person compared to the control person.

***Self-Report.*** Explicit attitudes towards the target and control individuals were measured on individual 7-point scales from "Dislike Very Much" to "Like Very Much", with "Neither Like nor Dislike" as the midpoint. Sympathy was measured on a 5-point scale ranging from "Not at all Sympathetic" to "Extremely Sympathetic"

***Manipulation Boost/Check***. After the attitude induction, but before the dependent measures, participants answered three questions about the information they read. They were asked to identify one of the individuals by selecting from a list of names and recognize where one individual grew up on a list of cities The final question asked about the information specific to their experimental condition (e.g., Which woman is a prostitute because she cannot feed her kids?). Excluding participants who answered any

or all of these questions incorrectly did not substantively change the results.[1] All participants were included in analyses regardless of manipulation check responses.

**Procedure.** After consenting to participate in a study on the Project Implicit Research website, participants were randomly assigned to one of two sets of scenarios: a version about cancer and drugs with a male target and control individual (Nathan and Henry) or a version about prostitution and starving children a female target and control individual (Janet and Alice). The identity of the target was randomized.[2] Participants first read about the control individual and then read about the target individual. The target description included the negative information for one of the four conditions (Negative 1, Negative 2, External attribution, or Internal attribution).

Participants then answered three questions about the information they just read. These questions were designed to both strengthen the manipulation and served as an attention check. Next, Participants completed the implicit and explicit attitude measures about the target and control. The order of the implicit and explicit measures was

---

[1] Thirty-eight percent of participants missed one or more manipulation check questions. A model predicting implicit attitudes from failure on the key check question, condition, and the interaction between failure and condition showed no effects of check failure, $F(1, 1924) = 1.76$, $p = 0.185$, $\eta^2 = 0.001$, and no interaction, $F(3, 1924) = 0.16$, $p = 0.924$, $\eta^2 = 0.0002$. The same was true for explicit attitudes predictions; check failure $F(1, 1645) = 0.79$, $p = 0.373$, $\eta^2 = 0.001$ and it's interaction with condition. $F(3, 1645) = 0.48$, $p = 0.698$, $\eta^2 = 0.001$, were not predictive.

[2] I found a main effect of the stimuli counter-balancing on implicit attitudes: Janet and Henry (the light haired figures) were preferred overall. Including target identity and it's interaction with condition in a model predicting implicit preference from condition demonstrated a large effect of participant identity, $F(1, 1924) = 200.88$, $p < 0.0001$, $\eta^2 = 0.094$, but no interaction, $F(3, 1924) = 0.58$, $p = 0.628$, $\eta^2 = 0.001$. This effect has been found in previous use of these drawings in implicit attitude research (see Ratliff and Nosek, 2008) and did not vary in magnitude based on condition. Target identity, $F(1, 1645) = 3.51$, $p = 0.0611$, $\eta^2 = 0.002$ were and it's interaction with condition, $F(3, 1645) = 0.74$, $p = 0.527$, $\eta^2 = 0.001$, not significantly predictive of explicit attitudes

randomized and did not affect the results.[3] Finally, participants were debriefed and given their IAT results.

**Results**

**Explicit Attitudes**. First, I tested to see if the causal order implying an external attribution influenced the formation of explicit attitudes by decreasing negativity towards the target. Explicit preference was calculated by subtracting liking for the control person from liking for the target. One-way analysis of variance indicated that difference in liking varied by condition, $F(3, 1649) = 3.31$, $p = 0.019$, $r^2 = 0.006$. Inspection of the individual means with dependent means *t*-tests showed that the target person was not evaluated more negatively than the control person with either individual piece of negative information (Negative 1: $M = 0.02$, $SD = 0.73$, $t(414) = 0.41$, $p = 0.685$, $d = 0.02$; Negative 2: $M = -0.07$, $SD = 0.70$, $t(406) = 1.85$, $p = 0.065$, $d = 0.09$). When those two negative pieces of information were combined into a causal relationship that was likely to elicit an internal attribution, participants showed a slight, but non-significant preference, for the control to the target ($M = -0.06$, $SD=0.65$; $t(399) = 1.70$, $p = 0.090$, $d = 0.09$). However, when the two negative pieces of information were combined in an internal attribution and sympathy eliciting causal relationship, participants preferred the target to the control explicitly ($M = 0.06$, $SD = 0.63$; $t(490) = 2.07$, $p = 0.039$, $d = 0.09$). Simple effects tests showed that the internal attribution order produced a stronger preference for

[3] In a model predicting implicit attitudes, the order of the measures and its interaction with condition were non-significant predictors of implicit attitudes, $F(1, 1924) = 0.19$, $p = 0.661$, $\eta^2 = 0.0001$; $F(3, 1924) = 1.54$, $p = 0.203$, $\eta^2 = 0.002$. A model predicting explicit attitudes also did not indicate significant effects of dependent measure order, $F(1, 1645) = 0.02$, $p = 0.889$, $\eta^2 = 0.0000$, or the interaction between measure order and attribution condition $F(3, 1645) = 0.99$ $p = 0.395$, $\eta^2 = 0.002$.

the control over the target person compared to the external attribution order, $t(829) = 2.66$, $p = 0.008$, $d = 0.18$.

A model testing for main effects and interactions between scenario set (the cancer/drugs scenarios or prostitute/kids scenarios) and condition accounted for 2% of the variance in predicting explicit attitudes, $F(7, 1645) = 3.95$, $p = 0.0003$, $r^2 = 0.017$. Results indicate a main effect of condition $F(3, 1645) = 3.43$, $p = 0.017$, $\eta^2 = 0.006$, but not of scenario set $F(1, 1645) = 0.18$, $p = 0.671$, $\eta^2 = 0.0001$, and a significant interaction between condition and scenario set $F(3, 1645) = 5.80$, $p = 0.0006$, $\eta^2 = 0.010$. Investigation of the interaction effect showed that the Negative 1 and Negative 2 conditions created different patterns of explicit attitudes based upon scenario information. The Negative 1 information that the target cannot feed her kids created more negative explicit attitudes than the information that the target has cancer $t(413) = 2.43$, $p = 0.015$, $d = 0.24$, but the Negative 2 information that the target is a prostitutes created less negative explicit attitudes than the information that the target uses illegal drugs $t(405) = 2.86$, $p = 0.028$, $d = 0.28$. See Table 1 for means by condition.

**Implicit Attitudes**. Overall, implicit and explicit attitudes were only slightly correlated, $r = 0.07$, $p = 0.004$. I tested whether implicit preferences between the target and control person differed between the four attitude formation conditions. A one-way analysis of variance yielded significantly different mean IAT scores, $F(3, 1928) = 4.28$, $p = 0.005$, $r^2 = 0.007$. Inspection of the individual means showed that the target person was negatively evaluated compared to the control person in the conditions with a single piece of negative information (for Negative 1: $M = -0.04$, $SD = 0.41$, $t(472) = 2.20$, $p = 0.028$, $d = 0.10$; but not for Negative 2: $M = -0.03$, $SD = 0.42$, $t(494) = 1.57$, $p = 0.118$, $d = 0.07$)

and when those two negative pieces of information were combined into a causal relationship that was likely to elicit antipathy ($M$ = -0.09, $SD$ = 0.41; $t$ (490) = 4.82, $p$ < 0.0001, $d$ = 0.22). However, when the two negative pieces of information were combined in a sympathy eliciting causal relationship, participants showed equal liking for target and control ($M$ = 0.01, $SD$ = 0.42), $t$(475) = 0.27, $p$ = 0.784, $d$ = 0.01. Simple effects tests indicated that the key difference between the Internal and External attribution conditions was significant, $t$(964) = 3.55, $p$ = 0.0004, $d$ = 0.22 with stronger implicit preferences for the target in the External versus Internal attribution conditions. See Table 1 for condition means and standard deviations.

A model predicted implicit attitudes with main effects and interactions between scenario set and condition, $F$(7, 1924) = 3.51, $p$ = 0.0009, $r^2$ = 0.013. Results indicate a main effect of scenario set $F$(1, 1924) = 5.97, $p$ = 0.015, $\eta^2$ = 0.003, and condition $F$(3, 1924) = 4.35, $p$ = 0.005, $\eta^2$ =0.007, but no interaction $F$(3, 1924) = 1.85, $p$ = 0.136, $\eta^2$ = 0.003. Means inspection reveals that the male cancer scenario set produced more positive implicit attitudes towards the target ($M$ = 0.06, $SD$ = 0.42) than the female prostitution set ($M$ = 0.02, $SD$ = 0.41); see Table 1 for means and standard deviations by scenario.

**Sympathy.** Finally, I tested to see how explicit measures of sympathy were related to explicit and implicit attitudes and the condition manipulations and if they mediated condition effects. Sympathy differences were calculated by subtracting sympathy for the control from sympathy for the target. Condition significantly predicted the difference in sympathy between the target and control individuals, $F($3, 2124) = 22.68, $p$ < 0.0001, $r^2$ = 0.031**.** Simple effects testing indicated that the difference in sympathy between the target and the control in the External attribution condition ($M$ =

0.83, *SD* = 1.34) was significantly less than the difference in the Internal attribution condition (*M* = 0.54, *SD* = 1.28), $t(1064) = 3.69$, $p = 0.0002$, $d = 0.23$.

Regression analyses were conducted for the subset of participants in the External and Internal attribution conditions to see how felt-sympathy influenced results. Condition significantly predicted the difference in sympathy between the target and control, ß = 0.30, $t(1064) = 4.40$, $p = 0.0002$, $\eta^2 = 0.01$, and the difference in sympathy predicted both the IAT score ß = 0.02, $t(953) = 2.24$, $p = 0.025$, $\eta^2 = 0.005$, and explicit attitudes ß = 0.08 $t(819) = 4.39$, $p < 0.0001$, $\eta^2 = 0.02$. Sympathy partially mediated the effect of condition on explicit attitudes (Sobel's Z = 3.11, $p = 0.002$) by decreasing the effect of condition on explicit attitudes from ß = 0.11, $t(829) = 2.66$, $p = 0.008$, $\eta^2 = 0.008$, to ß = 0.09, $t(818) = 2.22$, $p = 0.026$, $\eta^2 = 0.005$. However, explicit sympathy was not a significant mediator for the effect of condition on implicit attitudes, as adding sympathy into the model did not affect the relationship between condition and implicit attitudes.

**Discussion**

Study 1 provided initial evidence that attribution can influence the formation of both implicit and explicit attitudes. As hypothesized, explicit attitudes varied by condition. Participants rated the attitude target more positively then when the negative information was sympathetic than when it was not. This pattern was also found for implicit attitudes. The same information elicited more implicit positivity for the target person when the causal relation between the pieces of information elicited sympathy with external attributional processes compared to when it did not. These findings provide preliminary support for the perspective that propositional knowledge such as attribution can affect implicit attitudes. However, the effects – particularly for explicit attitude

formation – were rather weak, suggesting the need for additional investigation. In addition, Study 1 only examined attitude formation. Study 2a and 2b will investigate if propositional knowledge in the form of attribution influences both attitude formation and attitude change.

## Study 2a

### Study Overview

Study 1 provided initial support for the view that implicit attitude formation is sensitive to propositional knowledge – the causal relations between associations – beyond the meaning of the associations themselves. A causal relationship between negative information that created a sympathetic reaction evoked more implicit positivity than one that did not. For the first study, I gave the sympathetic interpretation the best chance to have an impact on implicit evaluations. The targets were novel, and the sympathetic interpretation was evident immediately on learning about the targets. However, I may have selected a special situation (i.e., attitude formation) for which the propositional knowledge could have an impact. In Study 2, I tested whether sympathy could have similar effects on implicit attitudes after they have been established.

Study 2 included a direct replication of the Internal and External attribution conditions for Study 1 for the prostitution scenario set. I also added two new conditions that establish attitudes and then try to change those attitudes with additional negative information framed with propositional knowledge. In these dual attitude induction conditions, participants read negative information about the target at Time 1 and then completed attitude measures. At Time 2, participants received additional negative explanatory information that should evoke more negativity (by adding an internal

attribution) or sympathy (by adding an external attribution). Then attitudes were measured again. For the replication conditions (single attitude induction), I predicted that the External attribution condition would elicit more implicit and explicit positivity than the Internal attribution condition. For the dual induction conditions, the predictions are less straightforward. If the effects of attribution on implicit attitudes are constrained to the initial formation, then implicit attitudes will not change from Time 1 to Time 2 in the External attribution condition. However, if attribution also influences implicit attitudes after formation, then implicit attitudes measured at Time 2 in the External attribution condition will differ from implicit attitudes formed at Time 1 or measured at Time 2 in the Internal attribution condition.

**Method**

**Participants.** Participants were volunteers on the Project Implicit Website (https://implicit.harvard.edu) randomly assigned to participate in this study, one of the several research studies concurrently running on the site. Of the 2543 participants who consented, 1642 (65%) completed the study. All participants who completed the relevant dependent measure were included in associated analyses (N = 1776).

Women composed 60% of the study sample. Participants were 12% Hispanic ethnically and 69% White, 10% Black, and 7% Asian racially. Eleven to 77 year-olds participated ($M$ = 30.7, $SD$ = 13.2). Study completion was largely unrelated to the demographics measured; however, participant drop out increased with age, $F(1, 1645) = 3.33$, $p = 0.02$, $\eta^2 = 0.002$.

**Design.** The design of the study was similar to a Solomon 4 group design. Two conditions gave single piece of negative information before Time 1, then additional

negative information at Time 2. The other two conditions gave all the information before any testing. These four conditions made up a 2 Condition (internal vs. external attribution) x 2 Attitude Induction (single vs. dual) between-subjects design. Unlike Study 1, Study 2a only used the prostitution scenario set. Target identity and measure order (between implicit and self-report and within self-report) were randomized.

**Materials**. Participants were given the following study explanation: "For this study, please imagine that you work at a community center. Because of this job, you encounter many new and diverse people in your everyday life. For instance, you recently met both Alice and Janet at the community center. Pay attention to what you learn about each of them on the next two pages; your memory will be tested." The scenario text was identical to that used in Study 1 for the prostitution scenario set. For the dual induction participants, the target's description ended with either "Janet is a prostitute" or "Janet is unable to feed her kids" (identical to the Negative 1 and Negative 2 condition information in Study 1). For the single induction participants, the Internal or External attribution condition information ended the target's description. Thus, all four negative information possibilities were used either for Time 1 (i.e., Negative 1 or Negative 2 conditions) or for Time 2 and the dual induction (Internal or External attribution conditions).

In the dual induction conditions, participants were given additional explanations after Time 1 measurement: "Now several weeks later you learn some more information about Alice and Janet. Pay close attention." The control was then pictured with this information, "As you may recall, this is Alice who is a social worker. Apparently, Alice also works part Time at a grocery store." On the following page the target was pictured with one of two captions: "As you may recall, this is Janet who is unable to feed her kids.

It turns out that Janet is unable to feed her kids because she is a prostitute" if participants were in the External attribution condition or "As you may recall, this is Janet who is a prostitute. It turns out that Janet is a prostitute because she is unable to feed her kids" if participants were in the Internal condition.

**Measures**

***IAT***. An IAT assessing preference between Alice and Janet was used to measure implicit attitudes. The IATs was shortened to 5 trial blocks to allow for same IAT to be complete twice in one 15-minute study session. To make a 5-block task, I removed Blocks 3 and 6 (each of which contained 20 practice trials) from the IAT in Study 1. Otherwise, the task was identical to that in Study 1. Three percent of IAT scores were excluded based on criteria (greater then 30% error rate overall or 40% error rate in one block).

***Self-Report.*** Study 2 used the same self-report items as Study 1 with one addition. Participants chose one of 7 statements that best describes them, ranging from "I strongly prefer Alice to Janet" to "I strongly prefer Janet to Alice" with "I don't have a preference between Alice and Janet" as the midpoint.

***Manipulation Boost/Check.*** Study 2 used the same items for checking and boosting the manipulation as Study 1 with one addition. After Time 2, Participants were asked to confirm the identity of that target based on the attribution condition manipulation (i.e., "Which woman is a prostitute because she cannot feed her kids?"). All

participants were included in analyses regardless of manipulation check results; exclusion of participants with incorrect answers did not affect research conclusions.[4]

**Procedure.** Participants provided consent and were randomly assigned to receive either a dual induction (2/3 of participants) or single induction of attitudes (1/3 of participants). Target identity was randomized.[5] Participants then read the introductory

---

[4] Thirty-three percent of participants who completed the study missed one or more manipulation check questions. For the dual induction participants, at Time 1, a model predicting implicit attitudes from failure on the key check question, condition, and the interaction between failure and condition showed no effects of check failure, $F(1, 1071) = 0.59$, $p = 0.441$, $\eta^2 = 0.001$, or the interaction, $F(1, 1071) = 2.45$, $p = 0.118$, $\eta^2 = 0.002$. The failure on the key question at Time 2, $F(1, 1006) = 0.17$, $p = 0.677$, $\eta^2 = 0.0002$, and its interaction with condition $F(1, 1006) = 1.43$, $p = 0.232$, $\eta^2 = 0.001$, did not predict implicit attitudes. Single induction participants showed no main effect or interaction with check failure for implicit attitudes, $F(1, 281) = 0.64$, $p = 0.432$, $\eta^2 = 0.002$, and $F(1, 281) = 0.05$, $p = 0.815$, $\eta^2 = 0.0002$, respectively. No effects of failure on the key question were found for explicit attitudes predictions at Time 1; analyses indicated no main effect: $F(1, 1184) = 0.87$, $p = 0.353$, $\eta^2 = 0.001$, or interaction: $F(1, 1184)=0.59$. $p = 0.441$, $\eta^2 = 0.001$. At Time 2, no main effect of failure was found, $F(1, 1063)=2.49$, $p = 0.115$, $\eta^2 = 0.002$, but condition did interact with check failure to predict explicit attitudes, $F(1, 1063) = 6.84$, $p = 0.009$, $\eta^2 = 0.006$. Likewise for single induction participants, the interaction between order and condition was significant for explicit attitudes, $F(1, 312) = 4.35$, $p = 0.038$, $\eta^2 = 0.013$, but no main effect of order was found, $F(1, 312) = 0.41$, $p = 0.523$, $\eta^2 = 0.001$. Mean comparisons indicate that answering the Time 2 check incorrectly impacted the Internal attribution condition more than the External attribution condition due to the extremity of the negative attitudes created by the Internal attribution condition.

[5] As seen in study 1, participants implicitly preferred the Janet to Alice. For the dual induction participants at Time 1, a model predicting implicit attitudes from target identity, condition, and the interaction between identity and condition showed a large effect of identity, $F(1, 1071) = 68.60$, $p < 0.0001$, $\eta^2 = 0.060$, but no the interaction, $F(1, 1071) = 0.37$, $p = 0.542$, $\eta^2 = 0.0003$. Target identity at Time 2 also predicted implicit attitudes, $F(1, 1006) = 28.84$, $p < 0.0001$, $\eta^2 = 0.028$, but its interaction with condition was non-significant, $F(1, 1006) = 1.86$, $p = 0.173$, $\eta^2 = 0.002$. Target identity did not impact explicit attitudes at Time 1; no main effect: $F(1, 1184) = 1.11$, $p = 0.293$, $\eta^2 = 0.001$, or interaction: $F(1,1184)=0.03$. $p = 0.854$, $\eta^2 = 0.0000$. At Time 2, I found no main effect of target, $F(1, 1063)=1.75$, $p = 0.186$, $\eta^2 = 0.001$, and condition did not interact with target to predict explicit attitudes, $F(1, 1063) = 0.35$, $p = 0.553$, $\eta^2 = 0.0003$.

explanation. After this point, procedure for the single induction was identical to Study 1. However, for the two new conditions in the dual attitude induction, I added a few additional procedural details. While all conditions were followed by three manipulation check questions and randomized implicit and explicit measures evaluating Janet and Alice, the dual induction participants continued on. Participants in these two conditions were given additional information about the control then target individuals. After this second attitude induction, participants completed the same implicit and explicit measures as before. The order of the implicit and explicit measures was randomized but kept constant between participants and did not affect the results.[6] After the study, regardless of condition, participants were debriefed and given their IAT results.

## Results

### Explicit attitudes

***Dual induction.*** If attribution influences explicit attitudes even after they have been established, then the difference in liking between the target and control should vary

[6] In a model predicting implicit attitudes at Time 1, the order of the measures and its interaction with Time 1 information were not significant, $F(1, 1071) = 0.11$, $p = 0.737$, $\eta^2 = 0.0001$, and $F(1, 1071) = 0.33$, $p = 0.568$, $\eta^2 = 0.0003$, respectively. In a model predicting implicit attitudes at Time 2, the order of the measures and its interaction with condition were also non-significant, $F(1, 1006) = 1.19$, $p = 0.276$, $\eta^2 = 0.001$, and $F(1, 1006) = 0.00$, $p = 0.960$, $\eta^2 < 0.0001$, respectively. For the single induction participants, order had no main effect on implicit attitudes, $F(1, 210) = 0.59$, $p = 0.445$, $\eta^2 = 0.003$, and the interaction between order and condition was non-significant, $F(1, 210) = 3.03$, $p = 0.083$, $\eta^2 = 0.014$. A model predicting explicit attitudes at Time 1 found no main effect of order, $F(1, 1182)=0.78$, $p = 0.376$, $\eta^2 = 0.001$, or its interaction with condition, $F(1, 1182) =1.70$, $p = 0.192$, $\eta^2 = 0.001$. At Time 2, a model explicit attitudes found no main effect of order, $F(1, 1063) = 3.40$, $p = 0.065$, $\eta^2 = 0.003$, or interaction between order and condition, $F(1, 1063) = 0.53$ $p = 0.466$, $\eta^2 = 0.001$. For single induction participants, I found no effect of order, $F(1, 223) = 2.16$, $p=0.143$, $\eta^2 = 0.009$, or its interaction with condition, $F(1, 223) = 0.56$, $p = 0.453$, $\eta^2 = 0.002$.

from Time 1 to Time 2 and by condition (see Table 2 for means).[7] First I compared the difference in liking between the target and control to 0 for each condition in order to test if explicit preferences were created at Time 1. At Time 1, participants who received the information that the target is a prostitute preferred the control person to the target ($M$ = -0.23, $SD$ = 1.36), $t(617) = -4.20$, $p < 0.0001$, $d = -0.17$. Participants who learned that the target can't feed her kids at Time 1 also preferred the control ($M$ = -0.15, $SD$ = 1.34, $t(569) = -2.63$, $p = 0.009$, $d = -0.11$. Time 1 explicit attitudes did not vary significantly by condition, $t(1186) = 1.05$, $p = 0.680$, $d = 0.06$.

