UNDERSTANDING AND REDUCING IMPLICIT MENTAL ILLNESS STIGMA: A CONTEMPORARY PREJUDICE PERSPECTIVE

REBECCA E. YOUNG

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN PSYCHOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

FEBRUARY 2018

Abstract

The stigma of mental illness is a serious social issue that exists across nations and cultures. Over the years, numerous anti-stigma campaigns have been developed to reduce the stigma associated with mental illness. However, stigmatizing attitudes still persist, which suggests that stigma not only exists in explicit and direct forms, but may also be expressed subtly and automatically causing it to remain unnoticed and thus unchanged. The purpose of this dissertation was to provide a deeper understanding of implicit stigma in order to determine an effective intervention to reduce it. Study 1 examined individuals' implicit attitudes using the Go/No-Go Association Task (Nosek & Banaji, 2001) and found that university student participants had more negative implicit attitudes toward mental illness than positive and tended to automatically associate mental illness with dangerous and helpless attributes. Consistent with implicit racial prejudice research, Study 1 also identified a contemporary form of stigma described as aversive stigmatization, which refers to when individuals explicitly report non-stigmatizing attitudes, but harbour implicit negative attitudes toward mental illness. This has implications for discrimination as aversive stigmatizers were found to be less avoidant and more willing to help individuals with mental illness compared to high stigmatizers, but more avoidant and less helping compared to low stigmatizers, suggesting that aversive stigmatizers express stigma more subtly. Study 2 developed and tested an intervention to reduce implicit stigmatizing attitudes toward mental illness, which had not yet been examined. Results demonstrated that the intervention, which contained education, bias awareness, and contact components, was effective overall in reducing negative implicit attitudes toward mental illness. Furthermore, the intervention was most effective for aversive stigmatizers (compared to low, high, and intentional stigmatizers) in improving prosocial behaviour toward individuals with mental illness. These

findings highlight the complex nature of stigma and illustrate the importance of continuing to examine implicit, contemporary forms of stigma that are subtle, yet harmful to individuals with mental illness. The findings are encouraging in that they demonstrate the possibility of reducing implicit stigmatizing attitudes and point to the continued need for specialized interventions that target all aspects of stigma in order to effectively reduce it.

Dedication

To my husband, for his patience and endless support throughout this process, and to my parents who have always supported and encouraged me to pursue my goals.

Acknowledgements

I would like to express my gratitude to Drs. Joel Goldberg, Ward Struthers, and Doug McCann, each of whom have been an invaluable influence and source of support through every stage of my dissertation and my entire graduate degree. I would also like to thank Mirka Ondrack and Dr. Rob Cribbie for their help in the statistical analysis of this research.

Table of Contents

List of Tables

List of Figures

Introduction

Mental Illness Stigma

The prevalence of individuals living with mental illness around the world is extremely high. It is estimated that one in five people will be diagnosed with a mental illness each year with approximately six percent being classified as severe (Kessler, Chiu, Demler, & Walters, 2005; World Health Organization, 2001a). Individuals affected by mental illness can experience impairment in emotional, social, and occupational functioning (Heatherton, Kleck, Hebl, & Hull, 2000; Stier & Hinshaw, 2007). However, the stigmatization that these individuals face because of their illness can be even more detrimental than the illness itself (Hinshaw, 2007; Markowitz, 1998). Although efforts have been made to increase awareness of and reduce the stigma associated with mental illness in recent years, it unfortunately continues to remain strong and pervasive (Guimon, Fisher, & Sartorius, 1999; Hinshaw & Cicchetti, 2000; Stier & Hinshaw, 2007). In fact, mental illness stigma occurs across nations and cultures around the world (Abdullah & Brown, 2011; Tsang, Tam, Chan, & Cheung, 2003; World Health Organization, 2001b), creating significant barriers to treatment-seeking. Individuals with mental illness often feel shame and are fearful of encountering discrimination (Whal, 1999), which prevents them from seeking needed treatment and ultimately leads to poorer prognosis and reduced quality of life (Markowitz, 1998).

A number of social psychological models of stigma have been described in efforts to better understand how stigma is developed, maintained, and expressed. Like most phenomena in psychology, at the most basic level, stigma is believed to consist of cognitive, affective, and behavioural components (McGuire, 1995). In their conceptualization of stigma, social psychologists distinguish between the constructs of stereotypes, prejudice, and discrimination

(Fiske, 1998). Stereotypes can be described as cognitive representations of particular groups of people that are stored in memory and tend to be formed through socialization (Crocker & Major, 2003). Cognitive representations of individuals with mental illness typically depict them as displaying certain negative characteristics or behaviours (e.g., incompetent, lazy, violent). Prejudice refers to a negative emotional reaction or evaluation that people have toward individuals with mental illness based on the stereotypes they endorse. Lastly, discrimination refers to negative behavioural responses to prejudice directed toward individuals with mental illness (e.g., refusing to hire someone with a mental illness for a job). Although the relations between stereotypes, prejudice, and discrimination may appear straightforward, stigma is considered to be a complex phenomenon that is influenced by various factors and can be broken down into public stigma and self-stigma (Corrigan et al., 2011; Ottati, Bodenhausen, & Newman, 2005). This dissertation focused on public stigma.

Public Stigma

Public stigma refers to discrimination by the general public against individuals with mental illness based on the endorsement of related stereotypes. Research indicates that individuals with mental illness are often perceived to be dangerous, violent, incompetent, and a drain on societal resources (Corrigan & Cooper, 2005; Corrigan, Edwards, Green, Diwan, & Penn, 2001; Wahl, 1999). These stereotypes, in turn contribute to the prejudice and discrimination that individuals with mental illness regularly face, limiting important life opportunities, such as obtaining employment or housing (Corrigan & Kleinlen, 2005; Page, 1995). Research on social distancing, measured by the extent to which people will come into varying degrees of contact with individuals who have a mental illness, suggests that people try to avoid individuals with mental illness across a number of situations (e.g., from sitting in the same

room to having a romantic relationship; Corrigan et al., 2001; Faulkner, 2010; Hartman, Michel, Winter, Young, Flett, & Goldberg, 2013). These preferences for social distance based on negative attitudes toward mental illness have shown to be present in children as young as four years of age and once formed, continue to remain strong years later (Weiss, 1986, 1994).

Self-Stigma

In addition to facing stigma from society, individuals with mental illness are prone to developing self-stigma, in which they internalize society's negative stereotypes about mental illness and apply them to themselves (Corrigan & Watson, 2002; Corrigan, Watson, & Barr, 2006). Self-stigma has been associated with shame, poor self-esteem, and a reduced sense of self-efficacy, making individuals less likely to seek and follow through with treatment (Moses, 2010). Further, due to shame and lack of self-confidence, self-stigma may cause individuals to avoid social contact in anticipation of rejection, which can lead to isolation and unemployment (Link, 1982). Taken together, the stigma of mental illness is prevalent and powerful and continues to cause deleterious effects on the lives of individuals affected by mental illness in various ways.

Mental Illness Stigma as a Contemporary form of Prejudice: Aversive Racism

Over the years, a number of anti-stigma campaigns have been developed and implemented in order to increase awareness of mental illness stigma and reduce its harmful effects. As a result, overt expressions of stigma toward mental illness have become less socially acceptable (Stier & Hinshaw, 2007). However, the fact that stigma still exists suggests that people may have deeper levels of stigma that are more resistant to change and continue to be expressed in indirect, yet nevertheless, harmful ways (Pescosolido et al., 2010; Phelan, Stueve, & Pescosolido, 2000; Stier & Hinshaw, 2007). Along this line of reasoning, research examining

racial prejudice has described a contemporary form of prejudice that is subtle and indirect in nature, and has been termed “aversive racism” (Gaertner & Dovidio, 1986). Whereas traditional forms of racism consist of outwardly expressing prejudicial behaviour based on having negative attitudes and feelings about a particular racial group, aversive racism refers to a type of prejudice experienced by individuals who consciously and genuinely endorse non-prejudiced values, but at the same time possess conflicting, unconscious, negative beliefs and feelings about a particular racial group (Gaertner & Dovidio, 1986; Kovel, 1970; Pearson, Dovidio, & Gaertner, 2009).

Because aversive racists are considered to sincerely aspire to be non-prejudiced and are believed to be unware of their negative unconscious attitudes, they will not act inappropriately in situations with strong social norms when discrimination would be obvious to others and to themselves. Rather this unconscious prejudice will typically be expressed subtly, indirectly, and in situations when it can be personally and explicitly justified by factors other than race (Dovidio & Gaertner, 2004; Gaertner & Dovidio, 1986; Nail, Harton, & Decker, 2003). For example, research has shown that White people tend to help a White or Black person equally in an emergency situation if they are the only witness present at the time. However, if they are among other witnesses and can therefore justify not helping, they are less likely to help a Black person than a White person (Gaertner & Dovidio, 1977; Saucier, Miller, & Doucet, 2005). Similarly, White people tend not to discriminate against hiring a White candidate over a Black candidate if the Black candidate clearly has better qualifications. However, when candidates’ qualifications for the position are less clear (e.g., only moderate qualifications) and the decision can be rationalized based on other factors, White individuals will tend to hire the White candidate more often than the Black candidate with the same credentials (Dovidio & Gaertner, 2000).

