Reducing Stigma Toward Individuals with Mental Illnesses: A Brief, Online Intervention

Abstract

**Purpose:** Our goal was to determine the effectiveness of a brief online intervention designed to reduce stigma-relevant attitudes toward mentally ill individuals. We examined whether the experience of completing a Brief Implicit Association Test (BIAT) measuring attitudes toward people with mental illnesses alters explicit stigma in two studies.

**Methods:** In Study 1, participants completed the BIAT and received feedback about their individual level of implicit bias. Study 2 added a condition in which the BIAT was administered without feedback on performance.

**Results:** Study 1 showed that the intervention was effective in reducing stigma-relevant attitudes toward mentally ill individuals. Study 2 showed that receiving feedback about one's own implicit bias was not necessary for the intervention to reduce stigma.

**Conclusion:** These studies show that the BIAT intervention may be effective at reducing explicit stigmatizing attitudes toward individuals with mental illnesses, at least in the short-term.

**Implications:** This brief, accessible, and cost-effective stigma reduction strategy may be of use to community organizations whose mission is to address issues surrounding stigmatization in real-world settings.

**Keywords:** bias, Brief IAT, implicit associations, Implicit Association Test, stigma attitudes

Introduction

One important aspect of stigma is the negative attitudes toward a particular group of people that pervasively taint interactions with members of that group (Goffman, 1963). Stigma tied to mental illnesses can lead to losses in earning, housing, and health (Link & Phelan, 2001). The current studies address whether experiencing an assessment of one's own implicit bias and learning that bias can exist outside of awareness or control can change explicit stigma toward people with mental illness.

Much of the early research on stigma of mental illnesses focused almost exclusively on explicit attitudes (e.g., Farina, Holland & Ring, 1966; Swanson & Spitzer, 1970). Recently, however, researchers have examined the value of also including implicit assessments of stigma, to reduce the need for conscious introspection and influence of social desirability in measurement (Greenwald & Banaji, 1995; Stier & Hinshaw, 2007). Teachman, Wilson and Komarovskaya (2006) showed that the general public, and even those with serious mental illness, were both implicitly and explicitly biased against other mentally ill people, compared to physically ill people (see also Rüscher, Todd, Bodenhausen & Corrigan, 2010; Rüscher, Corrigan, Todd & Bodenhausen, 2011). Peris, Teachman and Nosek (2008) also demonstrated the value of implicit stigma assessments, finding that implicit attitudes toward mentally ill people predict different clinical decisions rather than explicit attitudes. Participants read clinical vignettes and then rated expected prognosis and gave a diagnosis if they felt one was warranted. Explicit (but not implicit) attitudes predicted more severe prognoses, while implicit (but not explicit) attitudes predicted relatively more over-diagnosis.

In the current studies, we extend the findings about implicit bias toward people with mental illnesses and test whether learning about implicit bias and experiencing implicit measurement can modify explicit stigma about mentally ill people. Researchers have had some success in reducing stigma by revealing discrepancies between people’s ideal and actual attitudes (e.g., Devine, Monteith, Zuwerink & Elliot, 1991; Plant & Devine, 1998, 2009). For example, Devine et al. (1991) found that participants who reported a substantial difference in the way they ‘should’ act around a
stigmatized group member and the way they thought they ‘would’ act around a stigmatized group member were more likely to develop stronger feelings of discomfort, guilt, and self-criticism, relative to persons with less of a ‘should–would’ discrepancy.

