Impact of Implicit Association Experience and Motivation to Control Prejudice on Stigma

Kelly M. Shaffer
Distinguished Majors Thesis
University of Virginia
April 2011
Acknowledgement

I would like to extend the most sincere thanks to Dr. Fred Smyth and Professor Bethany Teachman for their extensive help and unfailing patience with me during my project from start to finish. I am so thankful to have had Fred’s guidance through BIAT programming and data management as well as his help with initial data analysis. To Professor Teachman, I am eternally grateful for her help with every stage of my thesis, with very many hours spent helping edit the project, guiding me through analysis of the data, and providing thorough feedback to help me improve my writing. Beyond help with my thesis, I have been so fortunate to have Professor Teachman as a trusted and respected advisor—I have sincerely appreciated your advice, thoughts, and encouragement over the past several years.
Abstract

The current study examined the effect of completing implicit association measures as an intervention to alter explicit stigma endorsement, when the intervention varies by mental illness content and implicit evaluation components. Volunteers (N=1549) aged 18-79 at an on-line research site completed either a mental illness, gay/straight, or non-evaluative Brief Implicit Association Test (BIAT) or were in a no intervention control condition before completing measures of explicit stigma toward mentally ill people and motivation to control prejudice. Contrary to hypotheses, intervention condition did not affect explicit stigma endorsement. However, as predicted, internal motivation to control prejudice did interact with intervention condition to alter explicit stigma endorsement. The direction of the effect was unexpected: Participants who were low in internal motivation endorsed significantly more stigma in the gay/straight BIAT condition compared to those in the non-evaluative BIAT condition and the mental illness condition (at the level of a nonsignificant trend). Results suggest the BIAT’s use as a psychoeducational measure to reduce stigma may be more limited than expected.
Impact of Implicit Association Experience and Motivation to Control Prejudice on Stigma

Stigma, the bias against a characteristic perceived as different, leads to the social devaluation of persons regarded as outside the social norm (Goffman, 1963; Hinshaw & Stier, 2008). Stigma can manifest as anything from interference in social interaction and increased social distance from others (Albrecht, Walker & Levy, 1982; Norman et al., 2010) to decreased housing and employment opportunities (Hinshaw & Stier, 2008). People with mental illness are frequently targets of stigma from society at large, often stereotyped as dangerous or responsible for their condition (Link, Phelan, Bresnahan, Stueve, & Pescosolido, 1999). Felt stigma has been shown to have deleterious effects on persons with mental illness, decreasing help-seeking behavior (Schomerus, Matschinger, & Angermeyer, 2008; Vogel, Wade, & Haake, 2006) and predicting poorer outcomes of mental illness even when disability is held constant (Hinshaw & Stier, 2008). This project aims to investigate a novel potential intervention—experience with implicit bias measurement—as a tool to reduce explicit bias against persons with mental illness.

Even among people who are not consciously biased, deeply engrained negative bias from socially constructed stereotypes can be automatically activated and result in discriminatory behavior. The Self-regulation of prejudice model (SRP; Monteith, 1993; Monteith, Mark, & Ashburn-Nardo, 2010) suggests that when a person realizes they have responded to a situation in a stigmatizing way that runs contrary to their desire to be non-prejudiced, cues for prejudice control are generated to check future discriminatory responses. For instance, if a person catches herself walking quickly past a homeless person because she assumes that person is mentally ill and therefore dangerous, the SRP model predicts this will evoke guilt and agitation if detected as being contrary to her desire to be non-prejudiced. Cues for prejudice control are thus formed in order to avoid revisiting the negative, self-punitive feelings following responses discrepant with
one’s values. Applied to stigma intervention, the SRP model implies that interventions that illuminate one’s implicit biases toward a stigmatized minority group and show how these biases may affect one’s behavior may generate cues for control, reducing explicit stigma.

**The Brief Implicit Association Test as a Stigma Intervention**

Recent evidence by Menatti, Smyth, Nosek, and Teachman (under review) suggests this implicit bias education may be achieved through completing a Brief Implicit Association Test (BIAT; Sriram & Greenwald, 2009). In this computer-based measure of implicit biases, one categorizes stimuli, such as “good” and “bad” words, with phrases about the stigmatized group being studied. Response latencies for categorizing the stimuli constitute the measure of strength of automatic associations in memory. Menatti et al. showed that simply taking one of these short tests was associated with less explicitly endorsed stigma. In their first study, participants who completed a BIAT measuring implicit evaluations of mentally ill people before completing an explicit attitude measure about this group reported significantly more positive attitudes than those who completed the tasks in the reverse order. In a follow up study, Menatti et al. examined whether receiving personalized feedback about one’s implicit biases was necessary to elicit the stigma-reduction effect. There was no significant difference in explicitly endorsed stigma between groups who received their personalized BIAT feedback message before completing the stigma questionnaire and those who did not receive their BIAT results.