Aversive racists are thus characterized as having non-prejudiced conscious, or explicit, attitudes, but negative unconscious, or implicit attitudes (Dovidio & Gaertner, 2004). Explicit attitudes operate in a conscious, controllable, and deliberative manner and can be assessed using traditional self-report measures. In contrast, implicit attitudes operate automatically without awareness and thus must be assessed using more indirect methods (e.g., Bargh, Chaiken, Govender, & Pratto, 1992; Devine, 1989; Greenwald & Banaji, 1995). The Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) is a commonly used reaction time measure of implicit attitudes whereby individuals pair positive (e.g., good) and negative (e.g., bad) attributes with a target variable of interest (e.g., race). It is based on the premise that people are typically faster at categorizing groups of stimuli presented to them that are consistent with their automatic, implicit attitudes. Thus, if an individual is faster to associate negative attributes with mental illness, for example, then they are assumed to have a negative implicit bias toward mental illness. The IAT has been used extensively in racial prejudice research and has illustrated a consistent finding of aversive racism in White individuals in which they tend to appear non-prejudiced on explicit self-report measures, but their responses on implicit measures indicate a negative racial bias that contrasts from their explicit views (Dovidio, Kawakami, & Beach, 2001; Hofmann, Gawronski, Gschwendner, Le, & Schmitt 2005). Despite the widespread use of the IAT in the implicit prejudice literature, there has been considerable debate over the years about the validity of the test, particularly whether the IAT can predict racial discrimination (e.g., Blanton, Jaccard, Klick, Mellers, Mitchell & Tetlock, 2009). Whereas one meta-analysis demonstrated a modest correlation between the IAT and race-related behavioural outcomes (r = .236; Greenwald, Poehlman, Uhlman, and Banaji, 2009), another illustrated a weaker correlation (r = .014; Oswald, Mitchell, Blanton, Jaccard & Tetlock, 2013). Furthermore, Carlsson and

Agerstrom (2016) recently conducted a meta-analysis on previous studies examining the relation between the IAT and discrimination and determined that many of the studies failed to appropriately measure or provide evidence of actual discrimination, thus making it difficult to draw meaningful conclusions about the predictive validity and reliability of the IAT in terms of discriminatory behaviour. Thus, although the IAT has been a highly useful tool in understanding implicit associations, its ability to reliably predict actual behavioural outcomes is less clear and should be considered when interpreting such findings from the literature.

Implicit versus Explicit Stigma

Findings regarding the differences in explicit and implicit prejudice highlight the importance of incorporating implicit measures into the assessment of mental illness stigma. Like racism and other forms of prejudice, it is becoming increasingly less socially acceptable to express stigma openly and given that explicit measures are prone to social desirability bias (Link & Cullen, 1983), it is likely that they will underestimate true levels of stigma (Stier & Hinshaw, 2007). In addition, research has shown that explicit and implicit measures appear to represent independent constructs with important differences found between these two types of processing (Bargh, Chaiken, Govender, & Pratto, 1992; Devine, 1989; Greenwald & Banaji, 1995; Greenwald & Farnham, 2000). For example, implicit measures have been shown to be more predictive of behaviours that are automatic or spontaneous (i.e., nonverbal behaviours), whereas explicit measures are more predictive of controlled behaviours (e.g., verbal communication) (Asendorpf, Banse, & Mucke, 2002; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997). Thus, in order to accurately measure mental illness stigma, it must be assessed both in terms of one's conscious, deliberate, and explicit attitudes as well as by tapping into their deeper, automatic, and implicit belief systems (Stier & Hinshaw, 2007).

Implicit Mental Illness Stigma

Despite the limitations of relying solely on explicit measures of stigma as well as support for the existence of implicit stigma toward mental illness, only recently has research been extended to begin to investigate mental illness stigma using implicit methods. One of the first studies to examine stigma indirectly was conducted by Graves and colleagues (2005) who measured psychophysiological responses to labels of mental illness (Graves, Cassisi, & Penn, 2005). They found that participants displayed increased physical reactivity when shown labels of schizophrenia compared to those with no diagnosis. These responses, in turn, predicted greater preferences for social distance against individuals labeled as having schizophrenia suggesting that exposure to the label of a serious mental illness triggers automatic responses that are likely to negatively influence subsequent behaviour (Stier, & Hinshaw, 2007). Teachman and colleagues (2006) later compared explicit and implicit stigma of mental illness in individuals both with and without a mental illness and were the first to measure implicit stigma using the IAT (Teachman, Wilson, & Komarovskaya, 2006). The results of their study showed that both individuals with and without mental illness displayed more negative explicit and implicit attitudes toward mental illness relative to physical illness as well as an overall stronger absolute negative implicit bias toward mental illness on the IAT. In addition, and consistent with previous research on explicit and implicit processing, they found that the explicit and implicit measures were uncorrelated, suggesting that they tapped independent constructs (Stier & Hinshaw, 2007).

A number of studies have since examined stigmatizing attitudes toward mental illness both explicitly and implicitly, which further illustrate the importance of assessing stigma using explicit and implicit measures. For example, Monteith and Pettit (2011) examined explicit and implicit stigmatizing attitudes about depression in a sample of undergraduate students and found

that overall, more negative attitudes about depression were shown on implicit, but not explicit measures, relative to physical illness. Studies have also investigated whether mental health professionals harbour negative attitudes toward individuals with mental illness compared to the general public as well as how these stigmatizing attitudes may influence subsequent clinical decision-making. Results from Peris and colleagues (2008) showed that compared to people without mental health training, mental health professionals demonstrated more positive implicit and explicit evaluations of individuals with mental illness (Peris, Teachman, & Nosek, 2008). However, within mental health professionals, negative bias was shown to predict clinical decision making such that explicit (but not implicit) stigma influenced these professionals to make more negative patient prognoses, whereas implicit (but not explicit) stigma influenced them to over-diagnosis patients. Similarly, Stull and colleagues (2013) found that the mental health professionals in their study exhibited both positive explicit and implicit attitudes toward individuals with mental illness (Stull, McGrew, Salyers, & Ashburn-Nardo, 2013). However, they also found that despite their overall positive attitudes toward those with mental illness, stigmatizing attitudes influenced subsequent clinical care whereby greater implicit (but not explicit) bias predicted greater endorsement of restrictive clinical interventions compared to those that allow the patient more control over their own recovery (e.g., medication monitoring by a professional versus self-monitoring). Kopera and colleagues (2014) conducted another study investigating explicit and implicit attitudes toward individuals with mental illness among non-professionals (medical students) with no previous contact with patients with mental illness and mental health professionals (psychiatrists and psychotherapists) who had at least two years of professional contact with patients with mental illness (Kopera, et al., 2014). Rather than using the IAT, the authors used the Go/No-Go Association Task (GNAT; Nosek & Banaji, 2001) to

measure implicit stigma. This is another type of implicit association task that does not require the use of a comparison category. Assessing implicit attitudes within the context of a comparison category can be problematic because measurement of implicit attitudes toward one group becomes biased by the selection of the contrasting category (e.g., implicit attitudes toward mental illness *relative to* physical illness). Consistent with the previous studies, the results illustrated that mental health professionals reported more positive attitudes toward mental illness compared to non-professional on explicit measures. However, both groups endorsed negative implicit attitudes towards mental illness on the implicit measure.

Taken together, these studies illustrate that even mental health professionals who have direct contact with and treat individuals with mental illness are not immune to the implicit biases about mental illness that much of the general public tend to endorse. Similar to aversive racists and consistent with a contemporary prejudice framework, it is conceivable that individuals who report that they are not stigmatizing, yet harbour implicit stigmatizing attitudes toward mental illness are unaware of their negative bias and genuinely strive to be non-stigmatizing (i.e., "aversive stigmatizers"). However, when these negative biases are expressed, they can impact the lives of individuals with mental illness in subtle, but significant ways (e.g., reduced patient care). Therefore, it is important for research to continue to examine stigma implicitly as well as aim to develop strategies and interventions to help reduce implicit stigma that continues to negatively affect individuals living with mental illness.

Interventions to Reduce Stigma

The stigma-reduction programs that have been implemented thus far have traditionally focused on changing conscious attitudes and overt discrimination (i.e., explicit stigma). These generally tend to include strategies such as education, which involves challenging the myths of

mental illness in order to enhance mental health literacy, and contact with an individual with mental illness. Although education can be effective in reducing explicit stigma, contact appears to be the most promising approach to stigma change and can augment the effects of education (Corrigan & Penn, 1999). However, because of the subtle and complex nature of implicit stigma, traditional interventions for reducing overt stigma are likely to be ineffective for combating implicit stigma (Pearson et al., 2009). Like aversive racists, individuals who report non-stigmatizing attitudes, while harbouring a negative implicit bias toward mental illness likely already believe stigma is harmful, they just do not believe that *they* are stigmatizing. Thus, other techniques are required in order to reduce and ultimately eliminate implicit forms of mental illness stigma. To my knowledge, no studies have yet been designed to reduce implicit mental illness stigma. However, numerous studies exist in the racial prejudice literature that have been effective at reducing implicit racial bias, which may also be applied to the implicit stigma associated with mental illness.