Monteith and colleagues extended this work, and found that participants who experienced self-criticism and guilt when confronted with ‘should–would’ discrepancies were likely to engage in self-regulating behaviors and develop ‘cues for control’ (Monteith, Ashburn-Nardo, Voils & Czopp, 2002; Monteith & Mark, 2005; Monteith, Mark & Ashburn-Nardo, 2010; see also Legault, Green-Demers & Eadie, 2009). Cues for control are environmental stimuli which activate self-regulating behaviors that help guard against unwanted biased actions. Cues for control can be anything from an encounter with a person from a stigmatized group, to seeing pictures or words associated with that group. We expect that experiencing implicit bias assessment can also activate cues for control; as Morris and Ashburn-Nardo (2010) note, recognizing one’s biases is a critical step in the successful control of prejudiced responding (p. 64). Along these lines, Monteith and colleagues (2002, experiment 4) found that having participants complete an Implicit Association Test (IAT) tied to race attitudes, a common measure of implicit bias, resulted in greater liking of stereotypically Black names. This effect was greater for individuals who were more upset about their implicit race bias than for those who were less upset. This promising finding suggests that experiencing implicit bias measurement can serve as a cue for control to reduce stereotypical responding.

The current research extends the work of these initial promising studies to develop a brief, web-based implicit bias awareness intervention to target explicit stigma toward persons with mental illnesses. These studies extend prior research in this area in a number of ways. To our knowledge, Monteith et al. (2002) report the only published study that involved examining the impact of taking an IAT on subsequent explicit bias. In their work, they focused on race attitudes and used only White undergraduate psychology students who indicated low prejudice at the start of the study. Also, they did not have a control group (who did not complete the race IAT in advance of the explicit stereotyping task). In the current study, we examine a new stigma domain (mentally ill people), introduce a control condition to more clearly determine the effect of the implicit bias assessment, and use two large, relatively more diverse general public samples (rather than only low prejudiced White undergraduates). Further, Monteith et al. (2002) told all participants their IAT score and discussed the significance of the score, leaving it unclear whether it was implicit bias assessment that elicited cues for control, or discussion about one’s implicit bias score (a question we test in Study 2).

Notably, our hypotheses also differ somewhat from those guiding the original cues for control model. While Monteith and colleagues expected that cues for control would only exert an effect on explicit bias when the person recognized and felt guilty about their own unacceptably negative implicit bias, we expect that the experience of implicit bias measurement can raise awareness about the importance of controlling prejudice even if the person’s own implicit score is not very negative. However, we agree that the potential moderating effect of implicit bias level is an interesting question, and so we tested whether the effect of the intervention varied across levels of implicit bias toward people with mental illness. On the one hand, people who learn that they have a high level of implicit bias may be especially alarmed at their results and may commit extra effort to self-regulation, resulting in weaker explicit stigma. On the other hand, individuals high in implicit bias may be more resistant to change because they may possess biases that are more entrenched in their self-concept.

Study 1

In Study 1, participants in the experimental condition completed a Brief Implicit Association Test (BIAT; Sriram & Greenwald, 2009), assessing implicit attitudes toward mentally ill people. After completing the BIAT, participants received feedback about their performance on the task along with some general information about the nature of implicit biases. Finally, they completed an explicit measure of stigma toward mentally ill people. In the control condition, participants completed the explicit measure of stigma before completing the implicit task. Both studies reported in this article received ethical approval from our University’s Institutional Review Board for Social and Behavioral Sciences.

Methods

Participants

Data were gathered through the publicly-available Project Implicit website (http://implicit.harvard.edu/). Eligibility was limited to U.S. citizens aged 18 or older. Project Implicit visitors can participate in studies and learn about implicit social cognition. Nine-hundred and eighty-five participants consented to participate. Of those, 62% persisted through the whole session and earned valid scores on both the BIAT and the CAMI-SR (59% of those in the BIAT-first condition, 66% in the CAMI-first condition). These ‘completers’ were similar to ‘incompleters’ in terms of gender and ethnicity (Hispanic origin) ratios, but differed somewhat in terms of race (\(\eta_p^2=22.6, p<0.01\))
Reducing Stigma

The BIAT (Sriram & Greenwald, 2009) is similar to the Implicit Association Test (IAT; Greenwald, McGhee & Schwartz, 1998), but uses a briefer format and, unlike the IAT, does not use an explicitly labeled comparison category to contrast with the mentally ill people target category. The task requires rapid categorization of exemplars into designated categories. BIAT instructions direct participants to be on the lookout for exemplars of two ‘focal’ categories at a time (e.g., Coke and good; ‘Press the “I” key when you see representative items.’) and specify that ‘anything else’ be identified with the ‘E’ key. If a person correctly categorizes exemplars more quickly when Coke and good are the focal categories than when Coke and bad are focal, it indicates stronger associations between the former than the latter concepts (i.e., a positive implicit evaluation of Coke).