These findings suggest that the experience of taking the BIAT itself, rather than receiving explicit feedback about one’s level of implicit stigma, impacted participants’ later conscious evaluations of mentally ill persons. The researchers hypothesized that the experience of taking the BIAT triggered feelings of “should-would discrepancy” (Devine, Monteith, Zuwerink, & Elliot, 1991). This phenomenon concerns a person’s reflection on what they think they should do
versus what they would do in a given situation. As the difference between the two grows, negative self-feelings and agitation-related emotions rise. While taking a BIAT, a person may “feel” their implicit biases through finding one category pairing condition more difficult than another, suggesting that the more difficult condition contradicts their automatic associations in memory.

For example, if a person felt himself taking longer to correctly sort “good” and “people with mental illness” stimuli together, he might come to understand he more typically associates mentally ill people with negative concepts. This may cause discomfort if he believes himself to be genuinely unbiased. The “should-would discrepancy” resulting from the experience of taking the BIAT is therefore hypothesized to activate the Self-regulation of prejudice process. “Feeling” one’s negative implicit biases could act as a discrepant response from the participant’s ideal non-prejudiced reaction, initiating self-reflection and the development of the cues for control, impeding later biased reactions. Regardless of valence of bias, it is possible that learning about the existence and effects of any implicit biases through the BIAT experience may encourage participants to be more aware of all of their biases and thus curb explicit endorsement of stigma.

**Mechanisms Underlying the BIAT’s Effect on Stigma**

The current study intends to evaluate the mechanisms involved in the reduction of explicit bias after taking a BIAT. The study was designed to examine whether the experience of completing an implicit bias assessment, exposure to mental illness content, or both in conjunction are sufficient to induce explicit bias change. If learning about one’s implicit bias initiates self-reflection and cues for prejudice control, it is possible that learning about one’s implicit biases toward any comparable stigmatized minority may successfully create widely generalizable cues for prejudice control. If the effect is more limited, one may need to learn
about one’s implicit biases toward the target stigmatized group specifically to elicit the effect for that group. The study included four conditions: A mental illness BIAT (exposing participants to both mental illness content and implicit bias assessment); a gay/straight BIAT (exposing participants to implicit bias assessment toward a stigmatized minority group but not to mental illness content); a non-evaluative mental illness pseudo-BIAT (exposing participants to mental illness content but not to implicit bias assessment); and a no intervention control (no exposure to mental illness content or implicit bias assessment), where participants completed the explicit stigma measure before completing a BIAT.

The gay/straight BIAT was chosen as an alternative to the mental illness BIAT because people have been shown to generalize empathy from members of one target group to those of another target group categorized under the same superordinate social category (Tarrant & Hadert, 2010). Evidence suggests that the public tends to consider both homosexuality and mental illness as subcategories of the broader “social disability” social category (Albrecht et al., 1982; Crandall, Eshleman, & O’Brien, 2002). We thus hypothesized that any “should-would discrepancy” activated by the gay/straight BIAT would similarly generalize to mentally ill people, a comparable socially stigmatized group, without participants being exposed to mental illness content directly.

A non-evaluative pseudo-BIAT was devised to remove the evaluation component from the BIAT task by using neutral circle and rectangle word stimuli instead of good or bad words, then pairing the circle and rectangle stimuli with the mentally ill people phrases. This removal of valence allowed participants to be exposed to the same BIAT task format, but only be primed to mental illness content without exposure to implicit bias assessment.
Mental illness thought content combined with implicit bias assessment in the mental illness BIAT condition is hypothesized to most directly activate cues for control of prejudice against persons with mental illness, and thus result in the lowest amount of endorsed stigma. Implicit bias assessment of a comparable social minority group (i.e., gay people) in the gay/straight BIAT condition is hypothesized to activate the SRP model as well, but not create as strong cues for prejudice control toward persons with mental illness, resulting in the second lowest endorsed stigma level. Priming of mental illness content alone in the non-evaluative pseudo-BIAT is not thought to necessarily activate the SRP model. However, as increased familiarity with out-groups has been shown to decrease negative evaluations of members of those out-groups (Gonsalkorale, Allen, Sherman & Klauer, 2010; Pettigrew & Tropp, 2006), we believe increased exposure to the mental illness target group through priming should lead to less negative explicit evaluations than the no intervention group, which provides no implicit bias or mental illness concept exposure.

**Moderators of the BIAT’s Effect on Stigma**

A second objective of the current study is to explore potential moderators of the BIAT effect, particularly participants’ motivation to control their prejudice. Plant and Devine (1998) describe two empirically distinct motivations to control prejudiced reactions: External and internal motivation. Persons high in external motivation seek to control themselves from making their biases explicit due to concern that an audience might react negatively to their prejudice. Those high in external motivation also tend to report higher levels of self-reported prejudiced attitudes than those who are internally motivated. Those with higher internal motivation have better-defined personal standards against expressing prejudice and judge themselves personally
for any prejudiced reactions. Persons high in internal motivation also tend to endorse lower levels of prejudice (Crandall et al., 2002; Plant & Devine, 1998).