Interventions to Reduce Implicit Racial Prejudice

In the prejudice literature, methods for reducing implicit racial prejudice have been implemented at both the unconscious and conscious levels of awareness, both of which have shown to be effective through different processes (Lai, Hoffman, & Nosek, 2013). Non-conscious methods tend to involve directly changing implicit attitudes through basic associative conditioning. This technique is based on the assumption that implicit attitudes are believed to reflect associations between concepts (e.g., Black/White) and evaluations (e.g., good/bad; Greenwald et al., 2002) whereby participants learn to associate concepts with attributes that differ from their preexisting attitudes to create alternative (non-prejudice) attitudes (Bar-Anan, De Houwer, & Nosek, 2010; Karpinski & Hilton, 2001; Olson & Fazio, 2006). For example,

Olson and Fazio (2006) found that briefly exposing participants to positive images and words paired with Black faces, and negative images and words paired with White faces reduced implicit racial prejudice immediately and remained effective two days later (Olsen & Fazio, 2006). In addition, based on the assumption that people tend to approach things that are good and avoid things that are bad, studies have shown that training participants with a computer program to engage in various approach behaviours toward Black individuals and avoidance behaviours toward White individuals reduced implicit prejudice toward Black individuals (presumably because the self is considered "good"; e.g., Kawakami, Phills, Steele, & Dovidio, 2007).

In contrast to the methods that work to change implicit attitudes without one's awareness, strategies at the conscious level aim to target individuals' non-prejudiced intentions and motivations by making them aware of their implicit bias (Pearson et al., 2009). Research has shown that when individuals low in explicit prejudice are told they may have a tendency to act with prejudice due to their implicit negative bias, the inconsistency between their behaviour and their personal standards produces feelings of guilt (Devine, Monteith, Zuwerink, & Elliot, 1991). This, in turn, triggers a type of self-regulatory process akin to Festinger's (1957) dissonance reducing theory, that motivates people to restore internal balance by not responding with prejudice in the future (Amodio, Devine, & Harmon-Jones, 2007; Devine & Monteith, 1993; Festinger & Carlsmith, 1959; Monteith, 1993; Montheith & Voils, 1998). Furthermore, continuing to act in ways that are consistent with their non-prejudiced standards can produce lasting changes in negative implicit attitudes and behaviour over time (Devine, Forscher, Austin, & Cox, 2012; Dovidio, et al., 2000; Pearson, et al., 2009). This effect has been found across various studies and appears to be strongest for individuals who score low in explicit prejudice and high in implicit prejudice (i.e., aversive racists). Compared to individuals who score high on

both explicit and implicit measures of prejudice (highly prejudiced) and those who score low on both measures (non-prejudiced), aversive racists tend to have the largest discrepancies between their behaviour (what they would do) and personal standards (what they should do), feel the most guilt, and are thus more motivated to engage in self-regulatory processes that facilitate balance and reduce their implicit negative bias (Dovidio et al., 2000; Green, et al., 2007; Son Hing, Li, & Zanna, 2002).

Devine and colleagues (2012) investigated the effects of an intervention designed to produce long-term reduction in implicit racial prejudice consistent with this line of research (Devine, Forscher, Austin, & Cox, 2012). Their intervention was based on a dual-processing model that views implicit biases as deeply rooted habits developed through socialization experiences. They argue that in order to be motivated to break this prejudice habit, people must first become aware of their biases and then feel concerned about the consequences of their biases (Devine, 1989; Devine & Monteith, 1993; Devine et al., 1991; Monteith, 1993). The intervention consisted of several components, including feedback about participants' implicit racial biases to make them aware of their biases as well as a bias education program intended to evoke concern about how their implicit racial biases can lead to discrimination. The intervention also contained a bias training program to train participants to eliminate their negative implicit biases through engaging in one of various bias-reducing strategies of their own choosing (e.g., stereotype replacement, perspective-taking, increasing opportunities for contact). Results of their study illustrated that the intervention was effective in reducing implicit racial bias and this effect lasted for up to eight weeks after the intervention was initially implemented. Specifically, the intervention helped increase participants' personal awareness of their racial bias and general

concern about discrimination in society, which highlights the importance of conscious awareness and effort in order to reduce implicit bias in the long-term.

In summary, although a large body of work has demonstrated the effectiveness of intervention strategies to reduce implicit racial prejudice at both the conscious and non-conscious levels of processing, no studies have investigated whether these or any intervention strategies will be effective in reducing implicit stigma toward mental illness. Given that the literature has shown increasing evidence for the existence of implicit stigma toward mental illness as well as its potentially harmful effects on behaviour, it is necessary for researchers to examine how to reduce people's implicit negative biases associated with mental illness that serve to perpetuate mental illness stigma and discrimination. It is comprehensible that the same underlying processes involved in the development and maintenance of implicit racial prejudice also apply to implicit mental illness stigma, therefore, making it appropriate to generalize the contemporary prejudice literature, theoretical frameworks, and intervention strategies to implicit mental illness stigma.

Current Program of Research

Using a contemporary prejudice framework, the present program of research aimed to fill gaps in the mental illness stigma literature by 1) examining stigmatizing attitudes both explicitly and implicitly in order to better understand how stigma operates and is expressed at both levels of processing and 2) creating a laboratory intervention to reduce implicit stigmatizing attitudes toward mental illness. The purpose of Study 1 was to investigate how explicit and implicit attitudes toward mental illness influence discrimination. As well, consistent with the contemporary prejudice literature, Study 1 examined the interaction between explicit and implicit stigmatizing attitudes toward mental illness to identify the "aversive stigmatizers" (individuals who score low on explicit measures and high on implicit measures of stigma) and how their

behaviour may differ from the high stigmatizers (high scores on both explicit and implicit measures of stigma) and low stigmatizers (low scores on both explicit and implicit measures of stigma).

The purpose of Study 2 was to build on the results of Study 1 and fill the current gap in the mental illness stigma literature by developing a laboratory intervention designed to reduce implicit negative attitudes toward mental illness. Based on the findings of the importance of conscious awareness in producing effective and long-term change in implicit attitudes (e.g., Devine et al., 2012), the intervention was aimed at attempting to change implicit attitudes at the conscious level rather than through unconscious processes (e.g., Kawakami et al., 2007; Olson & Fazio, 2006). The intervention in Study 2 was based on a modified version of Devine et al.'s, (2012) intervention to reduce implicit racial prejudice. Although Devine and colleagues' (2012) intervention was effective overall in reducing long-term implicit racial prejudice, their intervention incorporated a number of different bias-reducing strategies that participants engaged in based on their personal preferences, thus making it impossible to determine which strategies were more or less effective in reducing bias overall. Study 2 aimed to extend the work of Devine and colleagues (2012) by isolating and testing three of the intervention strategies that were utilized in their study in order to determine the strategies that are most effective in reducing implicit mental illness stigma.

Study 1

As discussed, the purpose of Study 1 was to examine the relation between explicit and implicit attitudes (and their interaction) and discrimination toward mental illness in terms of participants' tendency to avoid and help individuals with mental illness.

Hypotheses

Based on previous research examining explicit and implicit mental illness stigma (e.g., Koopera et al., 2014; Teachman et al., 2006), it was hypothesized that 1) overall, participants would have more negative implicit attitudes toward mental illness than positive and that 2a) stigmatizing implicit and explicit attitudes toward mental illness would predict greater levels of discrimination (i.e., more avoidance and less helping). Lastly, based on findings from the contemporary prejudice literature on aversive racism (e.g., Gaertner & Dovidio, 1986), it was hypothesized that 2b) explicit and implicit attitudes toward mental illness would interact to influence differences in behaviour. Specifically, the behaviour of aversive stigmatizers (high implicit, low explicit) would differ compared to high stigmatizers (high implicit, high explicit) and low stigmatizers (low implicit, low explicit) in terms of their level of discrimination toward mental illness. Aversive stigmatizers were expected to show less discrimination compared to high stigmatizers given that they will actually try not to be stigmatizing, but were expected to show more discrimination compared to low stigmatizers given that their negative automatic attitudes toward mental illness may still influence them to act in stigmatizing ways without their complete awareness.

Method

This study conforms to the standards outlined in the Canadian Tri-Council research ethics guidelines. It was also reviewed and granted ethics approval by the Human Participants Review Sub-Committee of the Office of Research Ethics at York University.