In the BIAT, Mentally Ill People was always one of the two focal categories, and Good or Bad, in counterbalanced alternating phases, was the other. Exemplars of physically ill people were used as non-focal items. Thus, when participants were instructed to focus on exemplars of Mentally Ill People and Good words, the anything else items included the designated Bad words as well as exemplars of physically ill people; when the focus was on Mentally Ill People and Bad words, Good words along with physically ill exemplars belonged in the anything else domain. Mentally Ill People exemplars were person with anxiety, person with depression, and person with schizophrenia; Physically Ill People exemplars were person with appendicitis, person with cancer, and person with pneumonia; Good exemplars were ‘excellent’, ‘great’, and ‘wonderful’; Bad exemplars were ‘horrible’, ‘nasty’, and ‘terrible’. We chose to use the term, person with…, to focus the participant on the individual as opposed to the more objective question of whether mental illnesses were negative from a health standpoint.

As noted, the BIAT uses only a single, labeled target category (i.e., Mentally Ill People), but we maintained the standard 7-phase structure of an IAT (Nosek, Greenwald & Banaji, 2007a), albeit with fewer trials in each phase. An outline of the procedure for this BIAT can be seen in Table 1. BIAT D scores were calculated according to the recommendations (Greenwald, Nosek & Banaji, 2003; Sriram & Greenwald, 2009). The D score is an effect size measure for each participant that indexes the difference in average response speed across the two critical comparison conditions (Mentally Ill paired with Good vs. Mentally Ill paired with Bad) while accounting for the participant’s overall response speed. Specifically, BIAT D scores were computed by taking the difference in mean response latency between the two 24-trial test phases (4 and 7) and dividing it by the participant’s standard deviation of all latencies in these phases. The possible range of D scores is −2 to 2, with scores of zero.
indicating no difference in average response latency between the two critical comparison conditions. To align interpretation with scoring of the CAMI-SR (i.e., higher scores indicating greater endorsement of discrimination toward mentally ill people), we scaled the BIAT scores so that positive values indicate relatively more negativity toward people with mental illness.

Immediately following the BIAT, participants were presented with personalized feedback about their BIAT performance (see example in Figure 1) and general information about implicit bias. The latter included a basic conceptual explanation of how differential speed of responding on the two conditions of the task is used to infer an implicit attitude, and examples of demonstrated relations between implicit attitudes and behaviors. The personalized feedback was one of seven characterizations of implicit association strength derived from the participant’s BIAT score (i.e., slight, moderate, or strong association of persons with mental illness with good; slight, moderate, or strong association of persons with mental illness with bad; or no difference in your implicit association of persons with mental illness with good vs. bad). The BIAT score ranges generating these classifications were established to approximate Cohen’s (1992) small, medium, and large effect size guidelines (Nosek et al., 2007b): scores with absolute values of 0.15 to 0.34 generated the ‘slight’ implicit association message in the given direction, 0.35 to 0.64 scores yielded the ‘moderate’ association message, and scores of 0.65 or higher produced the ‘strong’ association message. A ‘no difference’ message was received by participants with scores in the 0.0 to 0.14 range.

Results

Descriptive statistics for the two primary measures, the CAMI-SR and the BIAT, as well as for the BIAT feedback score, are listed by condition in Table 2. Only 2.1% of participants who answered any CAMI-SR items failed to answer all 10 items. These participants were excluded from these analyses. Following Sriram and Greenwald (2009), split-half reliabilities for the BIAT were based on D scores of subsets formed of alternating couplets of trials from the 24 test trials in each condition. Standardized Cronbach’s α internal consistencies were 0.78 and 0.73, respectively, for the CAMI-SR and BIAT. These internal consistencies compare favorably with those observed in laboratory administrations of other BIATs (Sriram & Greenwald, 2009) and of the CAMI-SR (Taylor & Dear, 1981). In both conditions (CAMI-SR first and BIAT first), CAMI-SR and BIAT means are in the direction of relative positivity toward mentally ill people. That is, CAMI-SR means are significantly <3.0, the neutral option, and BIAT scores are significantly or marginally (p=0.051 in the CAMI-SR first condition) <0.0, indicating slight relative positivity (vs. negativity) toward mentally ill persons.