In relation to the should-would discrepancy hypothesized to be at play in the BIAT’s effect on explicit stigma reduction, people high in internal motivation (as opposed to external motivation) report that they should be less prejudiced toward stigmatized minority groups. Those high in internal motivation also tend to react with strong guilt and compunction when they perceive incongruity between their “should” and “would” reactions (Devine et al., 1991). Therefore, for the current study, it was hypothesized that persons high (compared to low) in internal motivation to control prejudice would be more affected by the experience of “feeling” their implicit biases during the implicit bias assessment. This, in turn, should create stronger cues for prejudice control, leading to lower endorsed stigma against persons with mental illness.

This study examines two possible components of the BIAT experience that may alter explicit stigma endorsement—experience with an implicit bias measure and mental illness content—and considers how motivation to control prejudice may influence the impact of the BIAT on explicit stigma. As an easy to disseminate and quick intervention, the BIAT has considerable potential to be harnessed as a psychoeducational measure to reduce stigma. Understanding the mechanisms by which taking the BIAT reduces explicit bias and moderators of this effect may enhance our ability to reduce explicit stigma toward people with mental illness.

**Method**

**Participants**

Using the Project Implicit web platform for implicit attitudes research (http://implicit.harvard.edu), data were provided from 2597 volunteers, and analyzed from the
1549 participants with both a valid BIAT and explicit stigma measure score (see results for data reduction procedure). Participants for this study were limited to those 18 years of age and older and citizens of the United States. At Project Implicit, participants learn about implicit social biases by participating in research studies that include implicit bias measures and by receiving information and feedback on their implicit bias score after test completion. Participant samples are typically more heterogeneous than typical college convenience samples, but are not representative of any particular population.

Participants averaged 29 years of age, with 895 females and 651 males. Of those who reported ethnicity, 10.23% were Hispanic and 83.74% were non-Hispanic (ethnicity unknown for 6.03%). Race of participants was reported as: 0.65% American Indian/Alaska Native, 3.39% East Asian, 1.95% South Asian, 0.72% Native Hawaiian or other Pacific Islander, 8.79% Black/African American, 73.37% White/Caucasian, 1.04% More than one race—Black/White, 6.32% More than one race—Other, and 3.78% Other or Unknown. Participants tended to be highly educated, with 15.88% reportedly receiving a graduate degree, 27.45% receiving a bachelor’s degree or attending some graduate school, 47.12% receiving an associate’s degree or attending some college, 5.56% being high school graduates, and 3.99% having less than a high school degree. Participants also tended to report being liberal to neutral politically, with 13.93% self-identifying as strongly liberal, 22.60% moderately liberal, 9.60% somewhat liberal, and 31.67% neutral, while 3.53% self-reported as strongly conservative, 10.60% moderately conservative, and 8.27% somewhat conservative.

Materials

*Community Attitudes of the Mentally Ill—Social Restrictiveness Scale.* The CAMI-SR (Taylor & Dear, 1981) assesses stigma toward mentally ill persons. Only the 10-item social
restrictiveness subscale was used in the current study due to constraints on study length. This subscale evaluates beliefs about mentally ill people being a threat to society, including fear about them being dangerous (e.g., “The mentally ill are a danger to themselves and those around them”). This scale also measures desired social distance from persons with mental illness (e.g., “I would not want to live next door to someone who has been mentally ill”). Five response options range from Strongly Agree to Strongly Disagree, with Neutral in the middle. Responses were coded from one to five, with higher numbers indicating a greater propensity to stigmatize persons with mental illness. The CAMI-SR is a widely used index of attitudes toward mentally ill persons in society and has both high reliability and adequate construct validity (Taylor & Dear, 1981).