Paired t-tests comparing Time 1 explicit evaluations to Time 2 explicit evaluations indicated a decrease in positivity toward the target for participants in the Internal attribution condition, $t(504) = 10.69$, $p < 0.0001$, $d = 0.44$ and an increase in positivity toward the target for participants in the External attribution condition, $t(550) = 5.25$, $p < 0.0001$, $d = 0.22$. At Time 2, participants in the Internal attribution condition still preferred the control over the target ($M$ = -1.179, $SD$ = 1.722, $t(510) = -13.51$, $p < 0.0001$, $d = -0.60$, but External attribution condition participants no longer showed a preference between target and control ($M$ = -0.07, $SD$ = 1.25), $t(555) = -1.35$, $p = 0.176$, $d$

[7] Few differences by condition were found on the one-item preference measure. This item indicated a preference for the control at Time 1 for when the target was a prostitute ($M$ = 0.13, $SD$ = 1.10), $t(619) = 3.04$, $p = 0.003$, $d = 0.12$, and when the target could not feed her kids ($M$ = 0.13, $SD$ = 1.04), $t(577) = 2.99$, $p = 0.003$, $d = 0.12$, Participants did not show a preference difference by condition at Time 1, $t(1196) = 0.07$, $p = 0.94$, $d = 0.00$. Participant preference did not change from Time 1 to 2 in the Internal attribution condition ($M_{difference}$ = -0.07, $SD$ = 1.76), $t(518) = -0.92$, $p = 0.357$, $d = -0.08$, or the External attribution condition ($M_{difference}$ = 0.09, $SD$ = 1.46), $t(548) = 1.40$, $p = 0.162$, $d$ = 0.12. Both the Internal attribution ($M$ = -0.10, $SD$ = 1.44) and External attribution condition participants ($M$ = -0.03, $SD$ = 1.11) did not show a preference for the target or control at Time 2 (Internal attribution: $t(517) = -1.55$, $p = 0.12$, $d = -0.07$; External attribution: $t(552) = -0.69$, $p = 0.490$, $d = -0.03$. Time 2 explicit preference also did not vary by condition, $t(1069) = 0.84$, $p = 0.400$, $d = 0.04$.

= -0.06. Independent samples t-tests found a large effect of condition on Time 2 explicit attitudes, $t(1065) = 12.01$, $p < 0.0001$, $d = 0.74$ (see Table 2 for means by condition).

***Single Induction.*** Results from the single induction conditions largely replicate the findings in Study 1. On explicit measures, Internal attribution condition participants liked the control more than the target ($M = -0.45$, $SD = 1.43$), $t(161) = -4.01$, $p = 0.002$, $d = -0.32$, while External attribution condition participants showed no explicit preference between control and target ($M = 0.01$, $SD = 1.34$), $t(153) = 0.12$, $p = 0.905$, $d = 0.01$. The condition difference in explicit attitudes was significant, $t(314) = 2.96$, $p = 0.003$, $d = 0.33$.[8]

**Implicit Attitudes**

***Dual Induction.*** Implicit and Explicit attitudes were significantly correlated with each other at both Time 1, $r = 0.16$, $p < 0.0001$, and Time 2, $r = 0.25$, $p < 0.0001$, in the dual induction conditions. The correlation between condition and implicit attitudes was smaller. $r = 0.10$, $p = 0.002$, than the correlation between condition and explicit attitudes, $r = 0.35$, $p < 0.0001$, at Time 2, $z = 6.00$, $p < 0.0001$.

Next, I tested to see if implicit attitudes changed over time in response to the attribution manipulation (see Table 2 for means). I first compared IAT scores to 0 to see if implicit preferences were being created a Time 1. Participants who learned that the target is a prostitute at Time 1 implicitly preferred the control to the target, ($M = -0.07$,

[8] On the one-item explicit preference measure after the single induction, External attribution condition participants showed no preference for target or control ($M = -0.026$, $SD = 1.044$), $t(154) = 0.31$, $p = 0.759$, $d = 0.02$, while Internal attribution condition participants preferred the control to the target ($M = -0.291$, $SD = 1.204$), $t(164) = 3.10$, $p = 0.002$, $d = 0.24$. The Internal attribution condition produced greater participant preference for the control and therefore less positivity toward the target than the External attribution condition, $t(318) = 2.10$, $p = 0.038$, $d = 0.24$.

*SD* = 0.41), $t(563) = -3.74$, $p < 0.0001$, $d = -0.17$, whereas participants who learned that the target can't feed her children showed no implicit preference at Time 1 ($M = -0.03$, $SD = 0.42$), $t(510) = -1.61$, $p = 0.109$, $d = -0.07$. An independent samples t-test showed that implicit attitudes at Time 1 did not differ by condition ($M_{difference} = -0.07$, $SD = 0.41$), $t(1073) = 1.38$, $p = 0.167$, $d = 0.08$. I then conducted paired-samples t-tests to analyze whether implicit attitudes changed from the Time 1 to Time 2. Participants in the Internal attribution condition showed a decrease in implicit preference for the target from Time 1 to Time 2, $t(476) = 3.15$, $p = 0.002$, $d = 0.14$. Those participants in External attribution condition demonstrated a non-significant increase in positivity from Time 1 to Time 2, $t(517) = 1.75$, $p = 0.081$, $d = 0.08$. I used independent samples t-tests to test for mean differences in attitude change. The implicit attitude change from Time 1 to Time 2 in the Internal attribution condition ($M = -0.07$, $SD = 0.48$) was significantly different from the change in the External attribution condition ($M = 0.04$, $SD = 0.45$), $t(993) = 3.52$, $p = 0.001$, $d = 0.22$.

Next, I compared how additional information designed to elicit an internal or external attribution impacted implicit attitudes after initial evaluation. Participants in Internal attribution condition preferred the control over the target at Time 2 ($M = -0.10$, $SD = 0.40$), $t(485) = -5.44$, $p < 0.0001$, $d = -0.25$, but participants in the External attribution condition had no preference between the target and control at Time 2 ($M = -0.02$, $SD = 0.37$, $t(523) = -1.47$, $p = 0.142$, $d = -0.06$. At Time 2, the External attribution condition ($M = -0.10$, $SD = 0.39$) elicited more positivity toward the target than the Internal attribution condition ($M = -0.02$, $SD = 0.37$), $t(1008) = 3.09$, $p = 0.002$, $d = 0.19$.

***Single Induction.*** I found a small to moderate correlation between implicit and explicit attitudes in the single induction conditions, $r = 0.20$, $p < 0.0001$**.** Single induction participants in the Internal attribution condition implicitly preferred the control over the target ($M = -0.21$, $SD = 0.38$), $t(148) = -6.85$, $p < 0.0001$, $d = -0.56$. Participants in the External attribution condition had a smaller implicit preference for the control over the target ($M = -0.11$, $SD = 0.42$), $t(135) = -3.15$, $p = 0.002$, $d = -0.26$. Implicit preference for the target was less positive in the Internal attribution condition than in the External attribution condition ($M_{difference} = -0.11$, $SD = 0.38$), $t(283) = -2.10$, $p = 0.036$, $d = -0.25$.

**Sympathy**

***Dual Induction.*** Finally, I explored the importance of sympathy in response to the information over time and in predicting attitudinal outcomes. Participants felt more sympathy for the target than the control when they learned the target was a prostitute at Time 1 ($M = 0.37$, $SD = 1.23$), $t(615) = 7.35$, $p < 0.0001$, $d = 0.30$, and when they learned the target can't feed her kids at Time 1 ($M = 0.68$, $SD = 1.37$), $t(568) = 11.85$, $p < 0.0001$, $d = 0.50$. The Time 1 information that the target cannot feed her kids produced a bigger difference in sympathy between target and control than the prostitution information, $t(1183) = 4.18$, $p < 0.0001$, $d = 0.24$. Participants showed an increase in sympathy for the target as compared to control in the External attribution condition from Time 1 to Time 2 ($M_{difference} = 0.45$, $SD = 1.15$), $t(548) = 7.41$, $p < 0.0001$, $d = 0.32$. The Internal attribution condition, however, produced a decrease in sympathy for the target over the control from Time 1 to Time 2 for the ($M_{difference} = -0.60$, $SD = 1.67$), $t(505) = 7.62$, $p < 0.0001$, $d = 0.34$.

Participants in the External attribution condition felt more sympathy for the target than the control at Time 2 ($M = 0.88$, $SD = 1.21$), $t(552) = 15.99$, $p < 0.0001$, $d = 0.68$. However, participants in the Internal attribution condition did not show a difference in sympathy between the target and control at Time 2 (M=0.11, SD=1.48), $t(511) = 1.64$, $p = 0.101$, $d = 0.07$. At Time 2, the External attribution condition elicited a greater difference in sympathy between target and control ($M = 0.82$, $SD = 1.21$) than the Internal attribution condition, ($M = 0.11$, SD=1.48), $t(1063)=8.66$, $p < 0.0001$, $d = 0.56$. I conducted a series of regressions to see if sympathy mediated any of the effects of condition on explicit or implicit attitudes. Time 2 sympathy slightly mediated the condition differences in attitudes at Time 2 for both implicit and explicit attitudes. Using Time 2 measures, inclusion of sympathy in a model predicting explicit attitudes from condition changed the estimate of condition effects from ß = 1.10, $t(1065) = 12.01$, $p < 0.0001$, $\eta^2 = 0.12$, to ß = 0.79, $t(1055) = -9.01$, $p < 0.0001$, $\eta^2 = 0.071$, (Sobel's Z = 7.56, $p < 0.0001$). Likewise, including sympathy in the model predicting implicit attitudes from condition decreased the condition estimate from ß = 0.07, $t(1008) = 3.09$, $p = 0.002$, $\eta^2 = 0.009$, to ß = 0.05, $t(995) = 2.02$, $p = 0.044$, $\eta^2 = 0.004$, (Sobel's Z = 3.36, $p = 0.001$).

***Single Induction.*** Next, I compared explicit sympathy ratings in the single induction conditions. Participants in the External attribution condition felt more sympathy for the target than the control, ($M = 0.95$, $SD = 1.39$), $t(152) = 8.48$, $p < 0.0001$, $d = 0.69$, as did participants in the Internal attribution condition ($M = 0.61$, $SD = 1.29$), $t(161) = 6.02$, $p < 0.0001$, $d = 0.47$. An independent samples t-test shows that the External attribution condition produced a greater difference in sympathy for the target than the control than the Internal attribution condition $t(313) = 2.27$, $p = 0.024$, $d = 0.26$.

Finally, the addition of sympathy difference in a regression model predicting liking difference from condition decreased the relationship between condition and that explicit attitude measure from ß = 0.46, $t(314) = 2.96$, $p = 0.003$, $\eta^2 = 0.027$, to ß = 0.37, $t(310) = 2.41$, $p = 0.017$, $\eta^2 = 0.018$, by partially mediating the effect, (Sobel's Z = 2.80, $p = 0.005$). This mediating effect was not found for implicit attitudes.

**Single-Dual Induction Comparisons**. Participant reaction to the External or Internal attribution condition information at Time 2 was compared to learning that information at Time 1 to see if adjustment from an established evaluation was equivalent to evaluation formation. Since the information at Time 2 for the dual induction conditions matched the single induction information, the study design allowed me to test whether introducing the additional negative information and propositional causal link after an initial measurement affected the degree of negativity or positivity towards the target. To that end, I conducted independent samples t-tests comparing explicit and implicit attitudes by induction type.

The Internal attribution condition created a larger difference in explicit liking for the target as compared to control in the dual induction ($M = -1.03$, $SD = 0.37$) than the single induction ($M = -0.29$, $SD = 1.20$), $t(671) = 3.87$, $p = 0.0001$, $d = 0.30$. Participants in the External attribution condition did not vary in liking difference as a result of induction type ($M_{single} = 0.01$, $SD = 1.34$, $M_{dual} = 0.07$, $SD = 1.25$), $t(708) = 0.51$, $p = 0.611$, $d = 0.04$.[9] The Internal attribution condition produced a smaller sympathy difference for the dual induction participants ($M = 0.11$, $SD = 1.48$) than for single

[9] The one item explicit preference measure did not differ by induction type for either the External attribution condition, $t(706) = 0.07$, $p = 0.946$, $d = 0.01$, or the Internal attribution condition, $t(681) = 1.55$, $p = 0.122$, $d = 0.12$.

induction participants ($M = 0.61$, $SD = 1.291$), $t(672) = 3.89$, $p = 0.0001$, $d = 0.30$. The difference in sympathy between the target and control was unaffected by induction type for the External attribution condition, $t(704) = 1.15$, $p = 0.251$, $d = 0.09$.

The External attribution condition information produced less implicit negativity toward the target for dual induction participants ($M = -0.02$, $SD = 0.37$), than for single induction participants ($M = -0.11$, $SD = 0.38$), $t(658) = 2.46$, $p = 0.014$, $d = 0.19$. The dual induction participants also showed less implicit negativity in the Internal attribution condition ($M = -0.10$, $SD = 0.39$) than the single induction participants ($M = -0.21$, $SD = 0.42$), $t(633) = 3.13$, $p = 0.002$, $d = 0.25$. Learning the same information at once before measurement appears to increase the strength of implicit evaluations compared to learning that information at two points with multiple measures.

**Discussion**

The present study provided additional evidence that propositional information affects implicit attitude formation. Study 2a replicated the key effects from Study 1. In single inductions, the Internal attribution condition created more implicit and explicit negativity toward the target than the External attribution condition. However, the extension to multiple time points suggests some constraints on the effects of attribution on implicit attitude formation. Attribution impacted established attitudes explicitly, but only slightly influenced implicit attitudes. Internal attributions for negative events increased implicit negativity, but external attributions did not alter implicit attitudes. The resulting preferences differed from one another, indicating that attribution does impact implicit attitude change, just to a lesser extent than explicit attitude change.

The implicit and explicit attitudes differences based on induction type were not expected. Comparisons between a single attitude induction and a dual induction suggest that stronger implicit attitudes result from a single exposure to information than from conflicting multiple exposures. In contrast, explicit attitudes were more extreme when conflicting information was presented earlier in the study as opposed to a single learning experience. Though implicit attitudes can be changed from an initial starting point, the resulting attitudes do not appear to be adjusted enough from the slight negative initial attitudes to match the single induction attitudes. One shortcoming of the study that the Time 1 information was confounded with Time 2 condition (e.g., participants who learned that Janet was a prostitute at Time 1 always received the Sympathy condition information at Time 2—that Janet is a prostitute because she can't feed her kids. Study 2b attempted to replicate the findings of Study 2a with the male cancer scenario set while counterbalancing Time 1 information.

## Study 2b

### Study Overview

Study 2b is a conceptual replication of Study 2a. This study used the male cancer scenarios from Study 1 and replicated the design of Study 2a with two additional conditions. Unlike the previous study, I randomized which negative information appeared at Time 1 in the dual induction conditions. This design consideration removes a confound allowing for better evidence of differential change between the External and Internal attribution conditions.

## Method

**Participants.** Volunteers on the Project Implicit Website (https://implicit.harvard.edu) were randomly assigned to participate in this study, one of the several running on the site at the Time of data collection. 2184 participants consented, and 1389 (64%) completed the study. I included all participants who completed a relevant dependent measure in analyses, N = 1573.

Sixty-four percent of the participants were women. Participant racial identification included 71% White, 11% Black, and 8% Asian. 11% of participants were Hispanic. The age of participants ranged from 12 to 96 ($M = 31.2$, $SD = 31.6$). Of the demographics collected, only race, $F(8, 1852) = 2.19$, $p = 0.025$, $\eta^2 = 0.009$ and ethnicity, $F(2, 1852) = 4.54$, $p = 0.011$, $\eta^2 = 0.005$, were significant predictors of study completion. Black, Asian, and participants ethnically identified as "other" were slightly more likely to drop out of the study than White and Hispanic participants.

**Design.** The design was identical to Study 2b with the addition of Time 1 information counterbalancing. Study 2b used the cancer scenario set. Target identity and measures order (between implicit and self-report and within self-report) were randomized.

**Materials.** Participants were given the same study explanation as Study 2a. The scenario text was identical to that used in Study 1 for the cancer scenario set. For the dual induction participants, the target's description ended with either "Nathan has cancer." or "Nathan uses illegal drugs." (identical to the Negative 1 and Negative 2 condition information in Study 1). For the single induction participants, the Internal or External attribution condition information ended the target's description. Thus, all four negative

information possibilities were used either for Time 1 (i.e., Negative 1 or Negative 2 conditions) or for Time 2 and the dual induction (Internal or External attribution conditions).

In the dual induction conditions, participants were given the same additional study explanation after Time 1 measurement as in Study 2a. The description of the control at Time 2 was also the same as Study 2a, but target was pictured with one of four captions instead of two.

If participants learned that the target had cancer at Time 1, then in the Internal attribution condition they read, "As you may recall, this is Nathan who has cancer. It turns out that Nathan has cancer because he uses illegal drugs.", and in the External attribution condition they read, "As you may recall, this is Nathan who has cancer. It turns out that Nathan uses illegal drugs because he has cancer." Participants who learned about the target's drug use at Time 1 read, "As you may recall, this is Nathan who uses illegal drugs. It turns out that Nathan has cancer because he uses illegal drugs." in the Internal attribution condition and read "As you may recall, this is Nathan who uses illegal drugs. It turns out that Nathan uses illegal drugs because he has cancer." in the External attribution condition.

**Measures**

***IAT.*** Implicit attitudes towards the target were measured with an IAT assessing preference between Nathan and Henry. This 5-block IAT compared reaction Times to categorizing Nathan with Good and Henry with Bad to Henry with Good and Nathan with bad. The procedure was identical to that in Study 2a. Analyses of participant responses latencies used the D-algorithm. Positive scores indicate preference for the

target compared to the control. Two percent of IAT scores were excluded based on criteria (greater then 30% error rate overall or 40% error rate in a single block).

***Self-Report.*** Explicit measures were identical to those in Study 2a, assessing the evaluations of Nathan and Henry instead of Janet and Alice.

***Manipulation Boost/Check.*** The manipulation boost/check items were identical to Study 2a. All participants were included in analyses.[10]

**Procedure.** The procedure was identical to that in Study 2a except for the scenario set change and counterbalancing the Time 1 information by attribution condition. Target identity randomization did not influence results.[11] Dependent measure

---

[10] Thirty-two percent of participants who completed the study missed one or more manipulation check questions. A model predicting implicit attitudes from failure on the key check question, condition, and the interaction between failure and condition showed no effects of check failure, $F(1, 969) = 0.17$, $p = 0.678$, $\eta^2 = 0.0002$, or the interaction, $F(1, 969) = 0.44$, $p = 0.508$, $\eta^2 = 0.001$, at Time 1. The failure on the key check question at Time 2 did not predict implicit attitudes, $F(1, 888) = 0.01$, $p = 0.916$, $\eta^2 = 0.0000$, or interact with condition, $F(1, 888) = 2.96$, $p = 0.086$, $\eta^2 = 0.003$. A model predicting explicit attitudes at Time 1 showed a main effect of failure on the key check question, $F(1, 1059) = 5.37$, $p = 0.021$, $\eta^2 = 0.005$ but no significant interaction, $F(1, 1059) = 1.00$, $p = 0.317$, $\eta^2 = 0.001$. The model predicting explicit attitudes at Time 2 found no main effect of check failure, $F(1, 946) = 0.35$, $p = 0.552$, $\eta^2 = 0.0004$, or significant interaction, $F(1, 946) = 1.14$, $p = 0.286$, $\eta^2 = 0.001$. The attitudes of single induction participants were unaffected by check failure (Implicit: no main effect $F(1, 216) = 0.06$ $p = 0.811$, $\eta^2 = 0.0003$, or interaction $F(1, 216) = 1.53$, $p = 0.217$, $\eta^2 = 0.007$; Explicit: no main effect $F(1, 242) = 0.65$, $p = 0.420$, $\eta^2 = 0.002$, or interaction $F(1, 242) = 2.44$, $p = 0.136$, $\eta^2 = 0.009$.