Our primary question was whether explicit bias, indexed by CAMI-SR scores, would shift as a function of first completing the BIAT. As hypothesized, the mean CAMI-SR score for participants who first took the BIAT (M=2.09, SD=0.53) was lower than for those completing the CAMI-SR first (M=2.23, SD=0.53; see Figure 2). Regression analysis identified a significant standardized effect of condition, coded 1 if BIAT was first and 0 if CAMI-SR was first, as β=–0.13 (t₁=-3.33, p=0.001). Regression diagnostic analyses following Cohen, Cohen, West and Aiken (2003) confirmed that this result was not a function of unduly influential observations. This effect was not mediated or moderated by BIAT D score or the 7-level BIAT-feedback variable shown to participants, nor by the counterbalancing order within the BIAT (i.e., regardless of whether mentally ill people stimuli were first paired with good or with bad stimuli), nor by participant age, gender, race, or education level.

Table 1: Structure of the Brief Implicit Association Test (BIAT)

<table>
<thead>
<tr>
<th>Phase</th>
<th>Number of Trials</th>
<th>Function</th>
<th>Category Labels</th>
<th>‘Anything else’ Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>Practice targets</td>
<td>Mentally Ill People</td>
<td>Physically ill person exemplars</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>Practice valence</td>
<td>Good*</td>
<td>Bad words</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>Practice combination 1</td>
<td>Mentally Ill People and Good</td>
<td>Physically ill person exemplars and bad words</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>Test combination 1</td>
<td>Mentally Ill People and Good</td>
<td>Physically ill person exemplars and bad words</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>Practice valence reversed</td>
<td>Bad</td>
<td>Mentally Ill People and Bad</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>Practice combination 2</td>
<td>Mentally Ill People and Bad</td>
<td>Good words</td>
</tr>
<tr>
<td>7</td>
<td>24</td>
<td>Test combination 2</td>
<td>Mentally Ill People and Bad</td>
<td>Physically ill person exemplars and good words</td>
</tr>
</tbody>
</table>

Note: ‘Order of Good and Bad categories was randomized between participants.
Implicit association test feedback

The sorting task you just completed is a "Brief Implicit Association Test" or BIAT. It provides an estimate of your "implicit attitude" toward mentally ill people. By implicit attitude, we mean one that may exist outside of your conscious awareness or control. The idea behind the BIAT is that the more strongly associated two concepts are in memory, the more quickly you will be able to classify words into paired categories reflecting those concepts. For example, people are usually quicker to categorize items (such as rose, tulip, etc.) into a "flower" category when it is paired with the category "good" than when it is paired with the category "bad." This indicates that flowers are more implicitly associated with the concept "good" than with the concept "bad." The BIAT you completed was designed to assess the relative strength of your associations of people with mental illnesses with "good" versus "bad." Your result is reported below.

Your result

Your data suggest a strong implicit association of persons with mental illness with good (versus bad). **This is because you were much faster at the classification task when the concept 'Mentally Ill People' was paired with the concept 'Good' than when 'Mentally Ill People' was paired with the concept 'Bad', which we interpret as an indication of your implicit evaluation of persons with mental illness.

It is important to realize that people's implicit attitudes do not always match their explicit attitudes, and that this test cannot be used to diagnose any mental disorder.

At the same time, although we are sometimes unaware of our implicit biases, past research has shown that they can automatically influence our thoughts and actions in profound and fundamental ways. For instance, implicit associations have been shown to predict everything from political preferences, to how one interacts with persons from various minority or disadvantaged groups (e.g., desire to approach closely), to responses to therapy.