Motivation to Respond without Prejudice—Internal Motivation Scale and External Motivation Scale. The Motivation to Respond without Prejudice questionnaire (Plant & Devine, 1998) contains two scales to identify both internal and external motivations to not exhibit prejudice. The original scale evaluated prejudice toward African Americans, so for the current study, items were modified to read “person with mental illness” rather than “Black person.” The two scales show discriminant, convergent, and predictive validity, and have sound psychometric properties (Plant & Devine, 1998). A sample modified internal motivation item is “Being nonprejudiced toward mentally ill people is important to my self-concept;” a sample modified external motivation item is “I try to act nonprejudiced toward mentally ill people because of pressure from others.” Response options range from Strongly Agree to Strongly Disagree on a 9-point scale with Neither Agree Nor Disagree in the middle. After reverse coding items as necessary, higher scores on a scale indicate higher levels of the particular motivation (internal or external).
**Brief Implicit Association Test.** The Brief Implicit Association Test (BIAT; Sriram & Greenwald, 2009) is a modified version of the original Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). Unlike the IAT, the BIAT does not use an explicitly labeled comparison category (i.e., there is no explicit category label to contrast with mentally ill people). The task is designed to measure association strength between target constructs by analyzing latency in categorization responses. During the critical combined task blocks, the BIAT instructions ask participants to categorize two groups of target stimuli as “IN” (e.g., *Flowers* and *Good*; “Press the ‘I’ key when you see representative items”), while all other items are to be categorized as “OUT” with the ‘E’ key. If a participant classifies exemplars from one category pairing (e.g., *Flowers* and *Good*) faster than those from another category pairing (e.g., *Insects* and *Good*), it is inferred that the participant has a stronger implicit association between flowers and positive evaluation than between insects and positive evaluation. In the current study, we maintained the standard seven-phase structure of an IAT (Nosek, Greenwald, & Banaji, 2007), but used fewer trials per block. For an example of the BIAT block organization used in the current study, see Table 1. Three different BIATs were used for the experimental conditions: Mental illness BIAT, gay/straight BIAT, and mental illness pseudo-BIAT.

**Mental illness BIAT.**

For the mental illness BIAT, the target group is *Mentally Ill People*, which is paired with the descriptor categories *Good* or *Bad*. *Physically Ill People* items served as background stimuli to the *Mentally Ill People* stimuli. Whether participants received the pairing of *Mentally Ill People* and *Good* categories versus *Mentally Ill People* and *Bad* categories first for the critical blocks was randomly assigned to reduce order effects. *Mentally Ill People* stimuli are “person with anxiety,” “person with depression,” and “person with schizophrenia;” *Physically Ill People* items...
stimuli are “person with appendicitis,” “person with cancer,” and “person with pneumonia.”

Good stimuli are “excellent,” “great,” and “wonderful;” Bad stimuli are “horrible,” “nasty,” and “terrible.” The emphasis on the person in both the category labels and stimuli was intended to place focus on the person with the illness in question, rather than the illness itself, which is more objectively negative.

**Gay/straight BIAT.**

The gay/straight BIAT has the same format as the mental illness BIAT, where participants are asked to classify exemplars from one focal pairing (i.e., Gay People with either Good or Bad stimuli) as “IN,” while all other stimuli are classified as “OUT.” As with the mental illness BIAT, the order of pairing the category Gay People with Good versus with Bad was randomized. Straight People stimuli served as background stimuli for the Gay People target category. Stimuli for Gay People include pictures of wedding cake toppers and restroom sign figures with two males or two females, while stimuli for Straight People include pictures of wedding cake toppers and restroom sign figures with both a male and female (following Nosek, Greenwald, & Banaji, 2005). Stimuli for the Good and Bad categories are the same as those used in the mental illness BIAT as listed above.

**Mental illness non-evaluative pseudo-BIAT.**

The non-evaluative mental illness pseudo-BIAT uses the target category of Mentally Ill People (and Physically Ill People background stimuli, using the same stimuli for both categories as used in the valid mental illness BIAT), but removes the implicit bias assessment component of the BIAT by replacing the evaluative Good and Bad categories with neutral Circle and Rectangle descriptor categories. Circle stimuli include “curve,” “ring,” and “oval;” Rectangle stimuli include “edge,” “corner,” and “box.” By removing the evaluation of pairing mentally ill people
with good or bad, this pseudo-BIAT acts as a prime to mental illness concepts without providing participants with the experience of an implicit *evaluation* assessment.

BIAT *D* scores were calculated following Sriram and Greenwald (2009) and Greenwald, Nosek, and Banaji (2003). The *D* score can be interpreted similar to a Cohen’s *d* effect size, and reflects the difference in classification time across the two critical category pairing conditions (e.g., *Mentally Ill People* and *Good* versus *Mentally Ill People* and *Bad*) divided by the participant’s response latency pooled standard deviation. The task was scored so that higher scores indicate more negative implicit associations toward people with mental illness or gay people, depending on the task.

**Procedures**

Following informed consent, participants completed an Instructional manipulation check, where participants demonstrate whether they have fully read the opening instructions paragraph by clicking on the “Instructions” header rather than on the more obvious “Continue” button at the bottom of the page (see Oppenheimer, Meyvis, & Davidenko, 2009). This task was included to evaluate whether participants were reading all presented materials carefully. Next, participants are randomly assigned to one of the four conditions: Mental illness BIAT, gay/straight BIAT, non-evaluative mental illness pseudo-BIAT, or no intervention control.