[11] A model predicting implicit attitudes from target identity, condition, and the interaction between identity and condition showed a main effect of identity, $F(1, 969) = 113.50$, $p < 0.0001$, $\eta^2 = 0.105$, but no interaction, $F(1, 969) = 0.44$, $p = 0.508$, $\eta^2 = 0.001$, at Time 1. Participants preferred Henry to Nathan regardless of condition. Identity at Time 2 also predict implicit attitudes, $F(1, 888) = 75.57$, $p < 0.0001$, $\eta^2 = 0.078$, but did not interact with condition, $F(1, 888) = 0.03$, $p = 0.852$, $\eta^2 = 0.0000$. A model predicting explicit attitudes at Time 1 found no main effect of identity, $F(1, 1059) = 0.11$, $p = 0.738$, $\eta^2 = 0.0001$, and no significant interaction, $F(1, 1059) = 1.03$, $p = 0.331$, $\eta^2 = 0.001$. The model predicting explicit attitudes at Time 2 also found no main effect of identity, $F(1,$

order was randomized between participants, but the order of implicit and explicit measures largely did not affect results.[12]

## Results

### Explicit Attitudes

***Dual Induction.*** If attribution impacts explicit attitudes once they have been established, then the difference in liking between the target and control should vary from Time 1 to Time 2 and by condition (see Table 3 for means).[13] First, I inspected the means

---

946) = 0.49, $p$ = 0.485, $\eta^2$ = 0.0001, or significant interaction, $F(1, 946) = 0.81$, $p$ = 0.367, $\eta^2$ = 0.0001.

[12] A model predicting implicit attitudes at Time 1 showed no main effect of dependent measure order, $F(1, 969) = 0.79$, $p = 0.375$, $\eta^2 = 0.001$, but a significant interaction of order and Time 1 information, $F(1, 969) = 8.44$, $p = 0.004$, $\eta^2 = 0.001$. When participants learned that the target does drugs at Time 1, implicit attitudes were more negative when the IAT was completed before the explicit measure ($M$ = -0.11, $SD$ = 0.43) than when it was completed after them ($M$ = 0.00, $SD$ = 0.39). When participants learned that the target has cancer at Time 1, implicit attitudes were more negative when the explicit measures were first ($M$ = -0.07, $SD$ = 0.43) than when the IAT was first. ($M$ = 0.01, $SD$ = 0.42). Implicit attitudes at Time 2 were unaffected by measure order, $F(1, 888) = 2.97$, $p$ = 0.085, $\eta^2$ = 0.003, or the interaction of order and condition, $F(1, 888) = 1.03$, $p$ = 0.311, $\eta^2$ = 0.001. For single induction participants, I found no main effect of measure order on implicit attitudes, $F(1, 177) = 3.73$, $p$ = 0.055, $\eta^2$ = 0.020, and interaction between order and condition $F(1, 177)$ =0.46, $p$ = 0.500, $\eta^2$ = 0.003. At Time 1, a model predicting explicit attitudes did not indicate significant effects of dependent measure order, $F(1, 1058) = 0.27$, $p$ = 0.601, $\eta^2$ = 0.0003, or an interaction between order and Time 1 information $F(1, 1058) = 1.48$, $p$ = 0.224, $\eta^2$ = 0.001. Order $F(1, 946) = 0.09$, $p$ = 0.767, $\eta^2$ = 0.0001, and the interaction of order and condition, $F(1, 946) = 1.35$, $p$ = 0.246, $\eta^2$ = 0.001, were also not predictive of Time 2 explicit attitudes. For single induction participants, I found no main effect of order $F(1, 191) = 0.11$, $p$ = 0.743, $\eta^2$ = 0.001, or significant interaction, $F(1, 191) = 0.67$, $p$ = 0.414, $\eta^2$ = 0.004 predicting explicit attitudes

[13] The one item explicit preference measures yielded the following results: 1) Cancer information at Time 1 created preference for the target ($M$ = 0.155, $SD$ = 0.946), $t(523)$ = 3.74, $p$ = 0.0002, $d$ = 0.16; 2) Drug information at Time 1 drugs created preference for control ($M$ = -0.371, $SD$ = 1.115), $t(543) = 7.77$, $p < 0.0001$, $d$ = 0.33; 3) Time 1 preference varied by condition, $t(1066) = 8.30$, $p < 0.0001$, $d$ = 0.25. 4) No preference between target and control at Time 2 for the External attribution participants, ($M$ = -

to see if negative attitudes were formed based on Time 1 information. Participants who learned that the target uses illegal drugs liked the control more than the target ($M$ = -0.47, $SD$ = 1.38), $t(539) = -7.92$, $p < 0.0001$, $d = -0.34$. However, participants who learned that the target has cancer liked the target more than the control ($M = 0.15$, $SD = 1.10$), $t(522) = 3.01$, $p = 0.003$, $d = 0.13$. An independent samples t-test shows this Time 1 information difference in explicit attitudes was significant, $t(1061) = 8.01$, $p < 0.0001$, $d = 0.49$.

Then, I compared the difference in liking between target and control for the two conditions at Time 2 using dependent means t-tests. Participants in the Internal attribution condition explicitly preferred the control to the target at Time 2 ($M$ = -0.71, $SD$ = 1.53, $t(477) = -10.15$, $p < 0.0001$, $d = -0.93$, but participants in the External attribution condition showed no explicit preference between the target and control ($M$ = -0.02, $SD$ = 1.26), $t(471) = 0.40$, $p = 0.687$, $d = -0.02$. An independent samples t-test found that explicit attitudes at Time 2 were more negative towards the target for Internal attribution participants than for External attribution participants, $t(948) = 7.56$, $p < 0.0001$, $d = 0.56$.

Paired-samples t-tests within condition tested for changes in explicit attitudes from Time 1 to Time 2. Participants in the Internal attribution condition increased in explicit negativity toward the target from Time 1 to Time 2 ($M_{difference} = 0.52$, $SD = 1.62$, $t(475) = 6.97$, $p < 0.0001$, $d = 0.32$. However, participants in the External attribution condition showed no change in explicit attitudes from Time 1 to Time 2, $M_{difference} = 0.11$, $SD = 1.66$, $t(469) = 1.47$, $p = 0.14$, $d = 0.14$.

---

0.0485, $SD$ = 0.994), $t(474) = 1.06$, $p = 0.289$, $d = 0.05$, or Internal attribution participants ($M$ = -0.075, $SD$ = 1.182), $t(481) = -1.31$, $p = 0.191$, $d = -0.06$; 5) No change over time for Internal attribution, $t(480) = 0.78$, $p$=0.434, $d = 0.04$, or External attribution participants, $t(473) = 0.28$, $p = 0.781$, $d = 0.01$, or Time 2 condition difference, $t(955) = 0.31$, $p = 0.754$, $d = 0.01$.

Next, I examined the effects of Time 1 information to see if the two negative events (that the target has cancer or the target does illegal drugs) had different impact attitudes based on the time points. A model predicting explicit attitudes at Time 2 tested for main effects of Time 1 information and attribution condition and their interaction, $F(3, 946) = 25.18, p < 0.0001, r^2 = 0.074$. Results indicate a main effect of both Time 1 information, $F(1, 946) = 14.74, p = 0.0001, \eta^2 = 0.017$, and condition, $F(1, 946) = 55.87, p < 0.0001, \eta^2 = 0.055$, but no significant interaction, $F(1, 946) = 2.81, p = 0.094, \eta^2 = 0.003$ (see Table 3 for means).

When participants in the External attribution condition learned the target uses illegal drugs at Time 1, they showed the expected change with explicit attitudes becoming more positive over time, ($M_{difference} = 0.67, SD = 1.56$), $t(243) = 6.64, p < 0.0001, d = 0.43$. When the participants learned that the target had cancer at Time 1, however, those in the External attribution condition showed more negativity at Time 2 then Time 1 ($M_{difference} = -0.48, SD = 1.56$), $t(255) = -4.66, p < 0.0001, d = -0.29$. Also contrary to expectations, when participants in the Internal attribution condition learned that the target has cancer at Time 1, their explicit attitudes decreased in positivity at Time 2, ($M_{difference} = -0.90, SD = 1.70$), $t(246) = -8.34, p < 0.0001, d = -1.06$. If Internal attribution condition participants learned that the target does drugs at Time 1, however, then their explicit attitudes did not change at Time 2, $M = -0.10, SD = 1.43, t(228)=-1.11$, $p=0.268, d = -0.07$.

In a independent samples t-test, participants in the Internal attribution condition did not differ on explicit attitudes at Time 2 based on Time 1 information $t(476) = 1.39, p = 0.164, d = 0.06$. However, External attribution condition participants did show Time 2

differences based upon which information they were given at Time 1. Participants who learned the Cancer information at Time 1 showed more negative explicit attitudes toward the target at when they learned the target does illegal drugs because he has cancer at Time 2 ($M$ = -0.28, $SD$ = 1.32) than those who learned the drug information at Time 1 and found out the target does illegal drugs because he has cancer at Time 2 ($M$ = 0.22, $SD$ = 1.14), $t(470) = 4.38$, $p < 0.0001$, $d = 0.40$. These findings suggest that learning that the target does drugs at Time 2 disrupts the effects of external attributional framing on explicit attitude change while learning the target has cancer at Time 2 disrupts the influence of internal attributional framing.

***Single Induction.*** For participants in the single induction conditions, attitude results by attribution condition largely replicated those found in Study 1. Participants in the Internal attribution condition liked the control more than the target ($M$ = -0.30, $SD$ = 1.01), $t(114) = -3.12$, $p = 0.002$, $d = -0.29$, but participants in the External attribution condition showed no explicit preference between control and target, ($M$ = -0.12, $SD$ = 1.29), $t(130) = -1.08$, $p = 0.280$, $d = -0.09$. Explicit attitudes did not vary by condition, however, $t(244) = 1.16$, $p = 0.247$, $d = 0.07$.[14]

**Implicit Attitudes**

***Dual Induction.*** Implicit and Explicit attitudes were correlated at both Time 1, $r = 0.19$, $p < 0.0001$, and Time 2, $r = 0.21$, $p < 0.0001$, in the dual induction conditions. The correlation between condition and explicit attitudes was larger, $r = 0.24$, $p < 0.0001$,

[14] The one-item preference measure showed no preferences for target or control in either condition (External attribution: $M = 0.07$, $SD = 1.08$, $t(131) = 0.73$, $p = 0.469$, $d = 0.06$; Internal attribution: $M = -0.09$, $SD = 1.00$, $t(115) = 0.93$, $p = 0.335$, $d = 0.09$) or condition differences in preference, $t(246) = 1.16$, $p = 0.246$, $d = 0.15$.

than the correlation between condition and implicit attitudes, $r = 0.11$, $p = 0.001$, at Time 2, $z = 2.87$, $p = 0.004$.

Next, I tested to see if implicit attitudes changed over time in response to the External or Internal attribution conditions (see Table 3 for means). Inspection of the means indicate that participants who read that the target uses illegal drugs at implicitly preferred the control to the target at Time 1 ($M = -0.09$, $SD = 0.43$), $t(488) = -4.71$, $p < 0.0001$, $d = -0.21$. The Time 1 information that the target has cancer created no implicit preference for target over control ($M = 0.04$, $SD = 0.41$), $t(483) = 0.20$, $p = 0.841$, $d = 0.01$. An independent samples t-test found that Time 1 implicit attitudes were slightly more negative when participants learned the target uses illegal drugs than when they learned he had cancer, $t(890) = 2.02$, $p = 0.044$, $d = 0.14$. I used paired-samples t-tests to look for implicit attitude changes from Time 1 to Time 2. Participants in the External attribution condition did not alter their implicit attitudes from Time 1 to Time 2 ($M_{difference} = 0.01$, $SD = 0.45$), $t(454) = 0.59$, $p = 0.552$, $d = 0.03$, and participants in the Internal attribution condition showed a very small, non-significant, increase in implicit negativity toward the target from Time 1 to Time 2 ($M_{difference} = 0.04$, $SD = 0.49$), $t(431) = 1.69$, $p = 0.091$, $d = 0.16$.

Inspecting the IAT means at Time 2 indicated that Internal attribution condition participants preferred the control to the target at Time 2 ($M = -0.11$, $SD = 0.40$), $t(435) = -5.42$, $p < 0.0001$, $d = -0.26$; Participant in the External attribution condition showed no implicit preference for the target or control ($M = -0.02$, $SD = 0.38$), $t(455) = -1.18$, $p = 0.240$, $d = -0.06$. Independent samples t-tests show that participants in the External attribution condition were less implicitly negative towards the target at Time 2 ($M = -$

0.02, $SD = 0.39$) than participants in the Internal attribution condition at Time 2 ($M = -0.11$, $SD = 0.40$), $t(890) = 3.20$, $p = 0.001$, $d = 0.22$.

Next, I examined the effects of Time 1 information to see if learning about a negative event (cancer or illegal drug use) had a different impact based on time point. A model predicted IAT D-scores at Time 2 from attribution condition, the information at Time 1 (cancer or drug use) and their interaction accounting for 2% of the variance, $F(1, 888) = 6.68$, $p = 0.0002$, $r^2 = 0.022$. Results show a main effect for condition $F(1, 888) = 11.00$, $p = 0.001$, $\eta^2 = 0.012$, and for Time 1 information, $F(1, 888) = 4.92$, $p = 0.027$, $\eta^2 = 0.005$, and a significant interaction, $F(1, 888) = 5.00$, $p = 0.026$, $\eta^2 = 0.006$, (see Table 3 for means). The difference between Time 2 IAT D-scores for the Internal attribution and External attribution conditions was larger when the drug information was given at Time 1.

Paired samples t-tests show no Time 1 to Time 2 differences for the Cancer/Internal attribution, Drugs/Internal attribution, Cancer/External attribution, or Drugs/External attribution condition combinations. Participants in the Internal attribution condition had more negative implicit attitudes towards the target at Time 2 when the Time 1 information was about drugs and they learned the target does drugs because he has cancer at Time 2 ($M = -0.05$, $SD = 0.38$) than when it the Time 1 information was about cancer and they learned the target does drugs because he has cancer at Time 2 ($M = -0.17$, $SD = 0.42$), $t(434) = 3.03$, $p = 0.003$, $d = 0.29$. All other mean differences based on Time 1 information were non significant. These findings suggest learning about drug use at time 2 reverses any effects of internal attributions on implicit attitude change.

***Single Induction.*** I found no correlation between implicit and explicit attitudes in the single induction conditions, $r = 0.06$, $p = 0.194$. External attribution condition participants implicitly preferred the control over the target ($M = -0.12$, $SD = 0.37$), $t(118) = 3.36$, $p = 0.0004$, $d = -0.31$, as did Internal attribution condition participants ($M = -0.22$, $SD = 0.41$, $t(100) = 5.45$, $p < 0.0001$, $d = -0.55$. Participants in the Internal attribution condition had slightly more negative implicit attitudes towards the target than participants in the External attribution condition, but this effect was non-significant, $t(218) = 1.90$, $p = 0.059$, $d = 0.15$.

**Sympathy**

***Dual Induction.*** Next, I explored the importance of sympathy in response to the attribution manipulation over time and in predicting attitudinal outcomes. Sympathy score were calculated by subtracting sympathy for the control from sympathy for the target. When participants learned the target had cancer at Time 1, they felt more sympathy for the target than control ($M = 1.08$, $SD = 1.32$), $t(521) = 18.61$, $p < 0.0001$, $d = 0.82$. Participants who learned that the target uses illegal drugs at Time 1 showed no differences in felt sympathy between the target and control ($M = 0.06$, $SD = 1.09$), $t(537) = 1.27$, $p = 0.205$, $d = 0.05$. An independent samples t-test indicated that sympathy differed based on the Time 1 negative information, $t(473) = 4.33$, $p < 0.0001$, $d = 0.40$. I used paired samples t-tests to see if sympathy changed from Time 1 to Time 2. Sympathy for the target as compared to the control increased from Time 1 to Time 2 for participants in the External attribution condition ($M_{difference} = 0.24$, $SD = 1.63$), $t(468) = 3.18$, $p = 0.002$, $d = 0.15$. The sympathy difference decreased from Time 1 to Time 2 for Internal

attribution condition participants ($M_{difference}$ = -0.46, $SD$ = 1.63), $t(470) = -6.08$, $p < 0.0001$, $d = -0.56$.

Dependent means t-tests found more sympathy for the target that control in both the Internal attribution condition ($M = 0.13$, $SD = 1.30$), $t(474) = 2.05$, $p = 0.041$, $d = 0.09$, and the External attribution condition ($M = 0.79$, $SD = 1.29$), $t(473) = 13.34$, $p < 0.0001$, $d = 0.61$. Time 2 sympathy for the target as compared to the control was greater for External attribution condition participants than Internal attribution condition participants, $t(947) = 7.95$, $p < 0.0001$, $d = 0.52$. A model predicting the difference in sympathy at Time 2 between the target and control from Time 1 information, condition, and their interaction accounted for 8% of the variance, $F(1, 945) = 28.11$, $p < 0.0001$, $r^2 = 0.082$. Time 1 information had a main effect on the sympathy difference at Time 2, $F(1, 945) = 17.35.01$, $p < 0.0001$, $\eta^2 = 0.017$, as did condition, $F(1, 945) = 61.72$, $p < 0.0001$, $\eta^2 = 0.060$, but the interaction was not significant, $F(1, 945) = 2.61$, $p = 0.107$, $\eta^2 = 0.003$.

I conducted a series of regressions to see if sympathy mediated any of the effects of condition on explicit or implicit attitudes. Including Time 2 sympathy in regression models predicting Time 2 explicit and implicit attitudes from condition found partial mediation effects. The effect of condition on Time 2 IAT score, ß = 0.08, $t(890) = 3.20$, $p = 0.002$, $\eta^2 = 0.010$, was reduced when Time 2 sympathy difference was included in the regression model, ß = 0.05, $t(878) = 2.45$, $p = 0.014$, η2 = 0.007 (Sobel's Z = 3.47, $p = 0.0005$). The condition effect on explicit liking differences at Time 2 decreased from ß = 0.68, $t(948) = 7.56$ $p < 0.0001$, $\eta^2 = 0.056$ to ß = 0.38, $t(940) = 4.47$, $p < 0.0001$, $\eta^2 = 0.020$ (Sobel's Z = 7.07, $p < 0.0001$).

***Single Induction.*** Participants were more sympathetic towards the target than the control in both the External attribution condition ($M = 0.99$, $SD = 1.40$), $t(131) = 8.09$, $p < 0.0001$, $d = 0.70$, and the Internal attribution condition ($M = 0.44$, $SD = 1.14$), $t(103) = 4.42$, $p < 0.0001$, $d = 0.44$. The External attribution condition created more sympathy for target over control than the Internal attribution condition, $t(244) = 3.36$, $p = 0.001$, $d = 0.43$. Sympathy did not mediate implicit or explicit effects.

**Single/Dual Induction Comparisons.** I conducted a series of independent samples t-tests to compare resulting attitudes after the single or dual attitude induction. For those in the Internal attribution condition, participants who received a dual induction had more negative explicit attitudes towards the target over control than the single induction participants, $t(591) = 2.76$, $p = 0.006$, $d = 0.23$. External attribution condition participants did not vary in their explicit attitudes by induction type, $t(601) = 0.79$, $p = 0.429$. Participants in the Internal attribution condition who received the dual induction formed less negative implicit attitudes toward the target than those in the single induction, $t(535) = 2.66$, $p = 0.008$, $d = 0.23$. Dual induction participants in the External attribution condition also formed less negative implicit attitudes toward the target than single induction External attribution condition participants, $t(573) = 2.62$, $p = 0.008$, $d = 0.22$.

## Discussion

Study 2b provided more evidence that explicit attitude change is influenced by propositional knowledge in the form of attribution. However, explicit attitudes only changed in the Internal attribution condition and not the External attribution condition. Implicit attitudes appeared unchanged over time regardless of condition. When Time 1

information and its interaction with condition were included in a model predicting implicit attitudes, however, implicit attitudes varied by attribution condition at Time 2. Like in Study 2a, this difference at Time 2 shows an effect, albeit a small one, of attribution on established implicit attitudes.

Like in previous studies, attribution did seem to influence both explicit and implicit attitude formation. The difference in implicit attitudes between the single induction conditions was weak in this case, however. Power may have been insufficient to reliably detect the effect. The small effects for the cancer scenario found in Study 1 suggest that ~100 participants per cell may not have been sufficient.

Some differences in the results of Study 2a and Study 2b (e.g., the failure to replicate explicit change in the external attribution condition) seem to be related to the scenario set and counterbalancing of the Time 1 information. The content of the negative information learned at Time 1 proved influential on both implicit and explicit Time 2 attitudes. Presumably, the sympathy created by learning the target has cancer at either time point or the overwhelming negativity of learning that the target does drugs at Time 2 disrupted the influence of attribution in changing attitudes.

## Study 3

### Overview

The results from the studies conducted so far indicate that external attributions that create sympathy decrease implicit negativity in attitude formation. In Study 3, I attempted to uncover whether external attributions are preventing negative attitudes from being formed or forming good associations that are counterweights to the formed bad associations. The difference between these outcomes could have important theoretical

implications. If both negative and positive associations are created by sympathy, then it might produce a state of implicit attitude ambivalence. For instance, implicit attitude ambivalence causes more neutral attitudes and less confidence in automatic evaluation (e.g., Petty et al., 2006). However, if sympathy counteracts the formation of negative attitudes – then the timing of when sympathy is evoked may be critical to the formation (or not) of negative evaluations. For example, if external attributions create sympathy during attitude formation, negativity is attenuated; however, if the attribution is given after the implicit preference has already been formed than it won't affect negativity.