Without measures like the BIAT, we may not perceive the existence of implicit associations in our own minds. Now that you are aware of how implicit associations can influence your thoughts and feelings, how might your associations be affecting your own views and behaviors toward persons with mental illnesses?

Please bear in mind that these BIAT results are not perfectly reliable, and can change over time. On the other hand, it is very possible to possess an automatic evaluation that you would rather not have. If that is the case, one possible solution is to seek experiences that could undo or reverse the situations that may have created the unwanted evaluation by doing one or all of the following:

- Find opportunities to interact with persons who have a mental illness.
- Remain alert to the existence of the undesired evaluation, recognizing that it may intrude in an unwanted fashion into your judgments and actions.
- Expose yourself to material that opposes the idea that mentally ill people are bad.

Figure 1. Example of a BIAT feedback page.

### Table 2: Descriptive Statistics and Correlations by Study and Condition

<table>
<thead>
<tr>
<th>Measure</th>
<th>Study 1 Conditions</th>
<th>Study 2 Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CAMI-1st (n=318)</td>
<td>CAMI-1st (n=277)</td>
</tr>
<tr>
<td></td>
<td>BIAT-1st Immediate Feedback (n=292)</td>
<td>BIAT-1st Immediate Feedback (n=229)</td>
</tr>
<tr>
<td></td>
<td>BIAT-1st Delayed Feedback (n=245)</td>
<td></td>
</tr>
<tr>
<td>CAMI-SR</td>
<td>2.23* (0.53)</td>
<td>2.16* (0.52)</td>
</tr>
<tr>
<td></td>
<td>–0.05 (0.48)</td>
<td>–0.09* (0.50)</td>
</tr>
<tr>
<td>Feedback</td>
<td>–0.19 (1.76)</td>
<td>–0.28* (1.80)</td>
</tr>
<tr>
<td></td>
<td>2.09* (0.53)</td>
<td>2.03* (0.54)</td>
</tr>
<tr>
<td></td>
<td>–0.06* (0.47)</td>
<td>–0.05 (0.47)</td>
</tr>
<tr>
<td></td>
<td>–0.21* (1.74)</td>
<td>–0.18 (1.69)</td>
</tr>
<tr>
<td>BIAT D</td>
<td>–0.19* (1.76)</td>
<td>–0.35* (1.78)</td>
</tr>
<tr>
<td>Feedback</td>
<td>–0.01</td>
<td>–0.02</td>
</tr>
<tr>
<td></td>
<td>0.98*</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Correlations</td>
<td>Study 1</td>
<td>Study 2</td>
</tr>
<tr>
<td>(1) CAMI-SR</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>(2) Condition*</td>
<td>–0.13*</td>
<td>–0.02</td>
</tr>
<tr>
<td>(3) BIAT D</td>
<td>0.21*</td>
<td>–0.11*</td>
</tr>
<tr>
<td>(4) Feedback</td>
<td>0.19*</td>
<td>0.12*</td>
</tr>
<tr>
<td></td>
<td>–0.01</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>0.98*</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.97*</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Note: *Coefficient is significantly non-zero (or non-neutral; i.e., 3, in the case of CAMI-SR) at p<0.05. CAMI-SR is the score on the Community Attitudes Toward the Mentally ill – Social Restrictiveness Scale. BIAT D is the score on the BIAT, with positive scores indexing relative negativity toward persons with mental illness. Feedback is a 7-level variable (scored –3 to 3) indexing the feedback participants received about their BIAT performance. Condition* is a dummy coding indexes the effect of taking the BIAT first: 0 if CAMI-SR came first, 1 if BIAT came first. Code-1 is an orthogonal contrast code indexing the effect of either of the Study 2 BIAT-1st conditions (i.e., regardless of feedback timing), compared with the CAMI-SR-1st condition: –0.5 if CAMI-SR came first, 1 if BIAT came first. Code-2 is an orthogonal contrast code indexing the effect of the BIAT first with immediate feedback condition vs. that of the BIAT first without immediate feedback condition: 0 if CAMI-SR came first, 0.5 if BIAT came first and feedback was immediate, –0.5 if BIAT came first and feedback was delayed.
Finally, BIAT performance was not influenced by whether CAMI-SR came before or after, or by counter-balanced order of the good and bad BIAT evaluative categories. Two dummy-coded variables were created to test for order effects: Condition=0 for CAMI-1st, and 1 for BIAT-1st; BIAT good 1st=0 when ‘Bad’ was the first target, and 1 when ‘Good’ was first. With these two variables and their interaction as predictors in a regression model for BIAT score, neither the model nor any of the regression coefficients was statistically significant [R²=0.001, F(3, 665)=0.17, p=0.92].