In the three experimental conditions, participants complete the BIAT specific to their intervention condition and then proceed to complete the CAMI-SR, ten items from the Positive and Negative Affect Schedule—Extended Version\(^1\) (PANAS-X; Watson & Clark, 1994), and internal and external motivation to control prejudice questionnaires. The no intervention control condition participants do not complete a BIAT before responding to the questionnaires. Given

---

\(^1\) This measure was included as part of a larger study, and is included here only as part of a secondary analysis described in the Discussion.
that all participants visiting the Project Implicit research portal expect an implicit association test experience and personalized test feedback, participants in the mental illness pseudo-BIAT and the no intervention control groups receive the valid mental illness BIAT after completing all study-relevant measures.

Once participants in the mental illness BIAT and gay/straight BIAT conditions complete the questionnaires and participants in the non-evaluative pseudo-BIAT and control conditions complete the mental illness BIAT, a personalized feedback screen about the participant’s BIAT performance and overview information on implicit bias is presented. Participants are then debriefed.

**Results**

**Sample Characteristics**

We evaluated intervention condition differences at baseline to ensure accuracy of random assignment across the four conditions. Chi-square tests revealed no significant differences between conditions on gender ($\chi^2(3) = 2.05, p = .56$), ethnicity ($\chi^2(6) = 11.03, p = .09$), education ($\chi^2(12) = 7.30, p = .84$), or political identification ($\chi^2(18) = 11.49, p = .87$). To simplify analyses, race categories that fewer than 100 participants reported were combined into an ‘other/unknown’ group; race was thus reported as other/unknown ($N = 441, 18.4\%$), White/Caucasian ($N = 1718, 71.7\%$), and Black/African American ($N = 216, 9.0\%$). There were no statistical differences between conditions at baseline across the combined race categories ($\chi^2(6) = 4.76, p = .58$).

Finally, a univariate ANOVA showed no differences across conditions for age ($F_{(3, 2347)} = 0.78, p = .51, \eta^2_p = .001$). See Table 2 for descriptive statistics by condition.

---

2 In total, 2597 participants were directed to the current study. Data were eliminated for those who did not have a valid CAMI-SR score and for participants who did not have a valid BIAT if they were in the control condition in which the BIAT was completed before the CAMI-SR. Valid CAMI-SR scores were limited to three standard deviations from the mean, or equal to or less than 3.7. Valid BIATs were those not disqualified per standard
Impact of Instructional Manipulation Check

The Instructional manipulation check (IMC; Oppenheimer, Meyvis, & Davidenko, 2009) is used as a proxy for determining whether a participant is liable to read instructions carefully throughout the experiment. Of the 1549 participants with valid BIAT and CAMI-SR scores, 44.22% failed the IMC. According to a univariate ANOVA, there were no differences across conditions in participant accuracy on the IMC ($F_{(3,2344)} = 1.57, p = .19, \eta^2_p = .002$). However, an independent means t-test showed that IMC accuracy was significantly associated with level of endorsed stigma on the CAMI-SR ($t_{(1634)} = -3.11, p = .002, d = -.08$). Those who accurately completed the IMC measure reported significantly less bias ($M = 2.15, SD = 0.50$) compared to those who incorrectly completed the IMC ($M = 2.23, SD = 0.51$). Because IMC accuracy was significantly correlated with our primary outcome measure of explicit stigma endorsement, we used IMC accuracy as a covariate for all analyses using CAMI-SR score.

Condition Effect on CAMI-SR

We hypothesized that the mental illness BIAT, which provided exposure to both implicit bias assessment and mental illness content, would result in the lowest CAMI-SR scores, followed by the gay/straight BIAT, and then the non-evaluative mental illness pseudo-BIAT. The no intervention control participants were expected to report the highest levels of bias. However, contrary to our expectations, results from a univariate analysis of covariance (ANCOVA), with intervention condition as the Between Subjects factor and IMC accuracy as a covariate showed that there was no effect of condition on explicit stigma endorsement on the CAMI-SR ($F_{(3,1631)} = 0.52, p = .67, \eta^2_p = .001$).

---

cleaning criteria as defined by Greenwald et al. (2003), which eliminated participants with error rates above 30% or who took less than 300 ms to respond to 10% of trials or more. These data reduction procedures reduced the data to 1549 participants.
Motivation to Control Prejudice as a Moderator

We further hypothesized that level of internal (but not external) motivation to control prejudice would significantly interact with condition to moderate CAMI-SR scores. Again controlling for IMC accuracy, results from the ANCOVA examining internal and external motivation scores as continuous moderators revealed main effects for both internal motivation ($F(1, 1562) = 267.67, p < .001, \eta^2_p = .15$) and external motivation ($F(1, 1562) = 153.33, p < .001, \eta^2_p = .09$). In line with prior research, greater internal motivation (using the IMS) was correlated with less expressed stigma on the CAMI-SR ($r(1576) = -.38, p < .001$) while greater external motivation (using the EMS) was correlated with more expressed stigma ($r(1576) = .27, p < .001$). As expected, no significant condition by EMS interaction was found ($F(3, 1562) = 1.30, p = .27, \eta^2_p = .002$). However, as hypothesized, there was a significant interaction between level of IMS and condition ($F(3, 1562) = 2.80, p = .04, \eta^2_p = .01$).