Study 3 attempted to examine why participants in the External attribution condition have shown no preference for target or control in Studies 1, 2a, and 2b. The neutral implicit attitudes in the External Attribution condition have at least two possible explanations: 1) External attributions create sympathy which forms Good associations that balance out negativity toward the target; or 2) the external attributions prevents the negative associations from forming in the first place (see Figure 1 for illustration).

In order to compare good and bad associations, I needed to find a way to measure them separately. Recent theory and evidence has suggested that the Brief IAT (Sriram, & Greenwald, 2009) may be able to capture distinct "Good" and "Bad" components of evaluation (Axt and Nosek, in prep). Using these implicit measures, I investigated whether external attributions for negative events resulted in either good and bad association formation or neutralization of association formation.

**Method**

**Participants.** Participants were volunteers on the Project Implicit Website (https://implicit.harvard.edu) randomly assigned to participant in this study, one of the

several research studies concurrently running on the site. Of the 1864 participants who consented, 1172 (63%) completed the study. All participants who completed the relevant dependent measure were included in analyses.

Women comprised 65% of the study sample. Participants were 10% Hispanic ethnically and 71% White, 9% Black, and 7% Asian racially. Fourteen to 82 year-olds participated ($M$ = 31.4, $SD$ = 13.2). Demographics did not predict study completion, $F(25, 1620) = 1.31$, $p = 0.138$, $r^2 = 0.020$.

**Design.** Study 3 was a 2 (attribution condition) x 2 (implicit measure) between-subjects design. Target identity was randomized and the order of the dependent measures was randomized (between implicit and self-report and within self-report).

**Materials.** The materials were identical to those used in Study 1 for the prostitution scenario set's External and Internal attribution conditions.

**Measures**

***BIAT.*** The Brief IAT (BIAT; Sriram, & Greenwald, 2009) was used as the implicit measure. Behaviorally, the critical response blocks for the task are identical to the IAT; however, perceptually, participants' attention can be focused on the category associated with good or the category associated with bad. In the BIAT, a category and attribute (e.g., Janet and good words) are defined as "focal," whereas the other category and attribute (e.g., Alice and bad words) are defined as non-focal. "Focal" means that the category/attribute label is present on the screen along with the category exemplars. The "non focal" category labels do not appear on the screen, but the stimulus items are still presented and categorized by the participant as not being a member of the focal categories.

The participants concentrate on the focal categories (e.g., Good and Janet) and reject Bad words or pictures of Alice as NOT being either good or Janet. The BIAT has two versions, one of which has good words always as a member of the focal pair, whereas the other has bad words as a stable member of the focal pair. Which person is paired with the stable category switches after half of the trials; for instance, if Janet and Good are focal first then Alice and Good are focal second. In each version of the BIAT, pairing order of the stable focal category with the non-stable category (Janet or Alice) is randomized between participants. 8% of IAT scores were excluded based on criteria (greater than 30% error rate overall or more than 40% error rate in each block). Positive *D* scores indicate implicit preference for the target person compared to the control person for both Good and Bad versions of the BIAT.

***Explicit Attitudes.*** Four measures were used to assess explicit attitudes toward the control and target individuals. To parallel the implicit measures, separate questions were designed to measure positivity and negativity. For explicit positivity, participants were asked, "Considering just your positive (and ignoring any negative) thoughts and feelings—How positive do you feel towards Janet [Alice]?" about both individuals. Five response options ranged from "Not at all positive" to "Extremely positive". Explicit negativity was measured using the following question: "Considering just your negative (and ignoring any positive) thoughts and feelings—How negative do you feel towards Janet [Alice]?" for both individuals. The responses ranged from, "Not at all negative" to "Extremely negative".

***Sympathy.*** The same sympathy items from previous Studies were used.

***Manipulation Check/Boost***. The manipulation check items were identical to those used in Study 1 for the Internal and External attribution conditions of the prostitution scenario set.

**Procedure**. Besides use of the BIAT, the materials and procedure for Study 3 was nearly identical to Study 1. Participants were assigned to the External or Internal attribution condition for the prostitution scenario set. Identity of the target was randomized and did not affect results.[15] After participants read about the target and control, they completed the manipulation boost/check used in Study 1. Next, participants completed either a Good focal or Bad focal BIAT and 6 explicit items measuring positive feelings, negative feelings, and sympathy for the target and control. Type of implicit measure and order of implicit and explicit measures were randomized.[16] Finally, participants were debriefed and provided feedback on their BIAT scores.

---

[15] When included in a model predicting BIAT score from condition. identity of the target was unrelated to BIAT score $F(1, 1026) = 0.00$, $p = 0.978$, $\eta^2 = 0.0000$, and did not interact with condition $F(1, 1026) = 0.18$, $p = 0.673$, $\eta^2 = 0.0002$. Unlike in the previous studies, participants did not implicitly prefer the lighter haired woman. When identity was included in models predicting implicit attitudes from identity and it's interaction with condition for each version of the BIAT, no significant results were found, $\eta^2$ s $< 0.0003$. Target identity and its interaction with condition were also unrelated to explicit attitude measures, all $\eta^2$ s $< 0.002$.

[16] Overall, implicit attitudes were unaffected by measure order, $F(1, 1026) = 0.03$, $p = 0.863$, $\eta^2 = 0.0000$, or the interaction of order and condition, $F(1, 1026) = 0.00$, $p = 0.983$, $\eta^2 = 0.0000$. This was true for the Bad and Good versions of the BIAT, all $\eta^2$s $< 0.005$. The explicit attitude composite was not significantly predicted by dependent measure order, $F(1, 1240) = 0.44$, $p = 0.505$, $\eta^2 = 0.0004$, or the interaction between order and condition, $F(1, 1240) = 0.19$, $p = 0.661$, $\eta^2 = 0.0002$. Explicit positivity and explicit negativity individually were also unrelated to order or its interaction with condition, $\eta^2$s $< 0.0005$.

## Results

**Explicit Attitudes**. First, I tested whether attribution influenced the formation of explicit positivity and negativity towards the target. Explicit positivity was calculated by subtracting feelings of positivity for the control person from feelings of positivity for the target. The target person was evaluated less positively than the control person with in the External attribution ($M = -0.14$, $SD = 1.09$, $t(619) = -3.14$, $p = 0.002$, $d = -0.13$) and Internal attribution conditions ($M = -0.40$, $SD = 1.17$, $t(647) = -8.62$, $p < 0.0001$, $d = -0.34$). Comparing explicit positivity by condition revealed more positivity towards the target compared to the control in the External attribution condition than in the Internal attribution condition, $t(1266) = 4.07$, $p < 0.0001$, $d = 0.23$.

Explicit negativity was calculated by subtracting feelings of negativity for the control person from those for the target. The target person was evaluated more negatively than the control person in the External attribution ($M = 0.23$, $SD = 1.00$), $t(620) = 5.67$, $p < 0.0001$, $d = 0.23$) and Internal attribution conditions ($M = 0.45$, $SD = 1.247$, $t(641) = 9.20$, $p < 0.0001$, $d = 0.36$). Comparing explicit negativity by condition revealed less negativity towards the target compared to the control in the External attribution condition than in the Internal attribution condition, $t(1261) = 3.53$, $p = 0.0004$, $d = 0.20$.

I created an explicit attitude composite by subtracting the difference between target and control in explicit negativity from the difference between target and control in explicit positivity. Participants demonstrated more overall negative attitudes toward the target as opposed to the control the internal attribution condition ($M = -0.85$, $SD = 2.06$) than participants in the External attribution condition ($M = -0.35$, $SD = 1.78$), $t(1243) = 4.57$, $p < 0.0001$, $d = 0.26$. See Table 4 for means by condition.

**Implicit Attitudes**. Overall, implicit and explicit attitudes were slightly correlated, $r = 0.15$, $p < 0.0001$. I examined whether attribution condition influenced implicit attitude formation as measured by the two versions of the BIAT. Inspection of the Good BIAT means by condition indicated no preference among participants in the Internal attribution condition ($M = 0.002$, $SD = 0.33$), $t(263) = 0.07$, $p = 0.941$, $d = 0.00$, or External attribution condition ($M = -0.004$, $SD = 0.32$), $t(256) = -0.22$, $p = 0.827$, $d = -0.01$. However, preference for the control person over the target was found when the Bad BIAT measured attitudes regardless of condition. Participants favored the control over the target in both the Internal attribution ($M = -0.09$, $SD = 0.32$, $t(254) = -4.61$, $p < 0.0001$, d = -0.29) and External attribution ($M = -0.07$, $SD = 0.32$, $t(253) = -3.52$, $p < 0.0001$, d = -0.22) conditions.

I also tested whether implicit preferences between the target and control person as measured by the BIAT (regardless of focal attribute) differed between attribution conditions. Inspection of the BIAT means showed that the target person was negatively evaluated compared to the control person in the Internal attribution condition ($M = -0.05$, $SD = 0.33$), $t(518) = -3.08$, $p = 0.002$, $d = -0.14$. Contrary to expectation, the overall BIAT showed a preference for the control person over the target in the External attribution condition as well ($M = -0.04$, $SD = 0.33$), $t(510) = -2.63$, $p = 0.009$, $d = -0.12$.

A model predicting BIAT scores from BIAT type, attribution condition, and their interaction accounted for 2% of the variance, $F(3, 1026) = 5.46$, $p = 0.001$, $r^2 = 0.016$). Results indicate a main effect of the BIAT type, $F(1, 1026)=15.79$, $p < 0.0001$, $\eta^2 = 0.015$, but no main effect for attribution condition, $F(1, 1026) = 0.14$, $p = 0.708$, $\eta^2 = 0.0001$, and no interaction, $F(1, 1026) = 0.44$, $p = 0.506$, $\eta^2 = 0.0004$.

**Sympathy.** Finally, I tested to see how explicit measures of sympathy were related to explicit and implicit attitudes and the condition manipulations and if they mediated condition effects. Sympathy differences were calculated by subtracting sympathy for the control from sympathy for the target. Participants felt more sympathy for the target than the control in the External attribution condition ($M = 0.96$, $SD = 1.39$), $t(619) = 17.13$ , $p < 0.0001$, $d = 0.69$, and in the Internal attribution condition ($M = 0.76$, $SD = 1.49$), $t(649) = 12.91$, $p < 0.0001$, $d = 0.51$. Sympathy for the target compared to the control was larger in the External attribution Condition, $t(1268) = 2.53$, $p = 0.011$, $d = 0.14$. Sympathy was significantly but weakly related to explicit attitudes, $r = 0.17$, $p < 0.0001$, and unrelated to implicit attitudes, $r = 0.04$, $p = 0.198$. See Table 5 for all correlations between measures.

**Discussion**

Study 3 replicated the previous explicit attitude results that propositional knowledge affects attitudes formation using the alternative measures of felt positivity and negativity. External attributions for negative information created less negativity and more positivity towards the target compared to the control than internal attributions. However, the results that propositional knowledge in the form of attribution affects implicit attitudes were not replicated. Attitudes as measured by the Bad BIAT were equally negative regardless of propositional knowledge. The Good BIAT did not reflect any differences between target and control or by condition.

The implicit findings in this study could be accounted for by one of the following explanations: a) the BIAT is less sensitive than the IAT and not capable of capturing condition differences seen in the previous studies; b) the BIAT captures a different aspect

of implicit cognition than the IAT and that aspect is less influenced by propositional knowledge; or c) only bad associations were formed by the manipulations and propositional knowledge of external attribution would have only affected good associations. The present data cannot distinguish from these alternatives but does provide further evidence that the focal category of the BIAT may determine which types of associations it is measuring.

## Study Four

### Study Overview

The first three studies focused on whether propositional knowledge in the form of attribution influences the impact of negative information on attitude formation and change External attributions of negative events appear to create sympathy, decreasing the negativity of both implicit and explicit attitudes. The purpose of Study 4 was to expand scope of the results by examining the effects of propositional knowledge in the form of casual attribution on positive as well as negative events.

This study presented neutral information about Henry and Nathan and then additional information about the target person: getting or losing a scholarship and receiving good or bad grades. I used combinations of the same two pieces of information, either both positive, both negative, or one of each, in causal orders designed to create either external or internal attributions. For instance, getting the scholarship and getting good grades were combined in a causal order that suggest internal attributions in one condition (i.e., “Henry got the scholarship because Henry got good grades.”) and external attributions in another condition (i.e., “Henry got good grades because Henry got a scholarship”). I varied the valence and causal order of the scholarship and grade events to

create eight combinations. The IAT and explicit measures of preference and liking were used as relative dependent measures to assess attitudes toward Henry and Nathan.

Studies 1 and 2 provided evidence that propositional knowledge that influences explicit attitude formation will have a similar impact on implicit attitudes. In line with these findings, I predicted that both implicit and explicit attitudes created by the external attribution combinations will be less extreme than those created by the internal attribution combinations. For instance, the two positive pieces of information combined to elicit an internal attribution (i.e., "Henry got the scholarship because Henry got good grades.") will create more positive attitudes than when they are combined to elicit an external attribution (i.e., "Henry got good grades because Henry got a scholarship"). See Table 6 for the study conditions and predictions of their results.

**Method**

**Participants.** Participants were volunteers on the Project Implicit Website (https://implicit.harvard.edu) randomly assigned to participant in this study, one of the several research studies concurrently running on the site. 2860 participants consented, 1870 (65%) of who completed the study. Analyses included all participants who completed the relevant dependent measure (N = 2300).

Women comprised 61% of the study sample. The racial and ethnic composition of the sample was as follows: 10% Hispanic, 71% White, 8% Black, and 8% Asian. The age of participants ranged from 12 to 82 ($M$ = 31.7, $SD$ = 13.1). Gender and ethnicity predicted participant dropout such that men, $F(1, 2443) = 6.96$, $p = 0.008$, $\eta^2 = 0.003$, and those ethnically identifying as "other", $F(2, 2443) = 3.98$, $p = 0.019$, $\eta^2 = 0.003$, were slightly less likely to complete the study.

**Design.** I implemented a 2 (scholarship valence) x 2 (grade valence) x 2 (Internal/External attribution) between-subjects design. Target identity, order of implicit and explicit measures, and order of the explicit measures were each randomized.

**Materials**. For this study, I created new versions of the Henry and Nathan scenarios used in Study 1 and 2b. An introduction page read, "Pictured below are Nathan and Henry. They are both students at Central University." and included pictures of the two cartoon figures. An additional page included the following information about their University:

Central University is a four-year comprehensive university in the Midwestern United States. Each year, Central University offers scholarships for current students who show academic promise. These scholarships are renewable based upon student success and continued funding resources.

The filler descriptions of the individuals were identical to the ones used in Studies 1-3 excluding the job information. The target description included an additional sentence that described him getting or losing a scholarship and receiving bad or good grades, e.g., "While at Central University, Henry lost a scholarship even though he got good grades". The positive and negative versions of the events were combined to create different causal relationships designed to produce internal or external attributions. That is, the conditions aimed to produce inferences that the target person was responsible (or not) for a good outcome or a bad outcome. See Table 6 for the text of all eight conditions and predictions of their results.

**Measures**

***IAT***. A seven-block IAT assessing preference between Nathan and Henry was used to measure implicit attitudes. The procedure, materials, and analysis strategy were

identical to those used in Study 1. Participants with greater than 30% error rates overall or more than 40% error rates per block were excluded (3%).

***Self-Report***. The same items assessing Henry and Nathan used in Study 2b were included to measure explicit attitudes.

***Manipulation Boost/Check***. Two manipulation check items were used. As in previous studies, the a manipulation check item displayed a picture of Henry and asked participants to identify him. Excluding participants with incorrect answers on this item did not affect research conclusions, so they were left in for analyses[17].

A attribution check item repeated the manipulation information (e.g., Recall that "Henry got good grades because Henry got a scholarship.") and then assessed the attribution it created by asking why the first event occurred (e.g., "Why did Henry get good grades?"). Responses to the question ranged on a 7 point scale from “It was entirely because of Henry[Nathan] and not at all because of the situation or circumstance that Henry[Nathan] was in.” to “It was entirely because of the situation or circumstances that Henry[Nathan] was in and not at all because of Henry[Nathan].” with “It was partly because of Henry[Nathan] and partly because of the situation or circumstances that Henry was in.”

As predicted, I found a main effect of attribution condition on this item, $F(1, 2250) = 263.46$, $p < 0.0001$, $\eta^2 = 0.093$. Participants were more likely to choose responses that implicated the target as responsible for the event in the conditions

[17] Twenty-three percent of participants missed the manipulation check question. Adding check failure into the models predicting implicit attitudes found no effect of check failure on the IAT, $F(1, 1777) = 0.41$, $p = 0.52$, $\eta^2 = 0.0002$ and no significant interactions (all $\eta^2$ s < 0.002). The same was true for explicit attitudes predictions, $F(1, 2059) = 0.22$, $p = 0.622$, $\eta^2 = 0.0001$, except for an interaction between failure, grades, and attribution condition, $F(1, 2059) = 6.23$ $p = 0.013$, $\eta^2 = 0.003$ (all other $\eta^2$ s < 0.002).

designed to create internal attributions than in conditions designed to create external attributions. A model predicting the attribution check from scholarship valence, grades valence, casual order, and their interactions with one another, indicated significant influences of all the manipulations on the attribution check, $F(7, 2250) = 85.46$, $p < 0.0001$, $r^2 = 0.21$ (see Table 7 for model effects). Mean comparisons show that all four event combinations created significantly different attributions based on their causal order (see Table 8).

**Procedure**. Visitors to the Project Implicit Research website who consented to participate in a research study were randomly assigned to one of eight versions of the Nathan and Henry scenarios. Participants were introduced to the two men and read about the University they attend. The descriptions of the control and target individuals followed. The target identity was randomized and did not affect results.[18] Participants then answered two questions about the information they just read. These questions strengthened the manipulation and served as an attention check. Next, participants completed the implicit and explicit attitude measures in a randomized order. The order of dependent measures did not influence results[19]. Finally, participants were debriefed and given their IAT results.

---

[18] Including target identity in a model predicting implicit attitudes from the condition combinations produced a main effect of target identity, $F(1, 1777) = 126.79$, $p < 0.0001$, $\eta^2 = 0.064$, but no interactions with identity, all $\eta^2$s $< 0.0005$. The same pattern was true for explicit attitudes (main effect of identity: $F(1, 2059) = 43.06$, $p < 0.0001$, $\eta^2 = 0.020$, but no interactions, all $\eta^2$s $< 0.001$). Regardless of who was the target, participants showed an overall preference for Henry over Nathan implicitly ($M = 0.10$, $SD = 0.40$), $t(1793) = 11.03$, $p < 0.0001$, $d = 0.25$, and explicitly ($M = 0.20$, $SD = 0.77$), $t(2074) = 9.43$, $p < 0.0001$, $d = 0.26$.

[19] Adding dependent measure order into the models predicting implicit attitudes from the condition combinations found no effect of order on the IAT, $F(1, 1777) = 2.32$, $p =$

**Results**

**Explicit Attitudes**. First, I tested to see if explicit attitudes varied by scholarship valence, grade valence, and attribution condition. Explicit preference was calculated by subtracting liking for the control person from liking for the target. I created a model using the three condition differences (internal/external attribution; got/lost scholarship; received bad/good grades) and their interactions as predictors of explicit attitudes to see if the manipulations were successful in influencing explicit attitude formation. If the manipulations were successful in forming different explicit attitudes based on the causal order of the same two pieces of information, then attribution condition should interaction with the valence of both events. The model showed main effects for grade valence and attribution, but not scholarship valence, and no significant interactions, $F(7, 2071) = 9.60$, $p < 0.0001$, $r^2 = 0.031$. While receiving bad ($M = -0.32$, $SD = 1.23$) or good grades ($M = 0.06$, $SD = 1.06$), $F(1, 2071) = 59.64$, $p < 0.0001$, $\eta^2 = 0.028$, and internal ($M = -0.09$, $SD = 1.17$) or external ($M = -0.18$, $SD = 1.16$) attribution order, $F(1, 2071) = 4.12$, $p = 0.043$, $\eta^2 = 0.002$, were both significant predictors of explicit attitudes, getting ($M = -0.15$, $SD = 1.14$) or losing ($M = -0.12$, $SD = 1.14$) the scholarship was not, $F(1, 2071) = 0.19$, $p = 0.618$, $\eta^2 = 0.0001$, (see Table 9 for marginal means). Contrary to expectations, casual order did not interact with either or both of the events to predict explicit attitudes, $\eta^2$ s $< 0.001$. See Table 10 for model statistics.[20]

---

$0.128$, $\eta^2 = 0.001$ and no significant interactions (all $\eta^2$ s $< 0.002$). The same was true for explicit attitudes predictions, $F(1, 2058) = 0.14$, $p = 0.704$, $\eta^2 = 0.0001$ (all interaction $\eta^2$ s $< 0.001$).

[20] The one-item preference measure produced the same pattern of results, $F(7, 2072) = 10.32$, $p < 0.0001$, $r^2 = 0.034$. Grades valence, $F(1, 2076) = 59.07$, $p < 0.0001$, $\eta^2 = 0.029$, and attribution condition, $F(1, 2076)=4.31$, $p = 0.038$, $\eta^2 = 0.002$, were significant predictors of preference, but scholarship valence, $F(1, 2076) = 2.36$, $p = 0.125$, $\eta^2 =$

Inspection of the individual means showed a preference for the target over the control when he got a scholarship because he got good grades ($M = 0.22$, $SD = 1.23$), $t(238) = 2.76$, $p = 0.006$, $d = 0.18$. However, I found a preference for control over the target when the target (a) got bad grades even though he got a scholarship ($M = -0.34$, $SD = 1.13$), $t(258) = -4.48$, $p < 0.0001$, $d = -0.30$, (b) got a scholarship even though he got bad grades ($M=-0.38$, $SD=1.13$), $t(245) = -4.43$, $p < 0.0001$, $d = -0.34$, lost a scholarship because he received bad grades ($M = -0.35$, $SD = 1.25$), $t(271) = -4.65$, $p < 0.0001$, $d = -0.28$, and got bad grades because he lost a scholarship ($M = -0.24$, $SD = 1.17$), $t(272) = -3.40$, $p = 0.001$, $d = -0.21$. The other three conditions did not produce liking differences (see Table 11 for means).