Participants

Participants were recruited through York University's Undergraduate Research Participant Pool (URPP), an online system that provides introductory psychology students with

course credit in exchange for their participation. The sample size was determined by conducting an apriori power analysis using a moderate effect size, $f^2 = .15$, $\alpha = .05$, 2 predictors, and 80% power, N=68. In anticipation of exclusions, 114 participants registered and participated in this study, but only 65 participants were included in the analyses (see below for exclusion criteria). Of these 65 participants, 54 were female and 11 were male who were, on average, 21 years old (*SD* = 4.31). Approximately 33% of participants identified their ethnicity as South Asian, 26% East Asian, 19% White, 14% Black, and 3% Latin American. Approximately 75% were single, 20% in a committed relationship, 3% common law, and 1 % married.

Participant exclusion criteria. Participants were excluded from the analyses if they 1) did not complete both parts of the study (i.e. the implicit stigma reaction time measure and the explicit stigma questionnaire measure) and 2) if it was determined that they were not appropriately responding to both the implicit and explicit stigma measures. Specifically, participants were removed if they received negative response times (an impossibly fast response) on the implicit stigma reaction time measure (i.e., the GNAT), indicating that they failed to respond to the stimuli presented to them. Participants were also removed based on criteria from the Conscientious Responders Scale (CRS; Marjanovic, Struthers, Cribbie, & Greenglass, 2014). The CRS is a five-item validity measure that uses instructional items dispersed throughout the questionnaire to identify random responders (i.e., participants who respond to items without paying attention to what they mean; Marjanovic et al., 2014). Because each item instructs responders exactly how to answer that particular question, such as, “To answer this question, please choose option number four, neither agree nor disagree” (based on a 7-point Likert scale from 1=*strongly disagree*, 7=*strongly agree*), it is assumed that if participants answered three or more items incorrectly, they were not attending to the items appropriately. Of the participants

who were excluded from the analyses, 15 were due to not completing both components of the study (GNAT and questionnaire), 28 were due to inappropriate responses on the GNAT, and six were removed based on the CRS criteria.

Measures and Covariates

Explicit attitudes. Explicit stigmatizing attitudes toward mental illness were measured with the Mental Illness Stigma Scale (MISS; Day, Edgren & Eshleman, 2007). The MISS is a 28-item questionnaire measure that assesses different factors related to mental illness stigma with seven subscales, including Anxiety, Relationship Disruption, Hygiene, Visibility, Treatability, Professional Efficacy, and Recovery. The 7-item Anxiety subscale measures feelings of discomfort or fear when in the company of an individual with mental illness (e.g., "When around someone with a mental illness, I worry that he or she may harm me physically"). The 6-item Relationship Disruption subscale assesses beliefs about the influence of mental illness on relationships (e.g., "A close relationship with someone with a mental illness would be like living on an emotional roller coaster"). The 4-item Hygiene subscale targets stereotypes about whether individuals with mental illness maintain their hygiene (e.g., "People with mental illness do not groom themselves properly"). The 4-item Visibility subscale reflects beliefs that individuals with mental illness can be identified by the way they look or act (e.g., "I can tell that someone has a mental illness by the way he or she talks"). The 3-item Treatability subscale assesses beliefs about the effectiveness of pharmacological treatments for mental illness (e.g., "There are effective medications for mental illnesses that allow people to return to normal and productive lives"). The 2-item Professional Efficacy subscale examines whether respondents believe that mental health professionals can deliver effective care (e.g., "Mental health professionals, such as psychiatrists and psychologists, can provide effective treatments for mental illnesses"). Lastly,

the 2-item Recovery subscale assesses respondents' beliefs about the ability of individuals to recover from mental illness (e.g., "Once someone develops a mental illness, he or she will never be able to recover from it"). All items are measured on a Likert scale from 1 (strongly disagree) to 7 (strongly agree). The Treatability, Professional Efficacy, and Recovery subscales were reversed scored for interpretation purposes such that for all scale items, higher scores reflected more stigmatizing attitudes toward mental illness. Most of the MISS subscales have been shown to have acceptable internal consistency psychometric properties ($\alpha > .70$) (Day et al., 2007; Masuda et al., 2009). However, one study reported lower internal consistency for the Visibility and Treatability subscales ($\alpha = 0.68$ and $\alpha = 0.64$ respectively; Stone & Merlo, 2011). Similarly, in the current study, the psychometric properties of all of the subscales were within the acceptable range ($\alpha_{Anxiety} = .89$; $\alpha_{Relationship\ Disruption} = .91$; $\alpha_{Hygiene} = .82$; $\alpha_{Visibility} = .73$; $\alpha_{Professional\ Efficacy} = .88$; $\alpha_{Recovery} = .82$) except for the Treatability subscale ($\alpha_{Treatability} = .68$).

For the purposes of this study, a total MISS score was computed since it was of most interest to examine participants' overall level of stigmatizing attitudes toward mental illness rather than any particular stigmatizing belief. When computing the total MISS score, inter-item correlations among each of the subscales were examined, which revealed that the Professional Efficacy and Visibility subscales were weakly and/or negatively correlated with several of the other scale items. Factor analysis was then completed to further examine whether the subscale items loaded onto one factor solution (MISS total score). All of the subscale items except those from the Professional Efficacy and Visibility subscales loaded onto the MISS total score factor solution and thus the items from these subscales were removed from the MISS total score. After removing these items, Cronbach's alpha for the MISS total score was .91 and the total variance

of the remaining items explained by the MISS total score improved from 33.31% to 46.00% (with acceptable variance being considered to equal at least 40%).

Implicit attitudes. Implicit stigmatizing attitudes toward mental illness were measured with the Go/No-Go Association Task (GNAT; Nosek & Banaji, 2001). Like other measures of implicit attitudes, (e.g., IAT), the GNAT assesses implicit attitudes or beliefs based on the strength of association between a target category (e.g., mental illness) and two poles of an attribute dimension (e.g., dangerous-harmless). However, the GNAT differs from other measures in that it is based on Signal Detection Theory whereby the strength of association is measured by the degree to which items belonging to the target category and attribute can be discriminated from distracter items that do not belong to those concepts (Nosek & Banaji, 2001). One condition requires the responder to simultaneously identify stimuli that represent both the target category (mental illness) and one dimension of the attribute (dangerous) and a second condition requires simultaneous identification of stimuli that represent the same target category (mental illness) and the opposite dimension of the attribute (harmless). The GNAT requires the same response - "go" (press the space bar) to items that belong to the category (mental illness) and a particular evaluative attribute (dangerous) both of which serve as the “signal.” No response - "no-go" (do not press the space bar) is indicated when items appear that do not belong to the target category and attribute (noise). Whether participants more strongly associate the target category (mental illness) with one particular attribute dimension (e.g., dangerous) compared to the other (harmless) is based on their relative ease of discriminating that target category (mental illness) with one attribute (dangerous) versus the other (harmless) in these two conditions (Nosek & Banaji, 2001). For example, if participants have stronger automatic associations between mental illness and dangerous compared to mental illness and harmless, accuracy in

discriminating mental illness and dangerous from distracters should be higher than accuracy in discriminating mental illness and harmless items from distracters. The difference in accuracy of discriminating signal from noise (also known as "sensitivity") between these conditions is taken as the measure of automatic attitude (measured by *d* prime).

The GNAT has various methodological advantages over traditional implicit attitude measures, such as the IAT (Greenwald et al., 1998), namely that it does not require the use of a comparison category in its measurement of automatic attitudes (e.g., Whites compared to Blacks, men compared to women). According to Nosek and Banaji, “the IAT requirement of the presence of a second attitude object directly in the measurement context constrains the interpretation of the effect to the particular comparison object” (Nosek & Banaji, 2001, p. 629). For example, attitudes toward Blacks may differ when the comparison target category is White versus Asian versus Hispanic, etc. In addition, associations made between the target variable and a particular attribute (e.g., Black and bad) are measured at the same time as the comparison category (e.g., White and good) and therefore cannot be analyzed separately. Because associations to both variables occur simultaneously, it makes it difficult to determine whether faster responses are due to stronger associations between Black and bad or White and good as a positive attitude toward Whites does not necessarily imply an opposite negative attitude toward Blacks. Therefore, the GNAT was chosen to measure automatic attitudes toward mental illness in this study because it eliminates the aforementioned problems with the traditional IAT and it was of interest to examine attitudes toward mental illness as a single category.