Study 2

While the hypotheses of Study 1 were confirmed, it remained to be determined if the results of the study were attributable to the BIAT feedback, completing the BIAT itself (without feedback) or if both were necessary to reduce explicit stigma-relevant attitudes. Study 2 was designed to answer these questions.

Methods

Participants

U.S. citizens aged 18 and older volunteering through the Project Implicit Research portal between June 3 and July 12, 2010 were assigned to Study 2 in the same manner as in Study 1. Of the 1161 that consented to participate, 65% (n=751) validly completed both the CAMI-SR and the BIAT. The demographic characteristics were similar to those of Study 1: an average age of 29 (median 25), 69% female, and comparable ethnic and educational proportions. Data were excluded for one participant because his score on the CAMI-SR was an extreme outlier (>5 SDs above the mean and >2 SDs above the next highest score) and exerted undue influence on regression analyses. Regression diagnostics were again performed according to Cohen et al. (2003).

Materials and Procedure

The materials and procedures were identical to those of Study 1 except that a third condition was added in which BIAT score feedback and general information about implicit bias did not immediately follow performance of the BIAT, but was delayed until after participants completed the CAMI-SR. Thus, in Study 2, we tested whether the reduction in explicit stigma as a function of our BIAT intervention depended on whether direct feedback about implicit biases and one’s BIAT score was given.

Results

Descriptive statistics are listed by condition in Table 2. Standardized Cronbach’s α internal consistencies for the CAMI-SR and BIAT, 0.79 and 0.74, respectively,
were virtually identical to those observed in Study 1. It is worth noting that only 1.6% of participants in Study 2 who completed any CAMI-SR item failed to complete all 10 CAMI-SR items. These few participants were excluded from these analyses. Our primary results, graphed in the Study 2 section of Figure 2, are that both BIAT-first conditions (i.e., with and without immediate feedback) resulted in lower CAMI-SR scores than did the control condition (CAMI-SR-first), and that the effects of the two BIAT-first conditions did not differ from one another. We tested these condition effects by using two orthogonal contrast codes as predictors in a regression model for CAMI-SR score: Code-1 contrasted the effects of the two BIAT-first conditions (i.e., the effect of immediate vs. delayed BIAT feedback), while code-2 contrasted the effect of the BIAT-first conditions, combined, with that of the control condition. Coding details are shown in Table 2. With these two codes as simultaneous predictors of CAMI-SR, the effect of code-1 (immediate vs. delayed BIAT feedback), was non-significant (p=0.51), but the effect of code-2 (BIAT-first vs. control) was significant, p<0.01. We unpacked the code-2 effect by conducting two separate regression analyses, one directly replicating the analysis of Study 1 by comparing effects of the control with the BIAT-first-with-immediate-feedback condition, and the other comparing the control with the BIAT-first-without-immediate-feedback condition. In both analyses, the BIAT-first condition resulted in significantly lower CAMI-SR scores than the control condition: β=-0.13, t=-2.84, p=0.005 when feedback was immediate (directly replicating Study 1) and β=-0.10, t=-2.21, p=0.028 without feedback. These results suggest that the bias reduction effect observed in Study 1 is not dependent on being explicitly informed about implicit bias and one’s BIAT performance.