Simple effects tests indicated the only difference based on causal order for the same grade and scholarship combination occurred for good grades and getting a scholarship. The causal order suggesting an internal attribution created more positive attitudes ($M = 0.23$, $SD = 1.25$) than the external attribution order ($M = 0.004$, $SD = 0.91$), $t(510) = 2.32$, $p = 0.021$, $d = 0.21$. See Table 11 for explicit means by all 8 condition combinations.

**Implicit Attitudes**. Implicit and Explicit attitudes were correlated with one another to a small degree, $r = 0.17$, $p < 0.0001$., I tested whether implicit preferences between the target and control person differed base on condition differences by running a regression predicting IAT scores from scholarship valence, grades valence, attribution condition, and their interactions with one another, $F(7, 1785) = 6.42$, $p < 0.0001$, $r^2 = 0.025$. If the manipulations were successful in creating different implicit attitudes based

---

0.001, was not. All interactions with attribution condition were non-significant, $\eta^2$ s < 0.0005.

on the causal order of the same two pieces of information, attribution condition will interact with scholarship and grade valence. Receiving bad ($M$ = -0.08, $SD$ = 0.40) or good grades ($M$ = 0.03, $SD$ = 0.42), $F(1, 1789) = 37.68$, $p < 0.0001$, $\eta^2 = 0.021$, and a causal order that suggests internal ($M$ = -0.01, $SD$=0.41) or external ($M$ = -0.04, $SD$ = 0.41) attributions, $F(1, 1789) = 4.06$, $p = 0.044$, $\eta^2 = 0.002$, were both significant predictors of implicit preference, but getting ($M$ = -0.01, $SD$ = 0.41) or losing ($M$ = -0.04, $SD$ = 0.42) the scholarship was not, $F(1, 1789) = 1.69$, $p = 0.194$, $\eta^2 = 0.001$. See Table 9 for marginal means. Contrary to expectations, none of the interactions were significant in predicting implicit attitudes, $\eta^2$ s < 0.001 (see Table 12 for model).

Inspection of the individual means showed that the target person was negatively evaluated compared to the control person in the conditions where the target got a scholarship even though he got bad grades ($M$= -0.08, $SD$ = 0.41, $t(214) = -2.70$, $p$ = 0.008, $d$ = -0.20), got bad grades even though he got a scholarship ($M$ = -0.07, $SD$ = 0.38, $t(221) = -2.84$, $p = 0.005$, $d$ = -0.19), lost a scholarship because he got bad grades ($M$ = -0.11, $SD$ = 0.41, $t(239) = 3.99$, $p < 0.0001$, $d$ = -0.26), and got bad grades because he lost his scholarship ($M$=-0.08, $SD$=0.41, $t(235)=2.93$, $p = 0.004$, $d$ = -0.20). Participants showed a preference for the target over the control when he got a scholarship because he got good grades ($M$ = 0.10, $SD$ = 0.39, $t(197) = 3.62$, $p = 0.0004$, $d$ = 0.27. IAT scores indicated no preference in the other three conditions. See Table 11 for means by all 8 condition combinations.

Simple effects tests found only one difference in implicit attitudes based on the causal order of the same grade and scholarship combination. Getting a scholarship because the target got good grades created more positive attitudes ($M$ = 0.10, $SD$ = 0.39)

than getting good grades because he got a scholarship ($M = 0.003$, $SD = 0.42$), $t(443) = 2.52$, $p = 0.012$, $d = 0.24$. As predicted, the causal order that suggested an internal attribution of positive events created more positive attitudes than the order that suggested an external attribution.

**Discussion**

Like Studies 1-3, Study 4 tested whether causal orders that created external or internal attributions of the same events influence attitude formation. The study included both positive and negative versions of events combined to suggest either internal or external attributions of their cause. Though the combinations successfully created attribution differences on a manipulation check, these differences were not directly mirrored in implicit or explicit attitude formation. The only significant differences based on order occurred when the target got good grades and got a scholarship. This difference was found for both implicit and explicit attitudes, providing initial evidence that propositional knowledge can influence the formation of positive attitudes. The previous studies only found this for negative attitudes. However, this finding should not be overstated given the lack of differences for the other three event combinations.

Since explicit attitudes largely did not vary within matched events by attribution condition, few claims can be made about the influence of attribution on implicit attitudes from this study. Perhaps the large main effect of grade valence could have overwhelmed any attribution condition effects. Receiving bad grades created a preference for the control regardless of the scholarship valance or casual order. More likely, however, the manipulation content lacked the emotional impact needed to induce attitudes that could be influence by propositional knowledge. Alternatively, the large main effect of grade

valence could have overwhelmed any attribution effects. Study 5 will increase the emotional content of the stimuli to allow for stronger implicit and explicit attitudes to be formed that can then in turn be influenced (or not) by attribution.

Another potential shortcoming for this and previous studies is the causal order of the conditions being compared confounds attribution with which event is being explained. For example, in this study, the causal order that created external attributions (e.g., Nathan got good grades because Nathan got a scholarship) explained the a one event (good grades) and the order creating internal attributions (e.g., Nathan got a scholarship because Nathan got good grades) explained the other (getting a scholarship). Study 5 will use proposition knowledge to alter the meaning of positive and negative information by varying the attribution of the same outcome.

## Study Five

### Study Overview

The purpose of Study 5 was to examine the effects of attribution on already established positive and negative implicit attitudes. Studies 2a and 2b suggest that propositional knowledge likely affects existing explicit but not implicit attitudes. One limitation of drawing a strong conclusion from the earlier studies is that the attitude effects were small, especially implicitly. In Study 5, I altered the manipulation in an effect to increase the impact of the manipulation by creating more affectively laden scenarios to create stronger positive or negative attitudes. Further, in the research designs thus far, the manipulations of attribution varied the causal order. In contrast, the attribution manipulations in Study 5 explained the same outcome.

Participants were presented with one of four scenarios, two positive and two negative. The scenarios described events that occurred at the target's house, implying the target's involvement and responsibility. Attitudes were measured once for half the participants (a single attitude induction) and twice for the other half (a dual attitude induction). In the single induction conditions, after reading about the positive or negative events, participants immediately learned that the target was out of town and not responsible for the events. In the dual induction conditions, after reading about the positive or negative events, participants completed implicit and explicit attitude measures, and then learned that the target was not responsible before completing the attitude measures a second time.

The general expectation is that participants should form positive or negative attitudes toward the target based on the events that occurred at the home because they are intuitively attributed to the target. However, after learning that the target was not home, participants would shift the attribution away from the target. The question: does that attribution shift alter implicit attitudes? Further, does it matter if attitudes are "set" by measuring them prior to learning the alternative attribution (dual induction conditions) or not (single induction conditions)?

Given the results of Studies 2a and 2b showing that propositional knowledge is unlikely to influence implicit attitude change, I predicted that the attribution information given before Time 2 in the dual-induction condition would attenuate the explicit but not implicit attitudes. I also predicted that the induction type (dual or single) will matter more for implicit attitudes than explicit attitudes, and that prior measure will decrease the impact of the manipulation in the dual induction conditions. Participants who receive the

external attribution information during attitude formation (in the single induction conditions) will have little to no preference for either the target or control implicitly or explicitly. These single induction attitudes should be weaker both implicitly and explicitly than those measured at time one for the dual induction participants but only weaker explicitly than the attitudes measured at Time 2 for the dual induction participants.

**Method**

**Participants.** Visitors to the Project Implicit Website (https://implicit.harvard.edu) were randomly assigned to participate in the study, one of the multiple research studies running during date collection. Of the 1902 participants who consented, 1269 (67%) completed the study. All participants with data were included in analyses (N = 1441).

Women comprised 64% of the sample. Participants were 9% Hispanic ethnically and 70% White, 11% Black, and 7% Asian racially. Participants ranged in age from 13 to 85 ($M$ = 35.93, $SD$ = 13.24). Of the demographics measured, age, $F(1, 1657) = 9.02$, $p = 0.003$, $\eta^2 = 0.005$, and race, $F(8, 1657) = 2.00$, $p = 0.044$, $\eta^2 = 0.009$, predicted participant drop out. Older participants and Black and Asian participants were slightly less likely to complete the study than the rest of the sample.

**Design.** This study used a Solomon 4 group design. Half of participants read negative or positive information at Time 1 and the attribution information at Time 2. The other half of participants read all the information before completing any dependent measures. Four scenarios were used, 2 of each valence. Overall, the Study was a 2 Valence (positive vs. negative) x 2 Scenario (within valence) x 2 Attitude Induction

(single vs. dual) between-subjects design. Target identity and measure order were randomized (between implicit and explicit and within explicit).

**Materials.** Filler information identical to that used in Studies 1-3 was given for Janet and Alice. For the critical manipulation, I created 4 scenarios describing events that happened at the target's house. One positive event was a charity fundraiser being held at the target's house:

"Last weekend, Alice [Janet]'s house was the location of a fundraiser for underprivileged kids who cannot afford adequate health care. The fundraiser attracted some of the town's most generous philanthropists. The purpose of the event was to raise money to provide local children with regular check-ups and pay for doctor's visits and medicine for the chronically ill. In all, the attendees of the fundraiser at Alice [Janet]'s house donated five hundred thousand dollars to the cause. Because of the event at Alice [Janet]'s house, these children will receive the medical care they need to live long happy and healthy lives. Humanitarians of America is reviewing the event for recognition of excellence in fundraising for its upcoming annual awards."

The other positive event was a neighborhood coming together to help storm victims at the target's house:

"Last weekend, Alice [Janet]'s house served as a beacon of hope after a storm ravaged her neighborhood. A tornado had touched down a street over, destroying several houses and injuring their residents. An emergency triage and supply station was set up at Alice [Janet]'s house once the storm passed. At the house, the injured were bandaged and taken care of and others who had lost their homes were given food and shelter. The whole neighborhood came together at Alice [Janet]'s house to take care of one another and show the spirit and determination of their community in the face of hardship. Humanitarians of America is reviewing the event for recognition of excellence in disaster response for its upcoming annual awards."

One negative event was the house being the site of an out of control party that led to fatalities:

"Last weekend, Alice [Janet]'s house was the site of a party that led to tragic consequences. At least 40 teenagers and young adults were in attendance when neighbors complaining about the noise called the police. The police arrived at Alice [Janet]'s house to find the party had gotten out of control with physical violence, underage drinking, and

rampant illegal drug use. Upon noticing that several of the partygoers were unconscious and non-responsive, the officers called emergency services. Of the five individuals who were rushed to the hospital with drug and alcohol poisoning, only two survived. Several other people present at Alice [Janet]'s house were arrested. There was substantial evidence of underage drinking, assault, possession, contributing to the delinquency of a minor, and possibly involuntary manslaughter. Alice [Janet]'s house is currently a crime scene and an extensive investigation is underway with multiple felony counts likely to be distributed."

The other negative event was the target's house being used as a methamphetamine lab:

"Last weekend, police raided Alice [Janet]'s house. After reports of strange smells coming from the vents that led to the basement and people coming and going from her house at odd hours, the police suspected that Alice [Janet]'s house was being used to cook and sell meth. During the raid, the police found all the equipment and chemicals necessary to produce methamphetamine. They also found illegal weapons and large amounts of cash. Most disturbingly, an unattended toddler was found in one of the rooms of the house. The toddler was malnourished and showed signs of neglect and abuse. Alice [Janet]'s house is currently a crime scene and will not be released to her in the near future. An extensive investigation is underway with multiple felony counts likely to be distributed."

Information excusing the target from responsibility for the events that occurred was given on a separate page with a picture of the target. This attribution correction was created to override any attributions made after the initial scenario with an external attribution. The charity scenario was matched with this correction:

"Recall that this is Alice [Janet], whose house held a charity event that raised money for sick kids. Following a review of the event for their annual awards, the Humanitarians of America determined the following: Alice [Janet]'s house was indeed the location of a charity event where substantial funds were raised for children's healthcare. Alice [Janet] had no knowledge of this. Six months earlier, Alice [Janet] moved to Australia for a year and left a rental agency in charge of the space. The charity organizers rented Alice [Janet]'s house from this agency for the event."

This external attribution information followed the storm scenario:

"Recall that this is Alice [Janet], whose house was the site of a disaster relief effort. Following a review of the event for their annual awards, the Humanitarians of America determined the following: Alice [Janet]'s house was indeed the location of an

emergency triage, supply station, and housing for displaced victims. Alice had no knowledge of this. Six days earlier, Alice [Janet] had left on a trip to Australia for three weeks and had locked up the home for the time that she would be away. The relief organizers had used a neighbor's key to enter Alice [Janet]'s home and set it up to help victims of the tornado because the house was undamaged."

Both attribution corrections for the negative scenarios started with these sentences:

"Recall that this is Alice, whose house was the site of a police investigation. Following the investigation, the police determined the following:" The party scenario then continued, "Alice's house was indeed the location of an extensive and illegal party that had tragic consequences. Alice had no knowledge of this. Six days earlier, Alice left on a trip to Australia for three weeks and locked up her home for the time that she would be away. The party organizers had broken into Alice's house after she left and to use it for their party.". The meth scenario continued, "Alice's house was indeed the location of an extensive methamphetamine lab, and substantial child neglect and abuse had been occurring. Alice had no knowledge of this. Six months earlier, Alice moved to Australia for a year and locked up her home for the time that she would be away. The criminals had broken into Alice's house two months after she left and started using it for their drug lab."

The ending of all the attribution corrections was identical: Participants were told, "Alice [Janet] had no knowledge of the events and was not responsible for their occurrence. Please take this into consideration when evaluating Alice [Janet]."

**Measures**

***IAT***. A five-block IAT compared preference between Alice and Janet to measure implicit attitudes. See Study 2a for details of this IAT version. Two percent of IAT scores were excluded based on criteria (greater then 30% error rate overall or 40% error rate in one block).

***Explicit Attitudes***. Separate measures of liking for Alice and Janet measured explicit attitudes. An additional explicit item assessed relative preference. The items were identical to those used in previous studies.

***Manipulation Boost/Check***. A total of four possible manipulation check items were used. The first asked participants to identify which woman was pictured (i.e., the

same check item as used in previous studies). A second item used after the single induction and after Time 1 for the dual induction asked about the identity of the target by inquiring which woman whose house the positive or negative event occurred, (e.g., "Which woman's home held a charity event that raised money for sick children?"). Once participants knew that the target was not responsible for the event, after the single induction and after the Time 2 manipulation for the dual induction, they were asked the following: "Why is Alice [Janet] not responsible for the events that occurred at her home?" The response options included the correct explanation that she "had no knowledge of the events and was in Australia at the time." Excluding the participants who answered the check items incorrectly did not substantively change the results.[21]

---

[21] Forty-four percent of participants missed one or more manipulation check questions. For the dual induction participants, at Time 1, a model predicting implicit attitudes from failure on the key check question (which woman's home), valence, and the interaction between failure and valence showed no effect of check failure, $F(1, 630) = 0.01$, $p = 0.918$, $\eta^2 = 0.0000$, or interaction, $F(1, 630) = 1.78$ $p = 0.189$, $\eta^2 = 0.003$. The failure on the second key check question (responsibility) at Time 2, $F(1, 601) = 0.21$, $p = 0.650$, $\eta^2 = 0.0003$, and its interaction with valence, $F(1, 601) = 0.34$, $p = 0.559$, $\eta^2 = 0.001$, also did not predict implicit attitudes. Further, the change from Time 1 to Time 2 was not predicted by either key check failure question or their interactions with valence (all $\eta^2$s < 0.003). Single induction participants showed no main effect or interaction with responsibility check failure for implicit attitudes, $F(1, 628) = 0.03$, $p = 0.857$, $\eta^2 = 0.0001$, and $F(1, 628) = 0.89$, $p = 0.347$, $\eta^2 = 0.001$, respectively. Explicit analyses at Time 1 found a main effect of failure on the key check question, $F(1, 687) = 13.91$, $p = 0.0002$, $\eta^2 = 0.016$, and a significant interaction between valence and check failure, $F(1, 687) = 16.92$. $p < 0.0001$, $\eta^2 = 0.020$. Inspection of the means indicate that failing the check question impacted explicit attitudes in the negative conditions more than the positive conditions. At Time 2, failure on the second key check question did not predict explicit attitudes, $F(1, 619) = 0.31$, $p = 0.577$, $\eta^2 = 0.001$, but valence did interact with check failure to predict explicit attitudes, $F(1, 619) = 6.93$, $p = 0.009$, $\eta^2 = 0.011$. Like at Time 1, negative attitudes were more affected by manipulation check failure. Using the check failure on the responsibility question and it's interaction with valence to predict explicit attitude change found no main effect of check failure, $F(1, 617) = 3.64$, $p = 0.057$, $\eta^2 = 0.005$, but a significant interaction, $F(1, 614) = 12.64$, $p = 0.0004$, $\eta^2 = 0.016$. Mean comparisons indicate that answering the Time 2 check correctly increased the explicit attitude change in the negative conditions more than in the positive conditions.

***Attribution.*** A final item included after the single induction and at both Time 1 and Time 2 of the dual inductions assessed the attribution created by the manipulation. The item was measured on a 7-point scale ranging from "It was entirely because of Alice [Janet] and not at all because of the situation or circumstance that Alice [Janet] was in." to "It was entirely because of the situation or circumstances that Alice [Janet] was in and not at all because of Alice [Janet]." with "It was partly because of Alice [Janet] and partly because of the situation or circumstances that Alice [Janet] was in."

As predicted, attributions changed from Time 1 ($M$ = -0.60, $SD$ = 1.67) to Time 2 ($M$ = 2.61, $SD$ = 1.04) in the dual induction conditions regardless of valence, $t(634) = 42.91$, $p < 0.0001$, $d = 1.70$. Participants were more likely to choose responses that implicated the target as responsible for the event at Time 1 than at Time 2. Mean comparisons show that all four scenarios created more external attributions at Time 2 that at Time 1 in the dual induction conditions and that external attributions were created in all single induction conditions (See Table 13). The influence of attribution will be discussed more in the results.

**Procedure**. Visitors to the Project Implicit Research website who consented to participate in a research study were randomly assigned to either the dual or single induction condition, and orthogonally to receive one of two positive scenarios or one of two negative scenarios. First, all participants were told they would learn information

---

Using the check failure on the house question and it's interaction with valence to predict explicit attitude change found a main effect of check failure $F(1, 617) = 6.06$, $p = 0.014$, $\eta^2 = 0.008$ and a significant interaction, $F(1, 617) = 15.51$, $p < 0.0001$, $\eta^2 = 0.020$. Mean comparisons indicate overall less change when participants failed the first check; further, check failure had more impact on explicit attitude change in the negative conditions. For explicit attitudes, single induction participants showed no main effect, $F(1, 697) = 1.11$, $p = 0.293$, $\eta^2 = 0.002$, or interaction, $F(1, 697) = 3.33$ $p = 0.069$, $\eta^2 = 0.005$.

about Alice and Janet. They were presented with a picture of the control person with the filler description used in the previous studies. The identity of the target was randomized and did not influence results[22]. Then participants saw the target person with a picture, filler description, and an additional paragraph describing one of four events. These four possible conditions (two negative and two positive) described events in detail that occurred at the home of the target.

Next, in the dual induction conditions, participants completed manipulation check items that included the attribution question. These items were followed by implicit and explicit attitude measures. Then, those participants were told they would be presented with more information about Alice and Janet. A condition matched description provided information that the target was out of town during the event and did not know that it occurred. Then, these participants were presented with an additional manipulation check

---

[22] Including target identity in a model predicting implicit attitudes at Time 1 from condition valence produced a main effect of target identity, $F(1, 630) = 37.04$, $p < 0.0001$, $\eta^2 = 0.054$, and an interaction with identity, $F(1, 630) = 5.54$, $p = 0.019$, $\eta^2 = 0.008$. As seen in other studies, participants implicitly prefer Janet to Alice regardless of condition. At time 2, the main effect persisted, $F(1, 601) = 37.56$ $p < 0.0001$, $\eta^2 = 0.059$, but the interaction was no longer significant, $F(1, 601) = 2.21$, $p = 0.138$, $\eta^2 = 0.003$. The model predicting implicit attitudes from identity in the single induction conditions yielded a main effect of identity, $F(1, 628) = 8.00$, $p < 0.0001$, $\eta^2 = 0.100$, but no interaction, $F(1, 628) = 0.12$, $p = 0.728$, $\eta^2 = 0.0002$. Neither implicit and explicit change were impacted by target identity, $\eta^2$ s $< 0.003$. For explicit attitudes at Time 1, inclusion of target identity in a model with valence and their interaction revealed no main effect pf identity, $F(1, 687) = 0.32$, $p = 0.569$, $\eta^2 = 0.0004$, or interaction, $F(1, 687) = 1.63$, $p = 0.203$, $\eta^2 = 0.002$. At Time 2, I found no main effect of identity, $F(1, 619) = 0.16$, $p = 0.687$, $\eta^2 = 0.0003$, or interaction with valence, $F(1, 619) = 1.02$, $p = 0.312$, $\eta^2 = 0.002$, in predicting explicit attitudes. In the single induction conditions, target identity did not have a main effect or interact with valence in predicting explicit attitudes, $F(1, 697) = 0.00$, $p = 0.997$, $\eta^2 = 0.0000$, and $F(1, 697) = 2.16$, $p = 0.142$, $\eta^2 = 0.003$, respectively.

item and the same attribution question and implicit and explicit attitude measures used at time one again.