In order to examine the internal motivation by condition interaction, we conducted a median split on the continuous IMS variable, creating groups of participants high (scores greater than 6.06) and low in internal motivation (scores less than 6.06). Among participants low in internal motivation, pairwise comparisons across the four conditions revealed that participants in the gay/straight BIAT condition had significantly higher CAMI-SR scores than participants in the non-evaluative mental illness pseudo-BIAT condition ($p = .04$) and the mental illness BIAT condition (at the level of a non-significant trend; $p = .08$). No other condition comparisons reached significance (all $p > .10$). For those high in internal motivation, pairwise comparisons between groups revealed no significant differences in stigma endorsement across conditions (all $p > .10$). See Figure 1.
Taken together, these results indicate that, contrary to expectations, there was not a main effect of condition; instead, among individuals who are low in motivation to control prejudice, the gay/straight BIAT condition resulted in heightened explicit bias scores on the CAMI-SR.

**Discussion**

In this study, we sought to examine the impact of completing an implicit association measure on explicit bias endorsement, evaluating the contributions of implicit bias assessment and mental illness thought content, along with the moderating role of internal motivation to control prejudice. Contrary to expectations, exposure to implicit bias assessment toward people with mental illness versus people of another stigmatized minority group (i.e., gay people) or no implicit bias assessment at all had no significant effect on endorsement of stigma of people with mental illness. However, as predicted, internal motivation to control prejudice did significantly interact with intervention condition to influence explicit bias, although the direction of the effect was unexpected. Results indicated that condition differentially affected explicit stigma endorsement for those low in internal motivation, as opposed to those high in internal motivation as had been hypothesized. Participants low in internal motivation to control prejudice in the gay/straight BIAT condition reported higher levels of endorsed stigma as compared to levels reported after the non-evaluative mental illness pseudo-BIAT condition, and to a lesser extent, those in the valid mental illness BIAT condition.

**Lack of Intervention Effect**

It does not appear from these results that implicit bias assessment or priming mental illness content, either alone or in combination, positively affected participants’ explicit views of people with mental illness. Interestingly, examining the measure of positive and negative affect (PANAS) in a secondary analysis indicated that affect did not vary across condition for either
positive ($F_{(3, 1625)} = 1.52, p = .21, \eta^2_p = .003$) or negative ($F_{(3, 1656)} = 0.60, p = .62, \eta^2_p = .001$) affect. Thus, our conditions do not appear to have evoked the negative self-feelings that often accompany the should-would discrepancy or discrepant prejudice response detection from the Self-regulation of prejudice model. It is possible that the conditions did not provide a sufficient experience of (negative) implicit bias and how these biases can affect behavior to invoke personal reflection and cue for prejudice control in our participants.

With the failure to find condition effects on explicit bias endorsement, we did not replicate the findings from earlier research that showed taking a mental illness BIAT led to reduced explicit bias (Menatti et al., under review). Two factors that might have contributed to this discrepancy across studies include implicit bias feedback and participant demographics. First, while Menatti et al. showed that receiving personalized feedback about one’s implicit biases from the BIAT was not necessary to elicit the negative explicit evaluation reduction effect, the levels of endorsed stigma were lower for those who did receive feedback than those who did not (albeit non-significantly). In the current study, we chose to hold BIAT feedback until the end of the study in order to keep the conditions as similar as possible, because the two conditions without an initial valid BIAT would not have valid feedback. Including feedback might have reinforced the implicit bias education, providing a better chance to elicit cues to control prejudice. Therefore, while we still expected to see an effect, without the aid of the feedback we had less of an opportunity to see explicit stigma reduction.

Second, the failure to replicate the condition effect seen in the studies by Menatti et al. may also be partially explained by differences in sample characteristics between our respective study populations. While the samples were comparable on race, ethnicity, age, religious and political identification, levels of education, and mean CAMI-SR scores, the ratio of women to
men was higher in the Menatti et al. studies versus that seen in the current study (68.92% female versus 57.89% female, respectively). In the current study, a secondary analysis found gender differences in endorsed negative evaluations of mentally ill people ($F_{(2, 1625)} = 0.78$, $p < .001$, $\eta_p^2 = .01$), such that women reported less bias ($M = 2.15$, $SD = 0.50$) than men ($M = 2.24$, $SD = 0.52$). While neither study had significantly more females in any particular condition which would confound results, it is possible that having fewer females overall also decreased our opportunity to see decreased explicit stigma ratings across conditions.