The GNAT was constructed based on the recommendations of Nosek and Banaji (2001) when measuring attitudes toward a single category (i.e. mental illness). To assess automatic attitudes toward mental illness, participants were instructed to associate mental illness with the

attributes 1) negative versus positive, 2) dangerous versus harmless, and 3) helpless versus competent. These three attributes were chosen in order to examine participants' general automatic attitudes (i.e. negative or positive) toward mental illness as well as stereotypical beliefs commonly held about individuals with mental illness (i.e., that they are dangerous or incompetent). Automatic attitudes toward the target category, mental illness were assessed during the critical blocks when the target category and each attribute were simultaneously presented on the screen. During the critical blocks, participants were presented with two category labels on the computer screen at the same time (e.g., mental illness and negative), followed by a single stimulus presented in the middle of the screen. In this study, the stimuli presented in the middle of the screen were words belonging to either of the two categories as well as distracters. The distracters consisted of the opposite attribute for each attribute pair that was being assessed. For example, when assessing the association between mental illness and negative, positive was the distracter and vice versa. Participants were asked to determine whether or not the stimulus in the middle of the screen belonged to one of the two categories whose labels appeared at the top of the screen. If the stimulus belonged to either of these categories (e.g., mental illness and negative), participants were told to press the space bar (the 'Go' response). If the stimulus was a distracter item and did not belong to either of the two categories (e.g., positive), they were told not to press anything (the 'No-go' response). Error feedback was given after each trial indicating whether they correctly distinguished items belonging to the target versus distracter. Specifically, a green "O" appeared for a correct response (i.e., the participant chose the 'Go' response to targets and the 'No-go' response to distracters) and a red "X" for an incorrect response (i.e., the participant chose the 'Go' response to distracters or the 'No-go' response to targets). Stronger associations between mental illness and negative is indicated by relatively greater sensitivity (i.e.

accuracy) when the category mental illness was paired with the attribute negative, compared with to positive. The same procedure was applied when assessing participants' automatic associations between mental illness and dangerous/harmless and mental illness and helpless/competent. Each attribute pairing consisted of 60 trials (20 practice trials and 40 critical trials) for a total of 180 trials that were completed. The time deadline for participants to respond to the stimuli was 750 ms and the inter-stimulus interval (ISI; i.e., time between the presentation of each new stimulus) was 500 ms. Please see appendix A for the all of the stimuli used in the GNAT for the mental illness target category and the three attributes.

Social distance. The Social Distance Scale (SDS; Link et al., 1999) was used as a proxy measure for behavioural manifestations of mental illness stigma (i.e., discrimination). In the original measure, respondents were asked to read a vignette describing a person with mental illness and then asked to rate their willingness to interact with the person in various ways, such as living next door to the person, socializing with the person, and working with the person. Consistent with other research (Livingston, Tugwell, Korf-Uzan, Cianfrone, & Coniglio, 2013; Penn, Chamberlin, & Mueser, 2003), the current study omitted the vignette and had the participants only respond to the questionnaire items based on their willingness to interact with "an individual with a mental illness." The SDS consists of seven items (e.g., "How would you feel about working at the same job as someone with a mental illness?") that are traditionally rated on a 4-point scale from 0 (definitely not willing) to 3 (definitely willing). However, to keep all of the questionnaire scale measures consistent, items were instead rated on a 7-point scale, ranging from 1 (definitely not willing) to 7 (definitely willing). Items were reverse scored, such that higher scores indicated a greater tendency to engage in discriminatory behaviour toward individuals with mental illness related to avoidance. The total SDS score is the mean of the item

scores. The SDS has shown to have good psychometric properties (i.e., $\alpha > .80$; Cheon & Chiao, 2012; Livingston et al., 2013) and Cronbach's alpha in the current study was excellent ($\alpha = .90$).

Helping behaviour. In addition to obtaining a measure of participants' tendency to avoid individuals with mental illness (through social distance) as a proxy measure of discriminatory behaviour, it was also of interest to examine participants' actual willingness to help individuals with mental illness by allowing them to choose whether to donate to a mental illness versus a physical illness charity. Participants were told that the researchers would be donating $50 to support either the Canadian Mental Health Association or the Canadian Diabetes Association depending on the total number of votes that participants allocate to each organization. Specifically, participants were given a total of 10 votes that they could distribute to each organization however they wished (e.g., five votes each, 10 votes to one and none to the other, etc.) and that the organization that received the most votes at the end of the study would receive the $50 donation (a donation was actually made by the researchers to both organizations). Helping behaviour was measured based on the number of votes given to the Canadian Mental Health Association compared to the Canadian Diabetes Association (i.e., with a difference score), such that more votes indicated more helping behaviour.

Covariates. In addition to the measures described above, participants completed two scales that assess constructs known to be associated with reduced self-reported mental illness stigma, including familiarity with mental illness (Couture & Penn, 2003; Holmes, Corrigan, Williams, Canar, & Kubiak, 1999) and social desirability bias (e.g., Hartman, et al., 2013; Stier & Hinshaw, 2007).

Level of familiarity with mental illness (Measured with the Level of Contact Report (LOCR; Corrigan et al., 2005; Holmes, et al., 1999). To determine familiarity with mental

illness, participants were administered a version of the LOCR, which is a self-report measure that provides respondents with a list of situations in which they may have encountered individuals with mental illness. Situations are ranked in order of increasing familiarity with mental illness. For example, the item "I have watched a television show that included a person with mental illness" is ranked below the item "I live with a person who has a mental illness." For each item, respondents select "true" or "false" to indicate whether or not they have been in contact with an individual with mental illness in the situations described. Respondents are assigned a single rank order score, which reflects the most intimate interaction they have experienced with individuals with mental illness. Scores range from 0 (least familiar, e.g., "I have never observed a person with mental illness") to 11 (most familiar, e.g., "I have a mental illness, or have had one at some point in my life").

Marlowe-Crowne Social Desirability Scale – Short Form (Ballard, 1992; Crowne & Marlowe, 1960). Given that the expression of mental illness stigma is considered less socially acceptable in today's society, it is likely that participants will underreport stigmatizing attitudes they may hold about individuals with mental illness (Henderson, Evans-Lacko, Flach, & Thornicroft, 2012). To account for the effect of a socially desirable response style on explicit self-reported attitudes toward mental illness, a short version of the Marlowe-Crowne Social Desirability Scale (Ballard, 1992) was included. The Marlowe-Crowne Social Desirability Scale has been used in thousands of studies since its development. Although the measure's psychometric properties have occasionally been challenged due to low internal consistency (Hartman et al., 2013; Loo & Loewen, 2004; Loo & Thorpe, 2000, $\alpha = .61$), Ballard's (1992) short-form has been identified as the best of the short versions (Loo & Lowen, 2004) and was used in this study, which produced relatively good internal consistency ($\alpha = .76$). The scale

consists of 11-items that assess participants' tendency to respond in a socially desirable manner (e.g., "I am always willing to admit when I make a mistake") based on true or false responses. In an effort to maintain consistency across scale measures (i.e., continuous) and improve the psychometric properties of the measure by allowing for more of a range of responses, the items were measured on a 7-point scale, ranging from 1 (strongly disagree) to 7 (strongly agree), instead of the traditional true or false response options.

Procedure

Participants were provided a URL that allowed them to complete the entire study online. In order to avoid demand characteristics and reduce the likelihood of socially desirable responding, they were told that they were participating in a study examining attitudes and experiences about mental illness rather than stigma. After first providing their informed consent to participate in the study, participants were instructed to complete the GNAT, which was described as a reaction time task, followed by the questionnaire. After completing both parts of the study, they were directed to a debriefing screen explaining the full details about the nature and purpose of the study. They were also provided contact information to contact the researchers with any questions or concerns, none of which were expressed by any of the participants.

Data Analytic Strategy

The data were analyzed in SPSS (24). Paired samples t-tests were first conducted to examine hypothesis 1, whether participants had stronger negative implicit attitudes toward mental illness than positive as well as stronger implicit associations between mental illness and dangerous compared to harmless, and helpless compared to competent. Multiple linear regression was conducted to determine hypothesis 2, the extent to which explicit and implicit attitudes (and their interaction) toward mental illness predict discriminatory behaviour based on

each of the DV's (social distance and helping behaviour). To examine hypothesis 2, a composite implicit attitudes predictor variable (GNAT total score) was created and included in the regression analysis that combined the three implicit attitude categories (negative/positive, dangerous/harmless, helpless/competent) because it was of interest to examine the relation between general implicit attitudes (rather than specific categories) and discriminatory behaviour and also for interpretive purposes. To create the composite implicit attitudes variable, the difference scores between each attitude category were used (i.e., the difference between the *d* prime scores of negative versus positive, dangerous versus harmless, and helpless versus competent) in order to represent the complete dimension of each attribute. The difference scores are interpreted such that higher scores indicate stronger negative implicit attitudes toward mental illness. Each of the DV's were regressed on both the explicit and composite implicit attitude predictors as well as the interaction between the two while controlling for level of familiarity with mental illness and social desirability. Prior to conducting the analyses, a preliminary diagnostic assessment of the data was completed, which determined that the assumptions of conducting linear regression were met, including normality, homogeneity of variance, and linearity as well as ensured that no influential outliers were present. An alpha level of .05 was used for all tests of significance.

Results

Descriptive Statistics

A correlation matrix as well as the means and standard deviations of the explicit and implicit attitude variables, the level of familiarity and social desirability covariates, and the social distance and helping behaviour DV's are listed in Tables 1 and 2, respectively.