Participants in the single induction condition were immediately given the information about the target's whereabouts after the event was first described. They then completed all manipulation check items and the attitude measures once. The order of the implicit and explicit measures were randomized but kept consistent from Time 1 to Time 2 for dual induction participants.[23] Regardless of condition, participants were debriefed and given feedback from one of the implicit measures.

## Results

### Explicit Attitudes

[23] Including dependent measure order in a model predicting implicit attitudes from valence at Time 1 found no main effect of order or interaction with valence, $F(1, 630) = 0.43$ $p = 0.514$, $\eta^2 = 0.001$, and $F(1, 630) = 1.07$, $p = 0.300$, $\eta^2 = 0.002$, respectively. In a model predicting implicit attitudes at Time 2, the order of the measures and its interaction with valence were also non-significant, $F(1, 601) = 0.11$, $p = 0.743$, $\eta^2 = 0.0002$, and $F(1, 601) = 0.41$, $p = 0.524$, $\eta^2 = 0.0007$, respectively. Implicit attitude change was also not predicted by order or interaction with valence, all $\eta^2$s $< 0.005$. For the single induction participants, order had no main effect on implicit attitudes, $F(1, 628) = 2.15$, $p = 0.143$, $\eta^2 = 0.003$, and the interaction between order and condition was non-significant, $F(1, 628) = 0.18$, $p = 0.670$, $\eta^2 = 0.0003$. For explicit attitudes at Time 1, I found a significant main effect of order, $F(1, 662)=5.53$, $p = 0.019$, $\eta^2 = 0.007$, and an interaction between order and valence, $F(1, 662) =4.05$, $p = 0.045$, $\eta^2 = 0.005$. At Time 2, the model predicting explicit attitudes found no main effect of order, $F(1, 619) = 1.41$, $p = 0.236$, $\eta^2 = 0.002$, and no interaction with valence, $F(1, 619) = 0.82$, $p = 0.366$, $\eta^2 = 0.001$. Change in explicit attitudes from Time 1 to Time 2 was influence by order $F(1, 617) = 9.68$, $p = 0.002$, $\eta^2 = 0.013$, but not the interaction between order and valence, $F(1, 617) = 1.15$, $p = 0.283$, $\eta^2 = 0.002$. For single induction participants, I found no effect of order, $F(1, 697) = 0.11$, $p = 0.739$, $\eta^2 = 0.0002$, but did find a significant interaction between order and valence in predicting explicit attitudes, $F(1, 697) = 11.26$, $p = 0.001$, $\eta^2 = 0.015$. For the negative conditions, explicit preference for the target was greater when implicit measures were completed before explicit measures than when explicit measures were first.

***Dual Induction.*** If the external attribution information influences explicit attitudes after they have been established, then the difference in liking between the target and control should vary in the dual induction condition from Time 1 to Time 2 and by valence (see Table 2 for means).[24] First, I compared the difference in liking between the target and control to 0 for each condition in order to test if explicit preferences were created at Time 1. At Time 1, participants who received the negative scenarios preferred the control person to the target ($M$ = -1.21, $SD$ = 1.93), $t(346) = -11.66$, $p < 0.0001$, $d$ = -0.62. Participants who received the positive scenarios preferred the target to the control at Time 1 ($M$ = 0.21, $SD$ = 1.16, $t(343) = 3.31$, $p = 0.001$, $d = 0.18$. An independent samples $t$-test found a large effect of valence on Time 2 explicit attitudes, $t(689) = 11.68$, $p < 0.0001$, $d = 0.89$.

Paired $t$-tests comparing Time 1 explicit evaluations to Time 2 explicit evaluations indicated a decrease in positivity toward the target for participants in the positive scenario conditions, $t(307) = 5.24$, $p < 0.0001$, $d = 0.30$ and an increase in positivity toward the target for participants in the negative scenario conditions, $t(312) = 10.40$, $p < 0.0001$, $d = 0.59$. The explicit attitude change from Time 1 to Time 2 in the

---

[24] The one-item preference measure yielded a similar same pattern of results. This item indicated a preference for the control at Time 1 for the negative scenarios ($M$ = -0.83, $SD$ = 1.29), $t(350) = -11.98$, $p < 0.0001$, $d$ = -0.64, and a preference for target at Time 1 for the positive scenarios ($M$ = 0.16, $SD$ = 0.87), $t(345) = 3.52$, $p = 0.001$, $d = 0.19$. The difference by valence at Time 1 was significant, $t(695) = 11.86$, $p < 0.0001$, $d = 0.90$. Participant preference for the target decreased from Time 1 to 2 in the positive valence conditions ($M_{difference}$ = -0.26, $SD$ = 0.88), $t(308) = -5.13$, $p < 0.0001$, $d$ = -0.32, and increased in the negative valence conditions ($M_{difference}$ = 0.80, $SD$ = 1.39), $t(316) = 10.33$, $p < 0.0001$, $d = 0.58$. Both the positive scenario ($M$ = -0.08, $SD$ = 0.74) and negative scenario participants ($M$ = 0.00, SD = 0.84) did not show a preference for the target or control at Time 2 (Positive: $t(308) = 1.93$, $p = 0.054$, $d = 0.11$; Negative: $t(316) = 0.07$ $p = 0.947$, $d = 0.00$. Time 2 explicit preference did not vary by condition, $t(624) = 1.23$, $p = 0.255$, $d = 0.10$.

negative conditions ($M = 1.20$, $SD = 2.05$) was significantly larger than the change in the positive conditions ($M = -0.34$, $SD = 1.16$), $t(596) = -11.38$, $p < 0.0001$, $d = -0.93$ presumably because strong negative attitudes, but weakly positive attitudes, were formed at Time 1.

Contrary to the preference at Time 1, participants in the positive conditions preferred the control over the target at Time 2 ($M = -0.14$, $SD = 1.08$, $t(307) = -2.22$, $p = 0.028$, $d = -0.13$, but negative condition participants no longer showed a preference between target and control ($M = 0.00$, $SD = 1.17$), $t(314) = 0.00$, $p = 1.00$, $d = 0.00$. The difference in explicit attitudes at Time 2 based on valence was not significant, $t(621) = 1.51$, $p = 0.131$, $d = 0.12$ (see Table 14 for means by condition). Examining the means for the four scenarios separately found that both positive scenarios created positive attitudes that decreased over time and both negative scenarios created negative attitudes that became more positive over time.[25]

[25] I examined the means for all four scenarios separately. At Time 1, participants preferred the target to control when they received the storm relief scenario, ($M = 0.28$, $SD = 1.19$), $t(166) = 2.99$, $p = 0.003$, $d = 0.23$., but when they received the other positive charity scenario, ($M = 0.14$, $SD = 1.12$), $t(176) = 1.68$, $p = 0.096$, $d = 0.13$. Participants preferred the control to the target when they received either the party, ($M = -0.92$, $SD = 1.67$), $t(191) = -7.67$, $p < 0.0001$, $d = -0.55$, or meth scenarios, ($M = -1.56$, $SD = 2.17$), $t(154) = -8.98$, $p < 0.0001$, $d = -0.72$, at Time 1. For the positive scenarios, explicit preference for the target decreased over time in the charity condition, $t(155) = 3.85$, $p = 0.0002$, $d = 0.31$, and the storm relief condition, $t(151) = 3.59$, $p = 0.0004$, $d = 0.29$. Explicit preference for the target increased from Time 1 to Time 2 in both the party, $t(174) = 7.66$, $p < 0.0001$, $d = 0.58$, and meth conditions, $t(138) = 7.19$, $p < 0.0001$, $d = 0.61$. At Time 2, participants showed a preference for the preference for control over target in the charity condition, ($M = -0.22$, $SD = 1.00$), $t(155) = -2.80$, $p = 0.006$, $d = -0.22$, but no preference in the storm relief condition, ($M = -0.05$, $SD = 1.15$), $t(151) = -0.49$, $p = 0.623$, $d = -0.04$. No preference was found for those in the party, ($M = 0.11$, $SD = 1.15$), $t(174) = 1.31$, $p = 0.192$, $d = 0.10$. or meth, ($M = -0.14$, $SD = 1.18$), $t(139) = -1.43$, $p = 0.154$, $d = -0.12$, conditions either.

***Single Induction.*** Next, I examined the explicit attitudes created after the single induction conditions. On explicit measures, negative scenario condition participants liked the target more than the control ($M = 0.26$ $SD = 1.15$), $t(366) = 4.36$, $p < 0.0001$, $d = 0.23$, while positive scenario participants liked the control more than the target ($M = -0.13$, $SD = 1.09$), $t(333) = -2.16$, $p = 0.031$, $d = -0.12$. The difference in explicit attitudes by valence condition was significant, $t(699) = 4.60$, $p < 0.0001$, $d = 0.35$.[26] These results suggest that participants are overcorrecting explicitly in the single induction conditions– rewarding targets that actually had not been involved in something negative, and punishing targets that actually had not been involved in something positive. [27]

**Implicit Attitudes**

***Dual Induction.*** Implicit and Explicit attitudes were significantly correlated with each other at both Time 1, $r = 0.21$, $p < 0.0001$, and Time 2, $r = 0.21$, $p < 0.0001$, in the dual induction conditions. Next, I tested to see if implicit attitudes changed over time in response to external attribution information (see Table 13 for means). I first compared IAT scores to 0 to see if implicit preferences were being created at Time 1. Participants who received a positive scenario at Time 1 showed no implicit preference, ($M = -0.01$,

---

[26] On the one-item explicit preference measure after the single induction, positive scenario condition participants showed no preference for target or control ($M = -0.06$, $SD = 0.91$), $t(353) = -1.33$, $p = 0.184$, $d = -0.07$, while Negative scenario participants preferred the target to the control ($M = 0.09$, $SD = 0.913$), $t(369) = 1.99$, $p = 0.047$, $d = 0.10$. The difference based on valence condition was significant, $t(680) = 2.35$, $p = 0.019$, $d = 0.18$.

[27] Investigation of means for the positive scenarios indicated a significant preference for control in the storm condition, ($M = -0.20$ $SD = 1.01$), $t(159) = -2.49$, $p = 0.014$, $d = -0.20$, but not the charity condition, ($M = -0.06$ $SD = 1.15$), $t(173) = -0.73$ $p = 0.469$, $d = -0.06$. Both negative scenarios created preferences for the target. Participants in the party condition, ($M = 0.35$ $SD = 1.16$), $t(179) = 4.06$, $p < 0.0001$, $d = 0.30$, and meth condition, ($M = 0.17$ $SD = 1.12$), $t(186) = 2.09$, $p = 0.038$, $d = 0.15$, actually liked the target more then the control.

*SD* = 0.40), *t*(317) = -0.28, *p* = 0.783, *d* = -0.02, whereas participants who received a negative scenario preferred the control to the target Time 1 (*M* = -0.12, *SD* = 0.40), *t*(315) = -5.11, *p* < 0.0001, *d* = -0.29. An independent samples t-test showed that implicit attitudes at Time 1 varied by condition, *t*(632) = 3.43, *p* = 0.001, *d* = 0.27.

I then conducted paired-samples *t*-tests to analyze whether implicit attitudes changed from the Time 1 to Time 2. Participants who received negative scenarios showed a decrease in implicit preference for the control from Time 1 to Time 2, (*M* = 0.10, *SD* = 0.45), *t*(299) = 3.92, *p* < 0.0001, *d* = 0.23. Those participants in positive conditions demonstrated no change in implicit attitudes from Time 1 to Time 2, (*M* = -0.03, *SD* = 0.44), *t*(298) = -0.10, *p* = 0.920, *d* = -0.01. I used independent samples *t*-tests to test for mean differences in attitude change. The implicit attitude change from Time 1 to Time 2 in the negative conditions was significantly greater than the change in the positive conditions, *t*(573) = 3.08, *p* = 0.002, *d* = 0.26.

Next, I compared how external attributions impacted the negative as opposed to positive conditions after initial evaluation. Participants in positive conditions showed no implicit preference at Time 2 (*M* = -0.02, *SD* = 0.38), *t*(299) = -0.90, *p* = 0.368, *d* = -0.05. Neither did participants in the negative conditions at Time 2 (*M* = -0.02, *SD* = 0.35), *t*(304) = -0.99, *p* = 0.322, *d* = -0.06. Implicit Attitudes at Time 2 did not vary by condition, *t*(603) = 0.03, *p* = 0.987, *d* = 0.00. Looking at the scenarios separately found that participants only showed an implicit preference in response to one

scenario—participants who received the meth scenario preferred the control to the target; further, implicit attitude change only occurred for that scenario.[28]

***Single Induction.*** Implicit and Explicit attitudes were significantly correlated with each other in the single induction conditions, $r = 0.15$, $p < 0.0001$. Single induction participants did not show a implicit preference either in the positive scenario ($M = 0.00$, $SD = 0.42$), $t(297) = 0.02$, $p = 0.984$, $d = 0.02$, or negative scenario conditions ($M = 0.03$, $SD = 0.40$), $t(333) = 1.47$, $p = 0.141$, $d = 0.08$. Mean preference did not vary based on condition valence, $t(630) = 0.97$, $p = 0.334$, $d = 0.08$. Participants demonstrated no implicit preferences regardless of scenario, all $ds < 0.1$ (see Table 2 for means).

**Attribution**

***Explicit Attitudes***. Next, I explored the importance of attribution as measured by the attribution check items in predicting explicit outcomes. At Time 1, more external ratings on the attribution item predicted more explicit liking for the target over the control regardless of valence, ß = 0.260, $t(686) = 6.81$, $p < 0.0001$, $r^2 = 0.063$. Including attribution and its interaction with valence in a model predicting explicit attitudes from

---

[28] Participants showed no implicit preference at Time 1 in response to the charity scenario, ($M = -0.002$, $SD = 0.39$), $t(162) = -0.07$, $p = 0.944$, $d = -0.01$, storm relief scenario, ($M = -0.01$, $SD = 0.41$), $t(154) = -0.32$, $p = 0.753$, $d = -0.02$, or party scenario, ($M = -0.05$, $SD = 0.40$), $t(173) = -1.68$, $p = 0.094$, $d = -0.13$. Participants did prefer the control to the target at Time 1 in response to the meth scenario, ($M = -0.20$, $SD = 0.40$), $t(141) = -5.86$, $p < 0.0001$, $d = -0.49$. For the positive scenarios, implicit attitudes did not change from Time 1 to Time 2 in either the charity condition, $t(147) = 0.55$, $p = 0.581$, $d = 0.05$, or the storm relief condition, $t(149) = -0.37$, $p = 0.712$, $d = -0.03$. While I found a change from time 1 to 2 for the meth scenario, $t(137) = -4.06$, $p < 0.0001$, $d = -0.35$, implicit attitudes were not significantly altered over time for the party scenario, $t(161) = -1.62$, $p = 0.107$, $d = -0.13$. At Time 2, participants demonstrated no implicit preferences regardless of scenario—charity: ($M = -0.03$, $SD = 0.38$), $t(148) = -1.16$, $p = 0.246$, $d = -0.10$, storm relief: ($M = -0.00$, $SD = 0.38$), $t(150) = -0.11$, $p = 0.912$, $d = -0.01$, party: ($M = 0.01$, $SD = 0.34$), $t(165) = 0.38$, $p = 0.708$, $d = 0.05$, and meth: ($M = -0.06$, $SD = 0.37$), $t(138) = -1.80$, $p = 0.074$, $d = -0.15$.

valence produced a main effect of valence, $F(1, 661) = 71.29$, $p < 0.0001$, $\eta^2 = 0.033$, and attribution, $F(1, 684) = 30.13$, $p < 0.0001$, $\eta^2 = 0.078$, and a significant interaction, $F(1, 684) = 42.73$ $p < 0.0001$, $\eta^2 = 0.047$. For the positive scenarios, more external attributions yielded less positive attitudes, while for the negative scenarios, more external attributions yielded more positive attitudes. The total model accounted for 25% of the variance in explicit attitudes, $F(3, 684) = 76.38$, $p < 0.0001$, $r^2 = 0.259$.

At Time 2, degree of external attribution did not influence explicit preference when valence was not included in the model, $F(1, 619) = 067$, $p = 0.413$, $r^2 = 0.001$. However, a model predicting explicit attitudes at Time 2 from valence, attribution at Time 2, and their interaction, accounted for 3% of the variance, $F(3, 617) = 7.26$, $p < 0.0001$, $r^2 = 0.034$. A main effect of valence, $F(1, 617) = 12.38$, $p = 0.001$, $\eta^2 = 0.019$, significant interaction, $F(1, 617) = 18.47$ $p < 0.0001$, $\eta^2 = 0.29$, but no effect of attribution, $F(1, 617) = 0.45$, $p = 0.504$, $\eta^2 = 0.001$, resulted. At both Time 1 and Time 2, as stronger external attributions were made in response to the positive scenarios, participants showed less preference for the target. As stronger external attributions were made in response to the negative scenarios, participants showed more preference for the target, $\eta^2$ s $> 0.03$.

Including the change in attribution from Time 1 to Time 2 and its interaction with valence in a model predicting the change in explicit attitudes from Time 1 to Time 2 from valence decreased the effect of valence from ß = 1.55, $t(619) = 11.58$, $p < 0.0001$, $\eta^2 = 0.178$, to ß = -0.264, $t(612) = -1.05$, $p = 0.302$, $\eta^2 = 0.001$. The change in attribution, $F(1, 612) = 85.54$, $p < 0.0001$, $\eta^2 = 0.099$, and the interaction, $F(1, 612) = 67.27$, $p < 0.0001$, $\eta^2 = 0.078$, were both predictive of the change in explicit liking, $F(3, 612) = 78.93$, $p <$

0.0001, $r^2 = 0.285$. A closer look at the interaction indicated that the increase in external attributions over time predicted a decrease in explicit preference for the target for the positive scenarios, ß = -0.133, $t(304) = -3.68$, $p = 0.0003$, $r^2 = 0.043$, and a increase in preference for the negative scenarios, ß = 0.418, $t(308) = 7.64$, $p < 0.0001$, $r^2 = 0.159$.

For single induction participants, attribution did not predict explicit attitudes, $F(1, 671) = 0.14$, $p = 0.711$, $r^2 = 0.0002$. Including attribution in a model predicting explicit attitudes from valence, I found no effect of attribution, $F(1, 694) = 0.30$, $p = 0.584$, $\eta^2 = 0.0004$, but a significant interaction with valence $F(1, 694) = 12.75$, $p = 0.0004$, $\eta^2 = 0.018$. This model accounts for 5% of the variance in explicit attitudes, $F(1, 694) = 11.61$, $p < 0.0001$, $r^2 = 0.048$. For the negative scenarios, stronger external attributions predicted more liking for the target compared to the control, ß = 0.105, $t(363) = 1.98$, $p = 0.048$, $\eta^2 = 0.011$. For the positive scenarios, stronger external attributions predicted less liking for the target compared to the control, ß = -0.143, $t(331) = -3.17$, $p = 0.002$, $\eta^2 = 0.030$. The addition of attribution and it's interaction with valence in the model decreased the relationship between valence and the explicit attitude measure from ß = -0.387, $t(699) = -4.60$, $p < 0.0001$, $\eta^2 = 0.030$, to ß = 0.216, $t(694) = 1.14$, $p = 0.256$, $\eta^2 = 0.002$.

***Implicit Attitudes.*** Next, I considered the role of attribution as measured on the attribution check item in implicit attitude formation and change. At Time 1, more external attribution ratings predicted more implicit preference for the target regardless of valence, ß = 0.030, $t(629) = 3.19$, $p = 0.002$, $r^2 = 0.016$. Including attribution and its interaction with valence in a model predicting implicit attitudes from valence produced a main effect of valence, $F(1, 627) = 4.97$, $p = 0.026$, $\eta^2 = 0.008$, and attribution, $F(1, 627) = 7.33$, $p = 0.007$, $\eta^2 = 0.011$, but no interaction, $F(1, 627) = 1.66$ $p = 0.199$, $\eta^2 = 0.003$.

The total model accounted for 3% of the variance in implicit attitudes, $F(3, 627) = 6.70$, $p = 0.0002$, $r^2 = 0.031$.

At Time 2, degree of external attribution did not influence implicit preference, $F(1, 600) = 1.26$, $p = 0.262$, $r^2 = 0.002$. Including attribution and its interaction with valence in a model predicting implicit attitudes from valence did not yield any significant effects, all $\eta^2$ s $< 0.005$. Including the change in attribution from Time 1 to Time 2 and its interaction with valence in a model predicting the change in implicit attitudes from Time 1 to Time 2 from valence did not produce significant results for valence, $F(1, 599) = 0.04$, $p = 0.835$, $\eta^2 = 0.0001$, attribution, $F(1, 599) = 1.01$, $p = 0.316$, $\eta^2 = 0.002$, or the interaction, $F(1, 599) = 0.05$, $p = 0.823$, $\eta^2 = 0.0001$.