As in Study 1, the effect on CAMI-SR score of first taking a BIAT, with or without immediate feedback, was not mediated or moderated by counterbalancing order within the BIAT, by participant age, gender, race, or education level. However, unlike in Study 1, in this sample there was a small, but statistically significant, moderating effect of BIAT performance (whether indexed by the actual BIAT score, which is unknown to participants, or by the highly correlated BIAT feedback variable). This interaction effect was identified by adding three predictors to the baseline contrast code model described above: (1) BIAT score and (2 and 3) the product of BIAT score and each of the two contrast codes. The code×BIAT interaction was significant (p=0.026), indicating that the effect of the BIAT-first conditions vs. the control condition varied with level of BIAT score. Specifically, among those in the BIAT-first conditions, higher BIAT scores (greater relative negativity toward mentally ill persons) were associated with greater reductions in CAMI-SR scores.

Note, the same interaction pattern between condition and BIAT score held in each of the two separate regression models. That is, in the model contrasting the BIAT-first-immediate-feedback condition vs. the control condition, the standardized estimate of the interaction effect was β=-0.09, t=-2.01, p=0.045, and for the model contrasting BIAT-first-delayed-feedback vs. control, the estimate was β=-0.08, t=-1.80, p=0.072.

**Discussion**

The brief implicit bias awareness intervention conducted in these studies reduced explicitly reported stigmatizing attitudes toward mentally ill people. In Study 1 participants in the experimental condition completed a BIAT assessing their implicit attitudes toward mentally ill people and received feedback about their performance along with general information about implicit bias. Afterwards, participants completed a short explicit measure (CAMI-SR) that assessed explicit stigma toward mentally ill people. Control participants completed the CAMI-SR prior to taking the BIAT. In Study 2 an additional group was added that completed the BIAT first, but then went on to complete the CAMI-SR before receiving BIAT feedback. The timing of BIAT feedback, immediate or delayed, made no difference in explicit bias among the participants who took the BIAT first, suggesting that the act of completing the BIAT led to the reduction in explicit stigma. Also, in Study 2 (but not Study 1), the impact of the bias awareness manipulation was moderated by participants’ level of implicit bias, such that participants who were more implicitly biased were more strongly influenced by completing the BIAT. Together, these findings point to the potential value of personally experiencing implicit bias measurement for reducing explicit stigma. The findings also raise many questions that warrant investigation about potential moderators and mediators of the intervention’s effects.

Devine and colleagues’ research on the should-would discrepancy might provide a good theoretical basis for understanding the current intervention’s success in reducing explicit bias (e.g., Devine et al., 1991; see also later work by Monteith, Voils & Ashburn-Nardo, 2001; Monteith et al., 2002). According to this research, when an individual recognizes a difference between the way she perceives she should act toward a particular group and the way she anticipates actually acting toward the group, she experiences global feelings of discomfort, which would presumably lead her to engage in efforts to decrease this discrepancy. As participants complete the
BIAT, they may have a sense of how they are doing even if they have not received explicit feedback. Alternatively, the BIAT may provide a palpable experience that implicit attitudes (regardless of one's specific current performance) can reside outside conscious control and influence well-intended egalitarian views. This experiential awareness that arises from completing the BIAT may be sufficient to initiate self-regulating effects on explicit responses. Completion of a BIAT may thus engender concern about a possible should–would discrepancy that leads participants to monitor and try to reduce their explicit stigma. Given that the BIAT is usually used solely as a measurement tool, rather than as an intervention, researchers will want to consider how completion of the measure may alter responding on other tasks.