**Internal Motivation as a Moderator of Stigma**

While internal motivation to respond without prejudice did moderate stigma endorsement by intervention condition according to our hypothesis, the effect did not occur in the expected direction. We originally hypothesized that for those participants high in internal motivation, implicit bias assessment would correlate with the lowest amounts of endorsed stigma because of their strong affective reactions to should-would discrepancy and greater attention to explicit stigma endorsement. This particular interaction did not occur; participants high in internal motivation were not differentially affected by intervention condition.

Instead, it was for those low in internal motivation for whom condition differentially influenced explicit stigma endorsement. The higher stigma endorsement by participants low in internal motivation who took the gay/straight BIAT as compared to the non-evaluative pseudo-BIAT (and valid mental illness BIAT at the level of a trend) ran contrary to our hypotheses. Due to the implicit bias assessment exposure provided in the gay/straight condition, we hypothesized this intervention would have the second lowest explicit stigma endorsement (we expected the mental illness BIAT to lead to the lowest explicit stigma, because it provides exposure to both implicit bias assessment and mental illness thought content).
These surprising findings might indicate that the gay/straight BIAT elicited a reactionary response for those low in internal motivation to control prejudice, leading to higher endorsed prejudice against those with mental illness. As levels of reported positive and negative affect did not vary across condition, we do not believe that this BIAT was more emotionally charged than the other condition BIATs. One potential factor that could have contributed to the higher stigma endorsement in the gay/straight condition is differential demand effects across the conditions. In the informed consent, participants were directly informed they would be asked to make evaluations of persons with mental illness; this instruction was included in hopes of standardizing demand effects for the explicit measure across conditions. This focus on mental illness evaluation may have been reinforced in the mental illness content conditions (both the non-evaluative mental illness pseudo-BIAT and valid mental illness BIAT conditions), whereas this reinforcement would presumably not have occurred in the gay/straight BIAT condition (which lacked further mental illness priming).

Although speculative, this may have led to inconsistent demand effects across conditions. For those low in internal motivation to control their prejudice, the demand effect may have provided incentive to endorse less prejudice in the mental illness content conditions, which was not apparent in the other conditions, because without this external cue this group shows little propensity to guard against endorsing stigma. While this explanation does not fully account for the pattern of results, it would be helpful in future research to assess demand effects to determine if they vary across conditions and predict explicit mental illness stigma scores in this context.

Furthermore, Crandall et al. (2002) have shown that people report having prejudice toward gay people in various roles is slightly more acceptable than toward those presumed to be mentally ill. On a scale from 0 (not OK to feel negatively toward these people) to 2 (OK to feel
negatively toward these people), participants’ ratings for “mentally unstable people” averaged 0.41, 0.52 for gay soldiers, and 0.61 for gay people who raise children. Priming participants with content about a stigmatized minority group for which it is more socially acceptable to have negative evaluations (e.g., gay people) could have thus primed those with low motivation to control their prejudice to report more prejudice toward a comparable stigmatized minority group (e.g., mentally ill people).

Another possible reason for a negative reaction to the gay/straight implicit association measure was that the “Don’t Ask, Don’t Tell” law for gay people serving in the military was repealed during the middle of our data collection (repealed by the Senate December 18, 2010 with our project active from December 3, 2010-January 11, 2011). Examination of the news cycle during the time of our data collection revealed that debates over the policy and gay people in the military were prevalent throughout the time our project was actively collecting data. It is possible that these debates in the media provided negative priming of gay people, and perhaps even elicited a threat response from those low in internal motivation to control prejudice. For those high in internal motivation to control prejudice, their strong personal conviction against being prejudiced and lower personal levels of prejudice likely buffered against the effects of negative media priming. For those low in internal motivation to control prejudice, however, the inflammatory rhetoric in the media about homosexuality could have weakened their already low personal conviction against endorsing prejudice, perhaps even providing a climate that encouraged endorsement of stigma against any group perceived as outside the norm.

Limitations and Future Research

The current findings need to be considered in light of a number of limitations. For instance, our participants tended to be highly educated (with over 90% reporting attaining “some
college” education or more), they are not representative of the general American public so results may not generalize accordingly. As the explicit stigma questionnaire (CAMI-SR) may be vulnerable to self-presentation effects, the questionnaire scores may not accurately represent the full degree of prejudice a participant may hold toward mentally ill people. The questionnaire is also not a measure of discriminatory behavior. It is possible that the BIAT interventions may still have activated the self-regulation of prejudice model and created cues for prejudice control that would have affected participants’ behavior in a situation with a mentally ill person. This change would then not be reflected on a questionnaire.

The current study calls for further research into understanding when and how the BIAT reduces endorsed stigma. In order to determine whether the “Don’t Ask, Don’t Tell” debates in the news cycle had an effect on our data, the project should be re-run during a time where discussion of gay rights is less contentious and prevalent in the media. Even though homosexuality and mental illness are both considered as residing under the same “social disability” social category, running the project with another stigmatized minority group that might be less prone to elicit reactionary responses as a comparison might show clearer indications of the generalizability of implicit bias assessment experience on stigma reduction.