Finally, I investigated the influence of attributions on the implicit attitude outcomes in the single induction conditions. Attribution did not predict implicit attitudes, $F(1, 626) = 2.25$, $p = 0.134$, $r^2 = 0.004$. Including attribution and it's interaction with valence in a model predicting implicit attitudes from valence did not produce significant results for attribution, $F(1, 624) = 2.48$, $p = 0.116$, $\eta^2 = 0.004$, or the interaction, $F(1, 624) = 0.26$, $p = 0.612$, $\eta^2 = 0.0004$.

**Single-Dual Induction Comparisons.** Participant reaction to the attribution correction at Time 2 was compared to learning that information at Time 1 to see if adjustment from an established evaluation was equivalent to evaluation formation. Since the information at Time 2 for the dual induction conditions matched the single induction information, the study design allowed me to test whether introducing the propositional attribution after an initial measurement affected the resulting degree of negativity or

positivity towards the target. To that end, I conducted independent samples *t*-tests comparing explicit and implicit attitudes by induction type.

I added induction type and its interaction with valence in an model predicting explicit attitudes, $F(3, 1320) = 9.51, p < 0.0001, r^2 = 0.021$. Valence did not produce a main effect, $F(1, 1320) = 2.31, p = 0.129, \eta^2 = 0.002$. However, Induction type produced a main effect, $F(1, 1320) = 4.66, p = 0.031, \eta^2 = 0.035$, such that explicit attitudes resulting from the single induction were more positive overall than those resulting from the dual induction. The interaction was also significant, $F(1, 1320) = 4.14, p = 0.042, \eta^2 = 0.003$. The attitudes for the negative scenarios were more positive after the single induction as opposed to the dual induction, while attitudes for the positive scenarios were more negative after the single induction as opposed to the dual induction. Including induction type and its interaction with valence in a model predicting implicit attitudes from valence yielded no significant results, $F(3, 1233) = 1.28, p = 0.279, r^2 = 0.003$, all $\eta^2$ s $> 0.003$. Participants appear to overcorrect their explicit attitudes by forming opposite valenced attitudes in response to the external attribution more in the single induction conditions than the dual induction conditions.

**Discussion**

Study 5 tested whether external attributions of negative and positive events could influence attitude formation and change. Explicit attitudes decreased in strength from Time 1 to Time 2 when external attributions were given at Time 2 for both positive and negative events. Implicit attitudes decreased in strength at Time 2 in response to external attributions for negative but not positive events. The positive scenarios were not successful in creating positive implicit attitudes at Time 1, so it is not surprising that the

external attribution did not have an impact. However, since both implicit and explicit attitudes fully corrected in response to the external attribution manipulation, and the only apparent reason for the difference in effect size is that explicit induction was stronger than implicit induction effect. Because both implicit and explicit attitudes were mostly equivalent to zero at Time 2, this study shows little evidence of differential sensitivity to attribution information.

One unexpected result was the reversal of preference that occurred in response to the external attribution information. Participants over-adjusted their initial positive or negative evaluation of the target, especially when the attribution correction occurred directly after the initial information. Participants appear to be rewarding the targets for not actually being involved in negative events and punishing the targets for not actually being involved in positive events.

## General Discussion

The five studies reported in this dissertation suggest that propositional processes can affect implicit attitude formation and change (see Table 15 for summary of results). Studies 1, 2a, and 4 indicated that propositional knowledge in the form of attributions influenced implicit attitude formation. Studies 1 and 2a demonstrated that indicating an external attribution for a person's negative behavior or circumstances created less negative implicit and explicit attitudes toward the person than indicating an internal attribution. In Study 4, suggesting an external attribution for a person's positive behavior or circumstances created less positive implicit and explicit attitudes toward the person than suggesting an internal attribution. These results demonstrate that propositional

knowledge in the form of attributions can affect implicit attitudes, much like it does for explicit attitudes.

This dissertation also explored whether propositional processes can produce implicit attitude change. Studies 2a, 2b, and 5 used repeated measures designs to test whether attributions are less impactful on established implicit attitudes. In Study 2a, more negative attitudes resulting from the internal attribution manipulation than the external attribution manipulation indicated that attribution does have a small impact when additional negative information is learned about a target. This effect was also found in Study 2b when I accounted for differences based on initial negative information. In Study 5, negative implicit attitudes created at Time 1 were eliminated by a manipulation designed to create external attributions at Time 2.

However, the attributional manipulations in these studies were not always successful in influencing implicit attitudes. Propositional knowledge did not affect implicit attitude formation when a different measure was used to assess negative and positive associations separately (Study 3), power was insufficient (Study 2b—single induction), or the attribution manipulation failed to influence explicit attitude formation (Study 4). Attribution did not effect established implicit attitudes when the initial implicit preferences were not formed in the expected direction (Study 5; Study 2b—dual induction); These studies suggest that the strength and reliability of the manipulation and the dependent measurement are important for observing the effect of attributions on implicit attitude formation and change.

**When Attribution Matters**

The studies reported in this dissertation suggest four conclusions about the influence of attribution on implicit and explicit attitude formation and change. First, attributions only influence implicit attitudes when explicit attitudes are affected. Second, attributions may affect established implicit attitudes to a lesser extent than explicit attitudes; a preference for the target or control must be established before attributions change implicit attitudes. Third, a successful manipulation of attribution and its resulting emotional impact do not necessarily yield attitude differences. Fourth, manipulations of attribution are only successful in influencing attitudes when they drastically change the meaning of the associative information.

First, attributions influence implicit attitude formation only when they influence explicit attitude formation. In study 4, both positive and negative events were ordered to encourage internal or external attributions, but only the positive event combination was influenced by attribution manipulation. This pattern of results was identical for implicit and explicit attitudes, illustrating the parallel impacts of attributional processes on implicit and explicit attitude formation.

In Studies 2a, 2b, and 5, the impact of attributions on attitude change may be slightly stronger for explicit than for implicit attitudes, supporting the general statement of my second conclusion. In Studies 2a and 2b, correlations between attribution condition and attitudinal outcomes were larger for explicit than implicit attitudes. The design of these studies included a strong test of the influence of attribution because additional negative information was provided with the external or internal attribution manipulation. Learning that Janet is a prostitute because she cannot feed her kids at Time 2 suggested an external attribution for her prostitution, which may have attenuated the negativity of

associations; however, this same information also increased the negativity associations related to Janet being unable to feed her children. The non-significant decrease in implicit preference against the target combined with the significant difference based on attribution condition suggest that propositional knowledge promoted small but important changes in implicit attitudes. The changes in explicit attitudes were more pronounced, however, suggesting that attributional processes have a bigger impact on existing explicit attitudes than existing implicit attitudes, consistent with the second conclusion.

The latter part of my second conclusion suggests that implicit preferences need to be established for implicit attitude change to occur from manipulations of attribution. In Study 2b, Negative implicit attitudes were not initially present across conditions. External attributions of negative events would not influence existing neutral or positive attitudes. Similarly, Study 5 provided evidence that attributions are effective in changing explicit attitudes. However, implicit attitude change only occurred for negative event scenarios in Study 5. Because positive implicit attitudes were not created initially in Study 5, the external attribution manipulation had no impact on the existing implicit attitudes. The apparent differential affect of attribution on explicit and implicit attitudes should not be overstated, however. The IAT may just be a less sensitive attitude measure than liking difference in these studies, thus explaining the magnitude differences in effect sizes.

The results of Studies 2a, 2b, 4, and 5 support the third conclusion presented above: Successful manipulations of attribution do not necessarily lead to attitude differences. The attribution manipulations in Study 2a and 2b were effective in altering sympathy from Time 1 to Time 2, but the increase in sympathy in response to external attributions and decrease in sympathy in response to internal attributions did not

consistently yield implicit attitude change. In Study 4, self-report measures of attributions confirmed that the manipulations of attribution were successful in eliciting internal versus external attributions. Thus the null effects of attribution for negative events cannot be explained by a manipulation failure. Just because different orderings of the same information predisposed participants toward internal or external attributions did not mean that those attributions in turn influenced their attitudes. In Study 5, the changes in attribution from Time 1 to Time 2 did not affect implicit attitudes in the positive scenarios. A successful increase in external attribution did not decrease implicit preferences toward the target. The degree to which the manipulations were effective conceptually in these studies did not match their impact on implicit attitudes, further illustrating that successful manipulations of attribution do not automatically produce changes in attitudes.

The fourth conclusion proposed above: The propositional basis for an attribution must have a substantial effect on the meaning of target associations for that attribution to influence attitudes. For instance, learning that Henry lost his scholarship even though he got good grades led to more external attributions than learning that Henry got good grades even though he lost his scholarship. The meaning of the associations barely changed with additional information in the latter ordering of propositions; participant evaluations of the positivity of Henry's good grades did not decrease with the information that he lost his scholarship. In contrast, learning that Henry got scholarship because he got good grades had a much more positive emotional impact than learning that Henry got good grades because he got a scholarship. The positivity of the grades decreased with an external attribution that undermined the truth-value of the association.

Study 5 demonstrated how drastic changes in meaning, in this case fully absolving someone from responsibility for a negative event, have the largest effect on attitudes. In Study 5, external attributions eliminated implicit preferences and reversed explicit preferences, showing that substantial and emotionally evocative attributional processes are particularly effective in influencing attitudes.

**Implications for Implicit Attitude Theory**

To integrate the results of this dissertation with existing research and theory, let us return to the two perspectives outlined in the introduction that provide conflicting predictions on the influence of attribution on implicit attitudes. From one perspective, only associative information impacts implicit attitude formation and change. The propositional processing that integrates the meaning beyond the associations into explicit evaluation does not influence implicit evaluation. Evidence supporting this perspective includes research on contingency judgments (e.g., Baeyens, et al., 1992) and evaluative conditioning (e.g., Olsen & Fazio, 2001) showing that implicit attitudes can be formed in absence of relationship information. However, just because propositional processing is not required in implicit attitude formation does not be those processes never have an impact. This dissertation demonstrates that attributional processes can influence implicit attitude formation. Previous research on attitude change also points to implicit attitudes as purely associative. Explicit but not implicit attitudes change in response to negation (Gregg, Seibt, and Banaji, 2006) and cognitive dissonance (Gawronski and Strack, 2004). These results can be explained in part by the conclusions of my dissertation: first, explicit attitude change is easier to achieve via propositional processes than implicit attitude change, and second, propositional knowledge must drastically change the

meaning of the associative information. In Gregg, Seibt, and Banaji (2006), the manipulation telling participants that the information they were given was false may not have been emotionally impactful enough to yield implicit attitude change. In Gawronski and Strack (2004), the cognitive dissonance manipulation did not alter the meaning of the associations. Dissonance reduction affects explicit and not implicit attitudes because the subtle and likely unconscious influences on the meaning of the associations are not drastic enough to impact implicit attitudes.

This dissertation provides some evidence for the alternative perspective that propositional processes can, at least under some conditions, affect implicit attitudes. In line with the findings that accuracy information affects attitude formation (De Houwer, 2007, 2009; De Houwer et al., 2005), I found that attributional processes can influence attitude formation. Accuracy information can also lead to both explicit and to a lesser extent implicit attitude change (Petty et al., 2007). Smith, De Houwer, and Nosek (2013) found that features of persuasive information (i.e., source credibility) matter for implicit attitude change. These findings support my conclusions that propositional processes impact implicit attitudes less than explicit attitudes and that impact occurs when the meaning of associations are drastically altered by the propositional knowledge.

**Sympathy and Implicit Attitudes.** The research showing that sympathy and negative implicit attitudes can coexist is somewhat difficult to reconcile with this dissertation (e.g., Frantz, et al., 2004; Uhlmann, Brescoll, & Paluck, 2006; Andreychik & Gill 2012). For instance, Andreychik and Gill (2011), found that external attributions for poor circumstances predict negative implicit attitudes. In my studies, external attributions of negative events that caused sympathy were likely to influence implicit attitude

formation by decreasing, not increasing implicit negativity. My findings seem more in line with research showing that sympathy relating to external attribution decreases implicit negativity toward outgroups (e.g., Galinksy & Moskowitz, 2000; Huntsinger, Sinclair & Clore, 2009). Propositional processes seem to elicit emotional responses that impact implicit attitudes. In my dissertation studies, the attitudinal effects of the propositional processes were smaller than the emotion or attribution effects. However, sympathy and external attributions consistently decrease implicit negativity.

**Future Directions**

The studies reviewed here in combination with the existing literature only begin to address the question of when propositional processes matter in implicit attitude formation and change. Several lines of research could further inform these issues discussed in this dissertation.

One line of research could use a within subjects design to investigate the influence of attribution. Participants would encounter multiple targets described with same positive or negative information ordered to create internal or attributions. This study would provide direct comparisons of the two conditions present in many of my studies instead of a comparison to a control person. This design would also increase the impact of the attribution manipulation by making the internal/external distinctions obvious to participants. I predict that participants would show both implicit and explicit preferences for the person with externally attributed opposed to internally attributed negative events. Participants will prefer the person whose positive actions are attributed internally more than the person whose positive actions are attributed externally.

We could also learn more about how attributional processes affect associations by implementing a time delay between attribution manipulations and measurement. This would reveal whether propositional processes have a lasting impact on implicit attitudes. At Time 1, manipulations that create external or internal attributions with the same two pieces of information in opposite causal orders could be used to induce implicit attitude differences. A few days later, participants would be reminded of the two individual pieces of information (without the causal order), and then implicit and explicit attitudes would be measured a second time. Predictions for this research line are less clear. Perhaps propositional knowledge in the form of attribution has a limited impact over time and attitudes will converge on the associative information. However, whether that effect would happen because of a feature of attributional processing or because the attribution was forgotten would need to be determined.

**Conclusion**

The purpose of this dissertation was to test if propositional knowledge in the form of attribution can impact implicit attitudes. The evidence provided initial support for this conclusion. Under some conditions, however, attributions failed to impact implicit attitudes. Placed in a broader context, implicit evaluations of people can sometimes go beyond the good and bad events that occur, and also consider the causes, context, and meaning of those events.

**References**

Andreychik, M. R., & Gill, M. J. (2012). Do negative implicit associations indicate negative attitudes? Social explanations moderate whether ostensible "negative" associations are prejudice-based or empathy-based. *Journal of Experimental Social Psychology, 48(5),* 1082-1093.

Baeyens, F., Eelen, P., & Van den Bergh, O. (1990). Contingency awareness in evaluative conditioning: A case for unaware affective evaluative learning. *Cognition & Emotion, 4,* 3-18.

Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2004). No place for nostalgia in science: A response to Arkes & Tetlock. *Psychological Inquiry, 15(4),* 279-289.

Batson, C. D., Turk, C. L., Shaw, L. L., & Klein, T. R. (1995). Information function of empathic emotion: Learning that I value the other's welfare. *Journal of Personality and Social Psychology, 68(2),* 300.

Bem, D. J. (1972). Self-perception theory. *Advances in experimental social psychology, 6,* 1-62.

Castelli, L., Carraro, L., Gawronski, B., & Gava, K. (2010). On the determinants of implicit evaluations: When the present weighs more than the past. *Journal of Experimental Social Psychology, 46(1)*, 186-191.

De Houwer, J. (2006). Using the implicit association test does not rule out an impact of conscious propositional knowledge on evaluative conditioning. *Learning & Motivation, 37,* 176-187.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. Spanish *Journal of Psychology, 10,* 230-241.

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior, 37(1),* 1-20.

Dovidio, J. F., Johnson, J. D., Gaertner, S. L., Pearson, A. R., Saguy, T., & Ashburn-Nardo, L. (2010). Empathy and intergroup relations. *Prosocial motives, emotion, and behavior: The better angels of our nature,* 393-408.

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of personality and social psychology, 82(6),* 878.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences, 11(2),* 77-83.

Frantz, C. M., Cuddy, A. J., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The race implicit association test as a stereotype threat experience. *Personality and Social Psychology Bulletin, 30(12),* 1611-1624.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology, 78,* 708-724.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132,* 692-731.

Gawronski, B., & Strack, F. (2004). On the propositional nature of cognitive consistency: Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology, 40(4),* 535-542.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102,* 4-27.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological bulletin, 117(1),* 21.

Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey, (Eds.), *The handbook of social psychology* (4th ed.; pp. 89-150). New York: McGraw Hill.

Gill, M. J., & Andreychik, M. R. (2007). Explanation and intergroup emotion: Social explanations as a foundation of prejudice-related compunction. *Group processes & intergroup relations, 10(1)*, 87-106.

Gill, M. J., & Andreychik, M. R. (2009). Getting emotional about explanations: Social explanations and social explanatory styles as bases of prosocial emotions and intergroup attitudes. *Social and Personality Psychology Compass, 3(6)*, 1038-1054.

Greenwald, A. G., McGhee, D. E., Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74,* 1464-1480.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of personality and social psychology, 97(1),* 17.

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: asymmetry in the malleability of implicit preferences. *Journal of personality and social psychology, 90(1),*1-20.

Hassin, R. R. (2013). Yes it can on the functional abilities of the human unconscious. *Perspectives on Psychological Science, 8(2),* 195-207.

Huntsinger, J. R., Sinclair, S., & Clore, G. L. (2009). Affective regulation of implicitly measured stereotypes and attitudes: Automatic and controlled processes. *Journal of Experimental Social Psychology, 45,* 560-566.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions the attribution process in person perception. *Advances in experimental social psychology, 2,* 219-266.

Kelley, H. H. (1967). Attribution theory in social psychology. In *Nebraska symposium on motivation*. University of Nebraska Press.

Malle, B. F. (2011). Time to Give Up the Dogmas of Attribution: An Alternative Theory of Behavior Explanation. *Advances in experimental social psychology, 44(1),* 297-311.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134,* 565-584.

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2012). Implicit social cognition. In S. Fiske & C. N. Macrae (Eds.), *Handbook of Social Cognition* (pp. 31-53). New York, NY: Sage.

Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the Implicit Association Test: implicit and explicit attitudes are related but distinct constructs. *Experimental psychology, 54(1),* 14.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12,* 413– 417.

Olson, M. A., & Fazio, R. H. (2009). Implicit and explicit measures of attitudes. *Attitudes: Insights from the new implicit measures*, 19-63.

Olson, M. A., Kendrick, R. V., & Fazio, R. H. (2009). Implicit learning of evaluative vs. non-evaluative covariations: The role of dimension accessibility. *Journal of Experimental Social Psychology, 45(2)*, 398-403.

Petty, R. E., Tormala, Z. L., Briñol, P., & Jarvis, W. B. G. (2006). Implicit ambivalence from attitude change: An exploration of the PAST model. *Journal of Personality and Social Psychology, 90,* 21-41.

Ranganath, K. A., & Nosek, B. A. (2008). Implicit attitude generalization occurs immediately; explicit attitude generalization takes time. *Psychological Science, 19,* 249-254.

Ratliff, K. A., & Nosek, B. A. (2011). Negativity and outgroup biases in attitude formation and transfer. *Personality and Social Psychology Bulletin, 37,* 1692-1703.

Rudman, L. A. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science, 13(2),* 79-82.

Shanks, D. R., & Darby, R. J. (1998). Feature- and rule-based generalization in human associative learning. *Journal of Experimental Psychology: Animal Behavior Processes, 24,* 405-415.

Smith, C. T., De Houwer, J., & Nosek, B. A. (2013). Consider the Source Persuasion of Implicit Evaluations Is Moderated by Source Credibility. *Personality and Social Psychology Bulletin, 39(2),* 193-205.

Sriram, N., & Greenwald, A.G. (2009). The brief implicit association test. *Experimental Psychology, 56(4),* 283-294.

Stewart, T. L., Latu, I. M., Kawakami, K., & Myers, A. C. (2010). Consider the situation: Reducing automatic stereotyping through Situational Attribution Training. *Journal of Experimental Social Psychology, 46(1),* 221-225.

Todd, A. R., Bodenhausen, G. V., Richeson, J. A., & Galinsky, A. D. (2011). Perspective taking combats automatic expressions of racial bias. *Journal of Personality and Social Psychology, 100,* 1027-1042.

Trope, Y. (1989). Levels of inference in dispositional judgment. *Social Cognition, 7(3),* 296-314.

Uhlmann, E. L. Brescoll, V. L., & Paluck, E. L. (2006). Are members of low status groups perceived as bad, or badly off? Egalitarian negative associations and automatic prejudice. *Journal of Experimental Social* Psychology, 42, 491-499.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6,* 101–115.

Vescio, T. K., Sechrist, G. B., & Paolucci, M. P. (2003). Perspective taking and prejudice reduction: The mediational role of empathy arousal and situational attributions. *European Journal of Social Psychology, 33(4),* 455-472.

Weiner, B. (2001). Responsibility for social transgressions: An attributional analysis. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 331–344). Cambridge, MA: The MIT Press.

Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. M. (1972). Perceiving the causes of success and failure. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 95-120). Morristown, NJ: General Learning Press.

Weiner, B., Perry, R. P., & Magnusson, J. (1988). An attributional analysis of reactions to stigma. *Journal of Personality and Social Psychology, 55,* 738-748.