Hypothesis 1

According to hypothesis 1, participants were predicted to have stronger negative implicit attitudes toward mental illness compared to positive as well as stronger implicit associations between mental illness and dangerous compared to harmless, and mental illness and helpless compared to competent**.** As predicted, a paired samples t-test revealed that there was a significant difference in implicit attitude scores between negative (M = 2.12, SD = .45) and positive (M = 1.95, SD = .58), such that participants had significantly stronger negative attitudes toward mental illness compared to positive, $t(63) = 2.85$, $p<.01$, d = .32. In addition, a significant difference was found between the attributes dangerous (M = 1.85, SD = .40) and harmless (M = 1.41, SD = .52), such that participants had significantly stronger associations between mental illness and dangerous compared to harmless, $t(63) = 6.49$, $p<.001$, d =.95. Lastly, a significant difference was found between the attributes helpless (M = 2.01, SD = .55) and competent (M = 1.62, SD = .54), such that participants had significantly stronger associations between mental illness and helpless compared to competent, $t(63) = 5.25$, $p<.001$, d=.72.

Hypothesis 2a

According to hypothesis 2a, implicit and explicit attitudes toward mental illness were expected to predict discriminatory behaviour (based on more social distance and less helping behaviour). The results of the regression analyses illustrated that overall, stigmatizing attitudes significantly predicted discriminatory behaviour toward mental illness for both social distance, F (5, 53) = 14.26, p = <.0001, R^2=.533, and helping behaviour, F (5, 53) = 2.61, p = .035, R^2=.122 (see Tables 3 and 4, respectively). Consistent with hypothesis 2a, greater negative implicit and explicit attitudes predicted greater levels of social distance, b = 2.34, t = 2.15,

p <.05, b = 1.02, t = 6.38, p<.001, respectively, toward mental illness as well as less helping behaviour, b = -15.61, t = -2.78, p <.01, b = -1.78, t = -2.17, p <.05, respectively.

Hypothesis 2b

According to hypothesis 2b, implicit and explicit attitudes toward mental illness were predicted to interact to influence discriminatory behaviour such that aversive stigmatizers (i.e., those who score low on negative explicit attitudes, but high on negative implicit attitudes toward mental illness) would show less discrimination than high stigmatizers (those who score high on both negative implicit and explicit attitudes), but more discrimination than low stigmatizers (those who score low on both negative implicit and explicit attitudes). As predicted, there was a significant interaction between implicit and explicit attitudes toward mental illness on both social distance, b = -6.87, t = -2.02, p =.045, 95%CI [-1.37, -.003], and helping behaviour, b = 3.92, t = 2.23, p =.030, 95% CI [.393, 7.44] (see Figures 1 and 2, respectively). Specifically, participants characterized as aversive stigmatizers displayed significantly more social distance, b = 1.12, t = 2.40, p = .019, compared to participants characterized as low stigmatizers, F (1, 53) = 13.88, p<0.001, and less (but not significantly) social distance compared to high stigmatizers, F (1, 53) = 0.602, p = 0.441. Regarding the helping behaviour DV, aversive stigmatizers showed significantly less helping behaviour toward mental illness, b = -7.06, t = -2.91, p < .01, compared to both low stigmatizers, F (1, 53) = 8.13, p = 0.006 and high stigmatizers, F (1, 53) = 5.22, p = 0.026.

Table 1

Study 1 Correlation Matrix of Mental Illness Stigma Predictors, Outcome Variables, and Covariates

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. MISS Anxiety	1	.35**	.68**	.39**	.39**	.34**	-.16	.85**	.02	.08	.22	.17	.58**	-.04	-.20	-.11
2. MISS Visibility		1	.36**	.39**	.22	.25*	.04	.41**	-.13	.04	.28*	.12	.40**	-.20	-.08	-.10
3. MISS Rel'p Disruption			1	.54**	.50**	.55**	.10	.91**	.10	.10	.15	.17	.66**	-.18	-.13	-.10
4. MISS Hygiene				1	.29*	.36**	.17	.67**	.22	.10	.16	.23	.44**	-.20	-.10	-.04
5. MISS Recovery					1	.48**	.19	.61**	.01	.06	-.03	.01	.38**	-.05	.11	-.01
6. MISS Treatability						1	.31*	.62**	.19	.03	.06	.13	.33**	-.04	.01	-.02
7. MISS Prof Efficacy							1	.08	.15	.21	-.01	.16	-.10	-.38**	.15	.10
8. MISS Total Score [a]								1	.15	.16	.18	.24	.70**	-.15	-.17	-.07
9. GNAT Negative [b]									1	.05	.33**	.65**	.11	-.09	-.06	.33**
10. GNAT Dangerous [b]										1	.14	.59**	.25*	-.18	-.18	-.05
11. GNAT Helpless [b]											1	.76**	.11	-.35**	-.31*	.19
12. GNAT Total Score												1	.24	-.32*	-.28*	.23
13. Social Distance													1	-.21	-.23	-.11
14. Helping Behaviour														1	-.03	-.10
15. Level of Familiarity															1	-.17
16. Social Desirability																1

Note. MISS = Mental Illness Stigma Scale; GNAT = Go/No-Go Association Task; [a] excludes items related to the Professional Efficacy and Visibility subscales, [b] consist of difference scores of both attribute dimensions including negative/positive, dangerous/harmless, and helpless/capable; $p < .05$ (2-tailed); * $p < .01$ (2-tailed); ** $p < .001$ (2-tailed).

Table 2

Study 1 Means and Standard Deviations for Mental Illness Stigma Predictors, Outcome Variables, and Covariates

Variable	Mean	SD
MISS Anxiety	3.14	1.21
MISS Visibility	3.62	1.12
MISS Rel'p Disruption	3.02	1.26
MISS Hygiene	2.66	1.10
MISS Recovery	2.58	1.29
MISS Treatability	2.78	1.03
MISS Prof Efficacy	2.75	1.14
MISS Total Score [a]	2.94	.93
GNAT Negative [b]	.17	.49
GNAT Dangerous [b]	.44	.54
GNAT Helpless [b]	.39	.60
GNAT Total Score	.33	.36
Social Distance	2.85	1.12
Helping Behaviour	1.98	4.15
Level of Familiarity	8.60	2.36
Social Desirability	4.09	.88

Note. MISS = Mental Illness Stigma Scale; GNAT = Go/No-Go Association Task; [a] excludes items related to the Professional Efficacy and Visibility subscales, [b] consist of difference scores of both attribute dimensions including negative/positive, dangerous/harmless, and helpless/capable.

Table 3

Study 1 Main Effects and Interaction of Regression Analysis for Social Distance

Variable	*B*	*SE*(B)	*t*	*p*	95% CI
Implicit Attitudes	2.34	1.09	2.15	.036*	(.159, 4.53)
Explicit Attitudes	1.02	.16	6.38	.000***	(.699, 1.34)
Implicit x Explicit Attitudes	-.69	.34	-2.02	.049*	(-1.37, -.003)
Level of Familiarity	-.09	.05	-1.83	.072	(-1.82, .008)
Social Desirability	-.04	.12	-.33	.744	(-.287, .206)

Note. $*p < .05$, $***p < .001$

Table 4

Study 1 Main Effects and Interaction of Regression Analysis for Helping Behaviour

Variable	*B*	*SE*(B)	*t*	*p*	95% CI
Implicit Attitudes	-15.61	5.61	-2.78	.007**	(-26.87, -4.35)
Explicit Attitudes	-1.78	.82	-2.17	.035*	(-3.44, -.132)
Implicit x Explicit Attitudes	3.92	1.76	2.23	.030*	(.393, 7.44)
Level of Familiarity	-.33	.24	-1.36	.179	(-.824, .158)
Social Desirability	-.58	.64	-.92	.364	(-1.85, .692)

Note. $*p < .05$, $**p < .01$

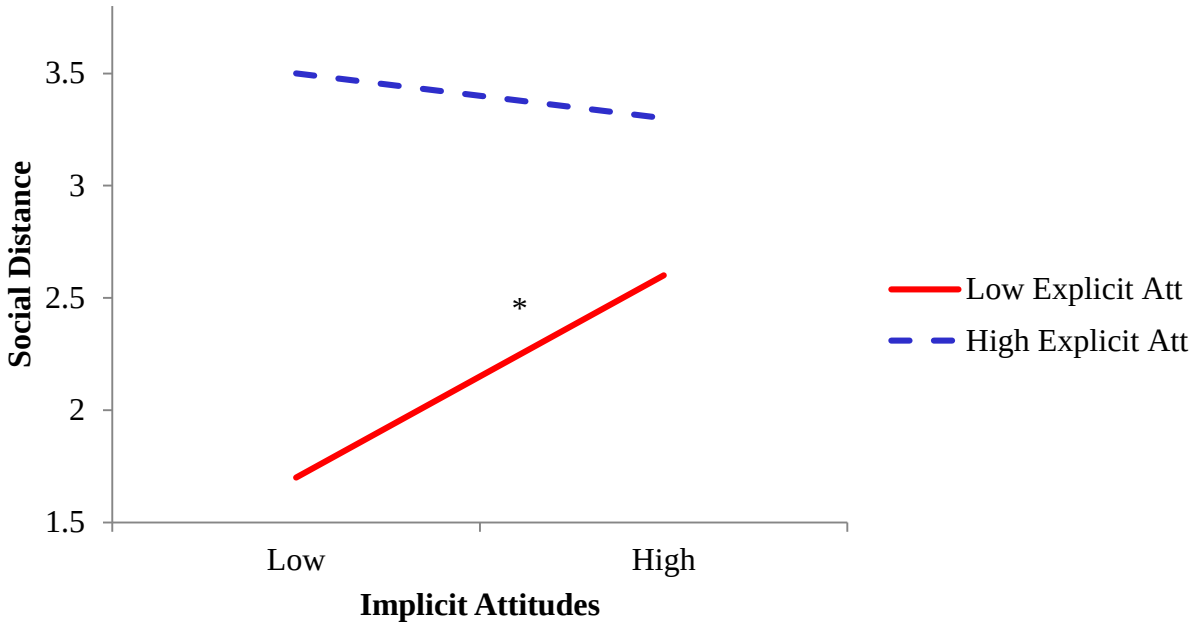

Figure 1. Interaction between implicit and explicit attitudes on social distance for Study 1 (controlling for level of familiarity and social desirabilty).
Note. *p<.05