To have a better chance of decreasing participants’ stigma endorsement, future studies should seek to modify the BIAT intervention to make participants’ should-would discrepancy from “feeling” their implicit biases more salient. Being a very brief intervention (with the entire study able to be completed in under 15 minutes), it would be constructive to determine whether the BIAT might need to be supplemented with other intervention material to be more effective in reducing stigma. While the BIAT showed no condition effect on its own in our study, it might
prove to be a beneficial addition other psychoeducational measures to reinforce concepts, or
could perhaps be repeated to increase intervention dosage.

In finding that the gay/straight BIAT condition might have in fact heightened explicit
bias toward those with mental illness for those low in internal motivation to control prejudice, we
are reminded that interventions can affect different subpopulations quite variably. Extensive
evaluation of interventions’ effect on diverse populations cannot be neglected. It is possible the
BIAT as an experiential intervention may still have application to reduce stigma toward
particular stigmatized minority groups. However, we cannot assume the intervention will work
effectively for all populations, especially for those who are the least motivated to control their
prejudice.
References


Table 1

*Example BIAT block order for the mental illness (good first) BIAT*

<table>
<thead>
<tr>
<th>Block</th>
<th>Number of Trials</th>
<th>Function</th>
<th>“IN” stimuli (I key)</th>
<th>“OUT” stimuli (E key)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>Practice</td>
<td>Mentally Ill People phrases</td>
<td>Physically Ill People phrases</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>Practice</td>
<td>Good words</td>
<td>Bad words</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>Test</td>
<td>Mentally Ill People phrases</td>
<td>Physically Ill People phrases</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+ Good words</td>
<td>+ Bad words</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>Test</td>
<td>Mentally Ill People phrases</td>
<td>Physically Ill People phrases</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+ Good words</td>
<td>+ Bad words</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>Practice</td>
<td>Bad words</td>
<td>Good words</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>Test</td>
<td>Mentally Ill People phrases</td>
<td>Physically Ill People phrases</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+ Bad words</td>
<td>+ Good words</td>
</tr>
<tr>
<td>7</td>
<td>40</td>
<td>Test</td>
<td>Mentally Ill People phrases</td>
<td>Physically Ill People phrases</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+ Bad words</td>
<td>+ Good words</td>
</tr>
</tbody>
</table>
Table 2

Demographic, stigma endorsement score, and internal and external motivation score information by intervention condition

<table>
<thead>
<tr>
<th></th>
<th>Mental Illness BIAT</th>
<th>Gay/Straight BIAT</th>
<th>Non-Evaluative Pseudo-BIAT</th>
<th>No Intervention Pseudo-BIAT</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>550</td>
<td>391</td>
<td>275</td>
<td>333</td>
<td>1549</td>
</tr>
<tr>
<td>% Female</td>
<td>56.73</td>
<td>58.31</td>
<td>55.63</td>
<td>60.66</td>
<td>57.78</td>
</tr>
<tr>
<td>Age Mean (SD)</td>
<td>28.35 (11.70)</td>
<td>29.76 (11.68)</td>
<td>29.78 (12.00)</td>
<td>28.88 (11.86)</td>
<td>29.07 (11.79)</td>
</tr>
<tr>
<td>CAMI-SR Mean (SD)</td>
<td>2.17 (0.51)</td>
<td>2.21 (0.52)</td>
<td>2.17 (0.51)</td>
<td>2.18 (0.50)</td>
<td>2.18 (0.51)</td>
</tr>
<tr>
<td>EMS Mean (SD)</td>
<td>4.09 (1.81)</td>
<td>3.87 (1.79)</td>
<td>4.16 (1.83)</td>
<td>4.07 (1.83)</td>
<td>4.04 (1.82)</td>
</tr>
<tr>
<td>IMS Mean (SD)</td>
<td>6.07 (1.16)</td>
<td>6.00 (1.12)</td>
<td>6.07 (1.19)</td>
<td>6.11 (1.10)</td>
<td>6.06 (1.41)</td>
</tr>
</tbody>
</table>

Note. BIAT refers to Brief Implicit Association Test; CAMI-SR refers to Community Attitudes toward Mentally Ill-Social Restrictiveness Scale; EMS refers to External Motivation Score; IMS refers to Internal Motivation Score.
Figure Caption

*Figure 1.* Average explicit stigma endorsement toward mentally ill people by participants high and low in internal motivation to control prejudice across intervention conditions.
Implicit Association Experience and Stigma

Mean CAMI-SR Score

Intervention Condition

Mental Illness BIAT | Gay/Straight BIAT | Non-Evaluative Mental Illness pseudo-BIAT | No Intervention Control

High IMS Low IMS | High IMS Low IMS | High IMS Low IMS | High IMS Low IMS