## Appendix

Pictures of the Targets/Controls


Nathan


Henry


Alice


Janet

**Table 1.** Means and Standard Deviations for Implicit and Explicit attitudes by condition and scenario set in Study 1

| **Condition** | **Explicit** | | **Implicit** | |
|---|---|---|---|---|
| | M | SD | M | SD |
| *Overall* | | | | |
| Negative 1 | $0.015_{abc}$ | 0.726 | $-0.042_{ab}$* | 0.414 |
| Negative 2 | $-0.064_{bd}$ | 0.695 | $-0.029_{ab}$ | 0.417 |
| Internal: 1 because 2 | $-0.055_{ad}$ | 0.647 | $-0.089_{a}$* | 0.408 |
| External: 2 because 1 | $0.063_{c}$* | 0.627 | $0.005_{b}$ | 0.416 |
| *Prostitution* | | | | |
| Negative 1: Can't feed kids | $-0.076_{a}$ | 0.675 | $-0.033_{abc}$ | 0.414 |
| Negative 2: Prostitute | $0.028_{a}$ | 0.638 | $0.032_{ab}$ | 0.400 |
| Internal: 1 because 2 | $-0.046_{a}$ | 0.566 | $-0.077_{c}$* | 0.387 |
| External: 2 because 1 | $0.014_{a}$ | 0.591 | $0.014_{b}$ | 0.426 |
| *Cancer* | | | | |
| Negative 1: Cancer | $0.097_{a}$ | 0.761 | $-0.052_{abc}$ | 0.415 |
| Negative 2: Uses illegal drugs | $-0.168_{b}$* | 0.744 | $-0.090_{a}$* | 0.426 |
| Internal: 1 because 2 | $-0.062_{b}$ | 0.713 | $-0.102_{ab}$* | 0.431 |
| External: 2 because 1 | $0.109_{a}$* | 0.659 | $-0.004_{c}$ | 0.405 |

*Note*. Means with the same subscript are not significantly different from one another (LSD within the ANOVA by condition) at $p < 0.05$. An asterisk indicates means significantly different from zero.

**Table 2.** Means and standard deviations for implicit and explicit attitudes by condition and induction type in Study 2a.

| **Condition** | **Explicit** | | | | **Implicit** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Time 1** | | **Time 2** | | **Time 1** | | **Time 2** | |
| | M | SD | M | SD | M | SD | M | SD |
| *Dual induction* | | | | | | | | |
| Can't feed / Internal | -0.147* | 1.336 | -1.029* | 1.722 | $-0.030_{b}$ | 0.418 | -0.098* | 0.398 |
| Prostitute / External | -0.230* | 1.359 | $0.072_{a}$ | 1.253 | $-0.065_{b}$* | 0.411 | $-0.024_{b}$ | 0.368 |
| *Single Induction* | | | | | | | | |
| Internal attribution | | | -0.451* | 1.432 | | | -0.214* | 0.381 |
| External attribution | | | $0.013_{a}$ | 1.343 | | | -0.114* | 0.422 |

*Note*. Means with the same subscript are not significantly different from one another (within Explicit/Implicit) at $p < 0.05$. An asterisk indicates means significantly different from zero.

**Table 3.** Means and standard deviations for implicit and explicit attitudes by Time 1 information, condition, and induction type in Study 2b.

| **Condition** | **Explicit** | | | | **Implicit** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Time 1** | | **Time 2** | | **Time 1** | | **Time 2** | |
| | M | SD | M | SD | M | SD | M | SD |
| *Dual induction by Time 1 information* | | | | | | | | |
| Cancer / Internal | $0.073_{a}$ | 1.037 | $-0.806_{c}$* | 1.599 | $0.008_{ac}$ | 0.380 | $-0.049_{cd}$ | 0.380 |
| Drugs / Internal | $-0.504_{bd}$* | 1.340 | $-0.610_{cd}$* | 1.455 | $-0.117_{be}$* | 0.444 | $-0.165_{b}$* | 0.420 |
| Cancer / External | $0.224_{a}$* | 1.168 | -0.282* | 1.324 | $-0.001_{af}$ | 0.439 | $-0.021_{dfh}$ | 0.382 |
| Drugs / External | $-0.435_{b}$* | 1.409 | 0.216* | 1.144 | $-0.067_{eg}$ | 0.408 | $-0.021_{gh}$ | 0.386 |
| *Dual induction* | | | | | | | | |
| Overall Internal | $-0.210_{a}$* | 1.229 | -0.711* | 1.532 | $-0.052_{a}$* | 0.416 | -0.105* | 0.403 |
| Overall External | $-0.122_{ab}$* | 1.340 | $-0.023_{bc}$ | 1.257 | $-0.035_{ab}$ | 0.423 | $-0.021_{b}$ | 0.378 |
| *Single Induction* | | | | | | | | |
| Internal | | | $-0.295_{d}$* | 1.017 | | | -0.224* | 0.413 |
| External | | | $-0.122_{cd}$ | 1.289 | | | -0.123* | 0.371 |

*Note*. Means with the same subscript are not significantly different from one another (within Explicit/Implicit) at $p < 0.05$. Marginal means (overall) and the Single Induction means are compared separately. An asterisk indicates means significantly different from zero.

**Table 4.** Correlations between measures in Study 3.

| Variables | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Overall BIAT | – | | | | | |
| 2. Good BIAT | – | – | | | | |
| 3. Bad BIAT | – | – | – | | | |
| 4. Explicit Composite | .15*** | .16*** | .15*** | – | | |
| 5. Explicit Positivity | .15*** | .12*** | .13** | .85*** | – | |
| 6. Explicit Negativity | -.12*** | -.10* | -.14** | -.85*** | -.46*** | – |
| 7. Attribution Condition | .04 | .02 | .05 | .17*** | .12*** | .16*** |

*Note*. All measures are comparisons between the target and control. All measures are coded such that positive numbers indicate a preference for the target expect for Explicit Negativity. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

**Table 5.** Means and standard deviations for implicit and explicit attitudes by condition and association type in Study 3.

| **Condition** | **Explicit** | | **Implicit** | |
|---|---|---|---|---|
| | M | SD | M | SD |
| Internal Attribution | -0.850* | 2.056 | -0.045 $_{a}$* | 0.331 |
| External Attribution | -0.351* | 1.778 | -0.038$_{a}$* | 0.325 |
| *Positive Associations* | | | | |
| Internal Attribution | -0.395* | 1.167 | 0.002 $_{b}$ | 0.334 |
| External Attribution | -0.137* | 1.085 | -0.004$_{b}$ | 0.324 |
| *Negative Associations* | | | | |
| Internal Attribution | 0.451* | 1.246 | -0.093 $_{c}$* | 0.321 |
| External Attribution | 0.223* | 0.998 | -0.072 $_{c}$* | 0.324 |

*Note*. Positive Associations were measured by a Good focal brief IAT and felt positivity. Negative Associations were measured by a Bad focal brief IAT and felt negativity. The top means are post-hoc combinations of the other implicit and explicit measures. Means with the same subscript are not significantly different from one another (within measure) at $p < 0.05$. An asterisk indicates means significantly different from zero.

**Table 6.** Causal order text and predictions for attitude formation in Study 4.

| **Condition** | | | | |
|---|---|---|---|---|
| **Attribution** | **Scholarship** | **Grades** | **Text** | **Attitude Prediction** |
| External | Got (+) | Good (+) | Nathan got good grades because Nathan got a scholarship. | Positive |
| Internal | Got (+) | Good (+) | Nathan got a scholarship because Nathan got good grades. | Very Positive |
| External | Got (+) | Bad (-) | Nathan got a scholarship even though Nathan got bad grades. | Slightly Negative |
| Internal | Got (+) | Bad (-) | Nathan got bad grades even though Nathan got a scholarship. | Negative |
| External | Lost (-) | Good (+) | Nathan lost a scholarship even though Nathan got good grades. | Neutral |
| Internal | Lost (-) | Good (+) | Nathan got good grades even though Nathan lost a scholarship. | Slightly Positive |
| External | Lost (-) | Bad (-) | Nathan got bad grades because Nathan lost his scholarship. | Negative |
| Internal | Lost (-) | Bad (-) | Nathan lost his scholarship because Nathan got bad grades. | Very Negative |

*Note*. Attitude predictions are for Explicit and Implicit Attitudes given the assumption that both are influenced by propositional knowledge (i.e., attribution).

**Table 7.** Model predicting attribution item ratings from condition in Study 4.

| **Variable** | ***F*** | ***p*** | $\eta^2$ |
|---|---|---|---|
| Scholarship (Got vs. Lost) | 27.49 | < 0.0001 | 0.010 |
| Grades (Good vs. Bad) | 3.86 | 0.049 | 0.001 |
| Attribution (External vs. Internal) | 263.46 | < 0.0001 | 0.093 |
| Scholarship*Grades | 257.32 | < 0.0001 | 0.090 |
| Attribution*Scholarship | 14.78 | 0.0001 | 0.005 |
| Attribution*Grades | 12.36 | 0.0004 | 0.004 |
| Scholarship*Grades* Attribution | 28.79 | < 0.0001 | 0.010 |

*Note*. Values from Type III sum of squares.

**Table 8.** Attribution Means by condition in Study 4.

| Condition | | | Attribution | |
|---|---|---|---|---|
| **Attribution** | **Scholarship** | **Grades** | **M** | **SD** |
| External | Got (+) | Good (+) | $-0.984_{af}$ | 1.757 |
| Internal | Got (+) | Good (+) | $-1.725_{b}$ | 1.551 |
| External | Got (+) | Bad (-) | $0.413_{c}$ | 1.735 |
| Internal | Got (+) | Bad (-) | $-0.588_{d}$ | 1.510 |
| External | Lost (-) | Good (+) | $1.161_{e}$ | 1.857 |
| Internal | Lost (-) | Good (+) | $-0.875_{af}$ | 1.763 |
| External | Lost (-) | Bad (-) | $-0.454_{d}$ | 1.692 |
| Internal | Lost (-) | Bad (-) | $-1.241_{a}$ | 1.446 |

*Note*. Negative numbers indicate the attribution was more internal than external. Means with the same subscript are not significantly different from one another.

**Table 9.** Marginal means by condition for explicit (liking difference) and implicit attitudes.

| Condition | Explicit | | Implicit | |
|---|---|---|---|---|
| | **M** | **SD** | **M** | **SD** |
| All Internal | $-0.092_{a}$* | 1.175 | $-0.009_{a}$ | 0.414 |
| All External | $-0.182_{b}$* | 1.165 | $-0.042_{a}$* | 0.414 |
| All Got Scholarship | $-0.125_{c}$* | 1.191 | $-0.013_{b}$ | 0.406 |
| All Lost Scholarship | $-0.150_{c}$* | 1.142 | $-0.039_{b}$* | 0.423 |
| All Good Grades | $0.057_{d}$ | 1.064 | $0.033_{c}$* | 0.419 |
| All Bad Grades | $-0.328_{e}$* | 1.228 | $-0.084_{d}$* | 0.403 |

*Note*. Means with the same subscript are not significantly different from one another (within Explicit/Implicit) at $p < 0.05$. Means with an asterisk are significantly different from 0.

**Table 10.** Model predicting the difference in liking between target and control from the conditions and their interactions in Study 4.

| **Variable** | ***F*** | ***p*** | $\eta^2$ |
|---|---|---|---|
| Scholarship (Got vs. Lost) | 0.19 | 0.618 | 0.0001 |
| Grades (Good vs. Bad) | 59.64 | < 0.0001 | 0.028 |
| Attribution (External vs. Internal) | 4.12 | 0.043 | 0.002 |
| Scholarship*Grades | 2.92 | 0.088 | 0.001 |
| Attribution*Scholarship | 0.33 | 0.564 | 0.0002 |
| Attribution*Grades | 0.27 | 0.601 | 0.0001 |
| Scholarship*Grades* Attribution | 1.62 | 0.204 | 0.001 |

*Note*. Values from Type III sum of squares.

**Table 11.** Means for explicit (difference in liking) and implicit attitudes by condition.

| Condition | | | Explicit | | Implicit | |
|---|---|---|---|---|---|---|
| **Attribution** | **Scholarship** | **Grades** | **M** | **SD** | **M** | **SD** |
| External | Got (+) | Good (+) | $0.004_{a}$ | 0.910 | $0.003_{a}$ | 0.415 |
| Internal | Got (+) | Good (+) | $0.226_{b}$* | 1.253 | $0.100_{b}$* | 0.039 |
| External | Got (+) | Bad (-) | $-0.382_{c}$* | 1.352 | $-0.075_{c}$* | 0.408 |
| Internal | Got (+) | Bad (-) | $-0.341_{c}$* | 1.133 | $-0.073_{c}$* | 0.384 |
| External | Lost (-) | Good (+) | $-0.011_{a}$ | 1.024 | $0.005_{a}$ | 0.419 |
| Internal | Lost (-) | Good (+) | $0.024_{a}$ | 1.055 | $0.036_{a}$ | 0.443 |
| External | Lost (-) | Bad (-) | $-0.353_{c}$* | 1.251 | $-0.105_{c}$* | 0.407 |
| Internal | Lost (-) | Bad (-) | $-0.242_{c}$* | 1.176 | $-0.079_{c}$* | 0.412 |

*Note*. Means with the same subscript are not significantly different from one another (within Explicit/Implicit) at $p < 0.05$. Means with an asterisk are significantly different from 0.

**Table 12.** Model predicting implicit preference for the target in Study 4.

| **Variable** | ***F*** | ***p*** | $\eta^2$ |
|---|---|---|---|
| Scholarship (Got vs. Lost) | 1.59 | 0.207 | 0.001 |
| Grades (Good vs. Bad) | 37.68 | < 0.0001 | 0.021 |
| Attribution (External vs. Internal) | 4.06 | 0.044 | 0.002 |
| Scholarship*Grades | 0.13 | 0.723 | 0.0001 |
| Attribution*Scholarship | 0.28 | 0.594 | 0.0002 |
| Attribution*Grades | 1.69 | 0.194 | 0.001 |
| Scholarship*Grades* Attribution | 1.32 | 0.251 | 0.001 |

*Note*. Values from Type III sum of squares.

**Table 13.** Attribution means by condition and induction type for Study 5.

| Condition | Time 1 | | Time 2 | |
|---|---|---|---|---|
| | **M** | **SD** | **M** | **SD** |
| *Dual Induction* | | | | |
| Positive: Charity | -0.568* | 1.523 | $2.718*_{b}$ | 0.857 |
| Positive: Storm Relief | -0.133 | 1.701 | $2.682*_{b}$ | 0.861 |
| Negative: Party | $-0.312*_{a}$ | 1.724 | $2.706*_{b}$ | 0.985 |
| Negative: Meth | -1.500* | 1.412 | 2.284* | 1.357 |
| Overall Positive | $-0.356*_{a}$ | 1.625 | $2.700*_{b}$ | 0.858 |
| Overall Negative | -0.858* | 1.693 | $2.517*_{b}$ | 1.183 |
| *Single Induction* | | | | |
| Positive: Charity | | | $2.331*\dagger_{a}$ | 1.411 |
| Positive: Storm Relief | | | $2.312*\dagger_{a}$ | 1.345 |
| Negative: Party | | | 2.653* | 0.911 |
| Negative: Meth | | | $2.328*_{a}$ | 1.409 |
| Overall Positive | | | $2.310*\dagger_{a}$ | 1.378 |
| Overall Negative | | | $2.485*_{a}$ | 1.205 |

*Note*. Means with the same subscript are not significantly different from one another (within Induction Type) at $p < 0.05$. Means with an asterisk are significantly different from 0. A † indicates a significant difference from the corresponding dual induction mean.

**Table 14.** Means for explicit (difference in liking) and implicit attitudes by condition and induction type.

| Condition | Explicit | | | | Implicit | | | |
|---|---|---|---|---|---|---|---|---|
| | Time 1 | | Time 2 | | Time 1 | | Time 2 | |
| | M | SD | M | SD | M | SD | M | SD |
| *Dual Induction* | | | | | | | | |
| Positive: Charity | 0.141 | 1.122 | -0.224*$_{b}$ | 1.000 | -0.002$_{ac}$ | 0.393 | -0.036$_{cd}$ | 0.379 |
| Positive: Storm | 0.275*$_{a}$ | 1.191 | -0.046$_{b}$ | 1.153 | -0.010$_{ae}$ | 0.413 | -0.003$_{de}$ | 0.377 |
| Negative: Party | -0.922* | 1.666 | 0.114 | 1.154 | -0.051$_{af}$ | 0.397 | 0.010$_{df}$ | 0.342 |
| Negative: Meth | -1.564* | 2.166 | -0.143 | 1.178 | -0.197*$_{b}$ | 0.400 | -0.056$_{d}$ | 0.367 |
| Overall Positive | 0.206* | 1.156 | 0.000*$_{a}$ | 1.170 | -0.006$_{a}$ | 0.403 | -0.020$_{ac}$ | 0.355 |
| Overall Negative | -1.207* | 1.929 | -0.136$_{a}$ | 1.139 | -0.118*$_{b}$ | 0.405 | -0.020$_{c}$ | 0.378 |
| *Single Induction* | | | | | | | | |
| Positive: Charity | | | -0.063$_{a}$ | 1.148 | | | 0.030$_{ac}$ | 0.404 |
| Positive: Storm | | | -0.200*$_{a}$ | 1.014 | | | -0.30$_{ac}$ | 0.434 |
| Negative: Party | | | 0.350*$_{b}$ | 1.155 | | | 0.066$_{b}$ | 0.398 |
| Negative: Meth | | | 0.171*†$_{b}$ | 1.118 | | | -0.000$_{bc}$ | 0.389 |
| Overall Positive | | | -0.128* | 1.087 | | | 0.001$_{a}$ | 0.419 |
| Overall Negative | | | 0.259*† | 1.139 | | | 0.032$_{a}$ | 0.394 |

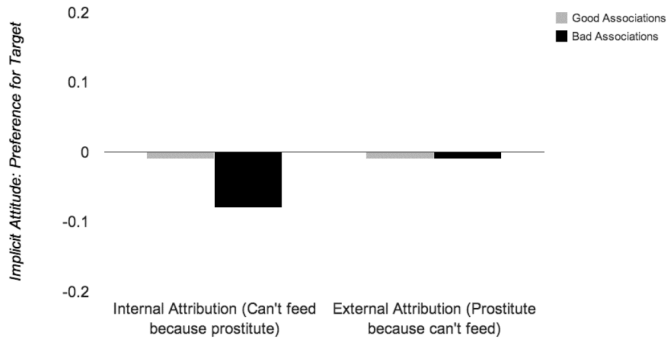
*Note*. Means with the same subscript are not significantly different from one another (within Explicit/Implicit and Induction Type) at $p < 0.05$. Means with an asterisk are significantly different from 0. A † indicates a significant difference from the corresponding dual induction mean.

**Table 15.** Results for explicit and implicit attitudes for all studies.

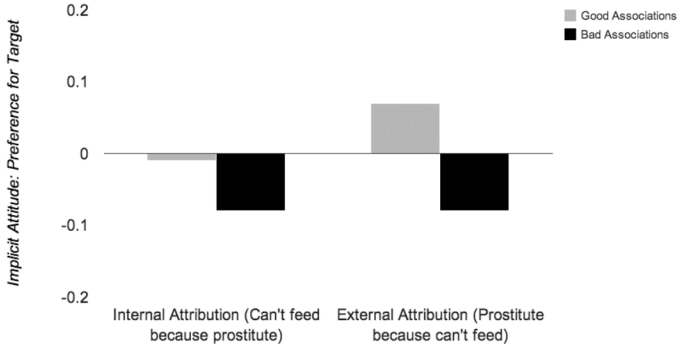
| | | **Explicit** | | **Implicit** | |
|---|---|---|---|---|---|
| **Study** | **Scenario/Method** | **Formation** | **Change** | **Formation** | **Change** |
| 1 | Prostitute & Cancer | $d = 0.18$* | -- | $d = 0.22$* | -- |
| 2a | Prostitute | $d = 0.33$* | $d = 0.74$* | $d = 0.25$* | $d = 0.19$* |
| 2b | Cancer | $d = 0.15$ | $d = 0.47$* | $d = 0.26$ | $d = 0.22$* |
| 3 | Prostitute/BIAT | $d = 0.26$* | -- | $d = 0.00$ | -- |
| 4 | Scholarship | 1 of 4: $d = 0.21$* | -- | 1 of 4: $d = 0.24$* | -- |
| 5 | House/External | $d = -0.12$*<br>$d = 0.23$* | $d = 0.30$*<br>$d = 0.59$* | $d = 0.02$†<br>$d = 0.08$† | $d = 0.01$<br>$d = 0.23$* |

Note. For Studies 1 through 4, effect sizes are condition comparisons; attitude change is defined as condition differences at Time 2. For Study 5, the first effect size is for the positive scenarios, and the second effect size is for the negative scenarios; formation is defined as Time 1 attitudes while change is the difference between Time 1 and Time 2. An asterisk indicates significances at $p < 0.05$. A † indicates that the attribution manipulation was successful by eliminating preferences.

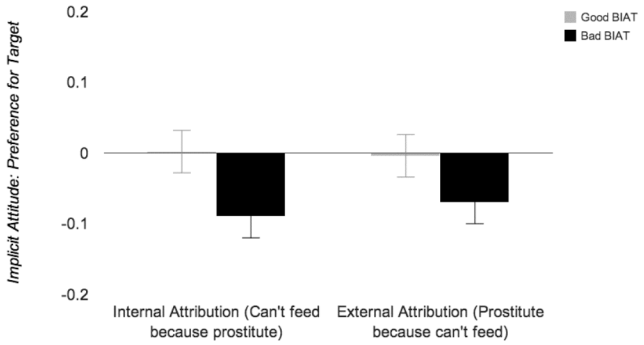
**Figure 1.** Possible influences of attributions of negative events on good and bad implicit associations and actual findings.




Alternative 1: External attributions suppress bad associations from being formed




Alternative 2: External attributions create good and bad associations




Actual Results: Internal and external attributions create equivalent associations