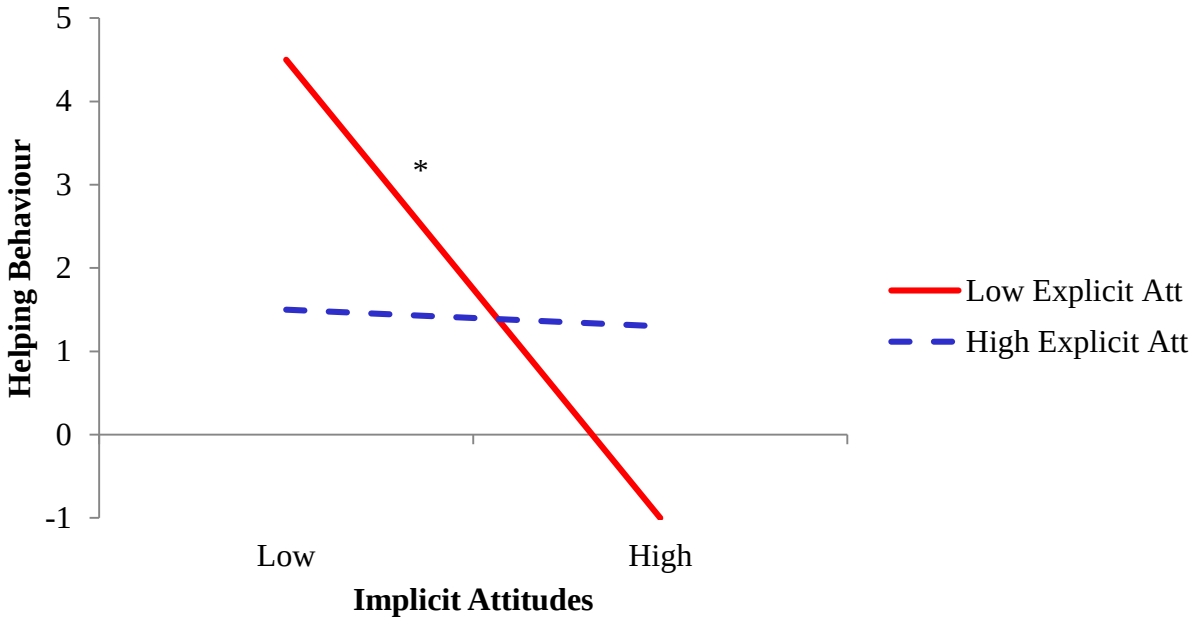

Figure 2. Interaction between implicit and explicit attitudes on helping behaviour for Study 1 (controlling for level of familiarity and social desirabilty).
Note. Helping behaviour is represented by a difference score between donations made to a mental illness versus a physical illness charity such that higher scores indicate more donations to the mental illness charity; *p<.05

Study 1 Discussion

The results of study 1 demonstrated that overall, participants tended to have a negative automatic bias toward mental illness as they displayed automatic associations with mental illness and negative, dangerous, and helpless attributes. As well, the results illustrated that both negative implicit and explicit attitudes predicted discriminatory behaviour toward mental illness, including the tendency to avoid individuals with mental illness as measured through social distancing as well as the tendency to be less willing to help individuals with mental illness by providing more donations to a diabetes charity compared to a mental illness charity.

These results also provide support for a contemporary form of mental illness stigma based on the finding that the interaction between participants' implicit and explicit attitudes toward mental illness predict how stigma may be expressed. Specifically, individuals who explicitly report to be non-stigmatizing, yet harbour negative automatic attitudes about mental illness (aversive stigmatizers) tend to be less avoidant of individuals with mental illness compared to high stigmatizers, but more avoidant compared to low stigmatizers. The behaviour of aversive stigmatizers also differs when it comes to their willingness to help individuals with mental illness, such that compared to the high and low stigmatizers, they were the least willing to donate to the mental illness charity compared to the diabetes charity. These results appear to be consistent with the contemporary prejudice literature on aversive racism whereby the behaviour of aversive racists tends to be subtle and is typically expressed in ambiguous situations when it can be justified (Dovidio & Gaertner, 2004; Dovidio & Gaertner, 2009, Gaertner & Dovidio, 1986). For example, perhaps aversive stigmatizers may only engage in social distancing when it does not appear overtly discriminatory (e.g., choosing not to sit next to an individual on the subway who may have a mental illness when other seats are available) and are less willing to

help individuals with mental illness in situations when it can be rationalized, such as choosing to donate to a physical illness rather than a mental illness charity given that they can still be perceived as helping.

Taken together, the results of Study 1 help provide insight into the complex nature of mental illness stigma by demonstrating that not only is negative automatic bias toward mental illness prevalent, but that the way in which it is expressed appears to depend on the combination of people's implicit and explicit attitudes. This may be particularly problematic for the aversive stigmatizers (high implicit, low explicit) who do not self-identify as stigmatizing, but because they are unaware of their automatic bias, will continue to express subtle forms of discrimination that reinforce and perpetuate stigma as well as violate their own personal standards and morals.

Study 2

Given the existence of implicit mental illness stigma and its negative influence on individuals with mental illness that often go unnoticed, the aim of Study 2 was to build on and extend the results of Study 1 by designing and implementing an intervention to reduce implicit stigmatizing attitudes toward mental illness, which has not yet been examined. The study was based on Devine and colleagues' (2012) intervention to reduce long-term implicit racial prejudice. As mentioned previously, although the results of Devine et al.'s (2012) study demonstrated that the intervention was effective overall in reducing implicit racial bias, it is not known which aspects of the intervention were more or less effective in the bias-reducing process. Thus, this study incorporated three aspects of Devine et al.'s (2012) intervention applied to mental illness stigma to determine which component is most effective in reducing negative implicit attitudes toward mental illness, including 1) Education, 2) Bias Feedback, and 3) Contact.

Intervention to Reduce Implicit Mental Illness Stigma

Education. The education component of the intervention involved educating participants about the existence and nature of negative automatic bias toward mental illness and how it can influence discrimination. Education about explicit mental illness stigma and its negative impact on individuals with mental illness has shown to be effective in reducing explicit negative attitudes toward mental illness (e.g., Corrigan & Penn, 1999; Hanisch, Twomey, Szeto, Birner, Nowak, & Sabariego, 2016; Hartman et al., 2013; Penn, Kommana, Mansfield, & Link, 1999). However, regarding implicit stigma, education is likely to be particularly important because most people tend to be generally unaware of automatic biases, how they are developed and expressed, and the impact that they have on affected individuals (Devine et al., 2012). Thus, the first step in changing automatic biases is to make people aware of them.

Bias feedback. The bias feedback component consisted of providing feedback to participants about their own implicit bias toward mental illness of which they are likely unaware. Once people are made aware of their own implicit biases, if they are inconsistent with their personal standards (i.e., they genuinely do not want to be biased), the feedback should create internal dissonance/discomfort and therefore motivation to reduce the bias (Amodio et al, 2007; Devine, 1989; Devine & Monteith, 1993).

Contact. The contact component involved providing virtual interpersonal contact with an individual with mental illness (through watching a video). Contact with mental illness has consistently shown to be one of the most effective methods at reducing explicit mental illness stigma among the general public (e.g., Corrigan, Morris, Michaels, Rafacz, & Rüsch, 2012; Corrigan & Penn, 1999; Hartman et al., 2013). However, no studies have examined if the effects

of contact will be similarly effective in reducing implicit negative attitudes toward mental illness, which was one of the goals of this study.

Four different conditions were created to examine the effectiveness of the different intervention components in reducing negative automatic attitudes toward mental illness. Condition 1 included the Education and Bias Feedback components. These components were combined into one condition rather than testing them separately given that previous research has illustrated that only making people aware of their bias without providing any context (e.g., education) is not enough to cause a significant reduction in negative bias (Devine et al., 2012). Condition 2 included the Education, Bias Feedback, and Contact components because it was of interest to examine whether contact would augment the effects of education and bias feedback. Condition 3 included Contact only to determine if contact with mental illness on its own would produce similar positive effects in reducing implicit attitudes as has been shown with explicit attitudes. Condition 4 consisted of a control condition to determine if the intervention components were at least more effective than no intervention at all (further detail about the intervention is described in the procedure section).