Also of interest, we observed some evidence (in Study 2) that the experience of completing an implicit measure of bias toward mentally ill people may be most effective in reducing explicit stigma for those with higher levels of implicit bias. That is, the effects of the intervention were somewhat larger for those with the strongest implicit bias in the first place. This interaction is unlikely to be an artifact of regression to the mean (i.e., to the extent that high BIAT scores have resulted by chance, CAMI scores would be expected to be lower, closer to the mean, because the random effects would not operate in the same way), because such an effect would be expected to operate in the same way regardless of condition. However, this interaction appeared only in Study 2, suggesting caution until it is replicated. Perhaps the most obvious question for future research has to do with the durability of these stigma-reduction effects. Explicit attitudes were assessed within several minutes of completing the BIAT and/or the feedback components. If the intervention is not potent enough to persist, then its practical application will be limited. Of course, such a simple 5-min intervention might not have lasting impact, but it may provide a gateway for developing stronger interventions that will be more durable. For example, an intervention that includes multiple sessions may produce more durable effects.

Other considerations for future research include replicating the intervention effects with other topics and samples. While the present sample was fairly diverse, and participants were randomly assigned to the study out of many possible studies, the sample was not representative of any particular population, and many participants arrived at the site with some expectation that they would complete measures about their implicit associations. Additionally, given that participants generally showed positive implicit and explicit attitudes toward individuals with mental illnesses compared to those with physical illnesses (regardless of experimental condition), future research should investigate the effectiveness of similar interventions among those with more negative implicit and explicit attitudes about individuals with mental illnesses. It is important, however, to emphasize that the implicit evaluation of mental illnesses is necessarily made relative to physical illnesses given the procedural characteristics of the IAT. That our sample was relatively more negative toward physically than mentally ill persons does not obviate the effect of the intervention on explicit stigma toward mental illness. The simple experience of completing a measure of implicit bias elicited less explicit negativity toward people with mental illness, suggesting that being faced with an implicit evaluation experience (regardless of the outcome of that implicit evaluation) generates more positive explicit evaluations. The extent that this is a self-regulatory strategy or operates via another mechanism requires further examination. Similarly, evaluating how the effect varies across mental illness diagnoses will also be useful given heterogeneity in the degree and nature of stigma directed toward different types of mental illnesses (see Crisp, Gelder, Rix, Meltzer & Rowlands, 2000).

It will also be important to investigate the extent to which the intervention effects reflect self-presentational vs. internalized changes in stigma. We think it is unlikely that self-presentation or demand effects account for these findings given that: (a) the group in study 2 that received direct verbal feedback about their level of implicit bias toward mentally ill people prior to taking the CAMI-SR, which would presumably create the strongest demand effects, did not perform differently vs. internalized changes in stigma. We think it is unlikely that self-presentation or demand effects account for these findings given that: (a) the group in study 2 that received direct verbal feedback about their level of implicit bias toward mentally ill people prior to taking the CAMI-SR, which would presumably create the strongest demand effects, did not perform differently than the group that completed the BIAT without receiving explicit feedback; (b) it is clear that the CAMI-SR is assessing endorsement of discrimination in which ever order it is presented, so it is unlikely that group differences on the measure could be fully explained by demand effects because any impact of knowing that bias is being assessed is present for all conditions; (c) all groups were alerted during consent that the study examined attitudes and issues related to mental health, so this topic was primed for all participants, yet condition differences still emerged; and, (d) because the study took place online with no experimenter present and anonymous responding, the situational features of the study likely minimized self-presentational and demand concerns (Rieps, 2002; Mangan & Rieps, 2007).

These results have potentially interesting implications for the future of basic and applied research on the stigmatization of different groups. While determining that a 5-min computerized intervention reduces stigma-related attitudes in a research sample is certainly of interest to academics, the breadth of the sample, as well as the brevity, accessibility and cost-effectiveness of the intervention, may potentially also be of interest to community organizations whose mission is to address issues of stigmatization in real-world settings. Similarly, individuals who frequently come into
contact with individuals with mental illnesses, such as mental health professionals, may not recognize the extent to which mental illnesses can be stigmatized, even among themselves. The experience of completing an implicit bias measure may thus be helpful to remind these individuals about how easy it is to experience and express bias.

Questions remain about how experiential education about implicit bias can reduce explicit stigma, but assuming these results extend across other topics and show some durability, these results provide a promising starting point for the development of stigma-reduction tools that are brief and amenable to wide dissemination.

References