Hypotheses

It was hypothesized that 1) Education, Feedback, and Contact (Condition 2) would be most effective compared to the other conditions in reducing automatic negative attitudes toward mental illness given that it combined all of the intervention components. Education and Feedback (Condition 1) was predicted to be more effective than Contact only (Condition 3) and the Control condition (Condition 4) given that in order for individuals to reduce their automatic bias, they must first become aware of these biases. Finally, the Contact only condition (Condition 3) was predicted to be more effective than the Control condition (Condition 4).

In addition to investigating an effective intervention to reduce negative automatic attitudes toward mental illness, another aim of this study was to examine whether the intervention would be more or less effective for certain types of people in reducing discrimination. Specifically, the second purpose of this study was to determine if the effectiveness of the intervention would differ for people characterized as high stigmatizers, low stigmatizers, and aversive stigmatizers in terms of their tendency to avoid or help individuals with mental illness. Previous research on implicit racial bias has found that interventions designed to reduce implicit prejudice have been most effective for aversive racists (i.e., those who score low on explicit measures, but high on implicit measures of racism) compared to individuals who are highly prejudiced (i.e., score high on both explicit and implicit measures of prejudice) and non-prejudiced (those who score low on both measures). This is believed to be because aversive racists tend to have the largest discrepancies between their behaviour and personal standards and are therefore more motivated to restore consistency and reduce their bias (Devine & Monteith, 1993; Dovidio et al., 2000; Green, et al., 2007; Son Hing, Li, & Zanna, 2002). Based on these findings, it was predicted that 2) the intervention overall would be most effective for aversive stigmatizers in reducing discrimination (i.e., less social distance and more helping behaviour) compared to high stigmatizers and low stigmatizers. The contemporary prejudice literature has not traditionally examined the group of individuals who score high on implicit prejudice, but low on explicit prejudice presumably because it is an unlikely combination as it would seem counter-intuitive to report prejudicial attitudes, while unconsciously endorsing non-prejudicial automatic attitudes. However, given that this was the first study to examine the effectiveness of an intervention to reduce discrimination for the interaction between implicit and explicit attitudes, the low implicit, high explicit attitude

combination (which is described as “intentional stigmatizers”) was included in the analysis for exploratory purposes, though no specific predictions were made about this particular group.

Thus, the aim of Study 2 was to answer two main research questions: 1) Will the intervention be effective in reducing negative automatic attitudes toward mental illness and if so, which components of the intervention are most effective? 2) How will the effectiveness of the intervention differ for people characterized as high stigmatziers, low stigmatizers, aversive stigmatizers, and intentional stigmatizers in terms of reducing discriminatory behaviour toward mental illness?

Method

This study conforms to the standards outlined in the Canadian Tri-Council research ethics guidelines. It was also reviewed and granted ethics approval by the Human Participants Review Sub-Committee of the Office of Research Ethics at York University.

Participants

As in Study 1, participants were recruited through York University’s URPP and received course credit in exchange for their participation. A power analysis was conducted prior to collecting the data, which determined that a total of 179 participants were needed based on 80% power to obtain a medium effect size. 220 participants registered and participated in this study and 195 were included in the analyses (see below for exclusion criteria). Of these 195 participants, 143 were female and 50 were male (two did not report their gender) who were, on average, 20 years old (*SD* = 4.79). Approximately 20% of participants identified their ethnicity as South Asian, 20% East Asian, 25% White, 15% Middle Eastern, 8% Black, and 2% Latin American. Approximately 65% were single, 30% in a committed relationship, 3 % married, and 1% common law.

Participant exclusion criteria. Participants were excluded from the analyses based on the same criteria as Study 1, including inappropriate responding to either the GNAT (based on an impossibly fast response time) or the questionnaire measures (based on criteria from the CRS; Marjanovic et al., 2014). Given that this study involved participating at two different time points (before and after receiving the intervention), participants were removed if they did not complete the study at both time points. Of the participants who were excluded from the analyses, 10 were due to inappropriate responding on the GNAT, three were based on the CRS criteria, and 14 were due to not completing both part 1 and 2 of the study.

Measures and Covariates

All of the measures used to assess the variables in Study 1 were also used in Study 2.

Explicit attitudes. Explicit stigmatizing attitudes toward mental illness were measured with the MISS (Day, et al., 2007). The psychometric properties of all of the seven subscales that comprise the MISS were within the acceptable range ($\alpha_{Anxiety} = .94$; $\alpha_{Relationship\ Disruption} = .89$; $\alpha_{Hygiene} = .84$; $\alpha_{Visibility} = .79$; $\alpha_{Professional\ Efficacy} = .78$; $\alpha_{Recovery} = .72$) except for the Treatability subscale ($\alpha_{Treatability} = .57$), which was also the case in Study 1. As in Study 1, a total MISS score was computed as it was of most interest to examine participants' general explicit stigmatizing attitudes toward mental illness rather than any particular stigmatizing belief. When computing the total MISS score, the inter-item correlations among each of the subscales were examined, which again revealed that the Professional Efficacy and Visibility subscales were weakly and/or negatively correlated with several of the other scale items. Factor analysis was then completed to further examine whether the subscale items loaded onto one factor solution (MISS total score). As was the case in Study 1, all of the subscale items except those from the Professional Efficacy and Visibility subscales loaded onto the MISS total score factor solution and thus the items from

these subscales were removed from the MISS total score. After removing these items, Cronbach's alpha for the MISS total score was .94 and the total variance of the remaining items explained by the MISS total score improved from 37.38% to 46.06%.

Implicit attitudes. Implicit stigmatizing attitudes toward mental illness were measured with the GNAT (Nosek & Banaji, 2001). Negative/positive, dangerous/harmless, and helpless/competent implicit attitudes toward mental illness were assessed and computed in the same manner as Study 1, including using *d* prime to compute difference scores for each attitude (e.g., negative *d* prime minus positive *d* prime, etc.), such that higher scores indicate more automatic negative bias.

Social distance. Participants' tendency to avoid individuals with mental illness to varying degrees was assessed with the Social Distance Scale (α=.93, as a proxy measure of behavioural forms of discrimination (Link et al., 1999).

Helping behaviour. Participants' tendency to help individuals with mental illness was measured based on their willingness to donate money to the Canadian Mental Health Association compared to the Canadian Diabetes Association (again, a donation was actually made to both organizations by the researchers). As in Study 1, difference scores were computed between the number of votes assigned to the mental health versus diabetes associations and interpreted such that higher scores indicate more helping behaviour toward mental illness.

Covariates. As in Study 1, participants' tendency to respond in a socially desirable manner as well as their level of familiarity with mental illness were controlled for in the analyses given that they are known to be associated with reduced self-reported stigma toward mental illness. Social Desirability was measured with the Marlowe-Crown Social Desirability Scale –

Short Form (α=.76; Ballard, 1992) and Familiarity with Mental Illness was measured with the Level of Contact Report (Corrigan et al., 2005).

Procedure

Part 1. Participants completed the study in the laboratory on a computer. An experimenter was present throughout the duration of the study to ensure participants were appropriately attending to the study and to answer any questions. Upon entering the lab, participants were told that the entire study would be completed on the computer and to follow the instructions provided. After indicating their student ID, they were directed to the informed consent screen where they were explained that the purpose of the study was to examine individuals' attitudes and experiences with mental illness. Participants were not initially told the true purpose of the study to avoid demand characteristics. After providing their informed consent to participate in the study, participants completed the implicit attitudes measure (GNAT), followed by an online questionnaire that contained the explicit attitudes measure, covariates, and measures of discrimination (i.e., social distance and helping behaviour). After completing part 1 of the study, participants were reminded to return to complete the second part of the study at a later date. To ensure the length of time between completing the time 1 and time 2 measures was consistent, all participants completed part 2 of the study approximately one week after part 1.

Part 2. Part 2 of the study was also completed in the laboratory on a computer with an experimenter present. After arriving to the lab, participants were randomly assigned to one of the four intervention conditions, which they received on the computer. In Condition 1 (Education and Bias Feedback), participants read information on how automatic negative attitudes are developed, how they contribute to subtle forms of discrimination toward mental illness, as well as how to begin to reduce negative automatic bias (see Appendix B for full description of

Condition 1). The Education component was based on Devine et al.,'s (2012) education portion of their intervention to reduce implicit racial prejudice and was modified to apply to mental illness for this study. After the Education component, participants were told that they would receive feedback on their GNAT score to indicate the extent to which they have a negative automatic bias toward mental illness. Given that most people have shown to have a negative automatic bias toward mental illness relative to positive, a pattern which was verified in Study 1, all participants were given the same feedback regardless of how they scored on the GNAT. The feedback message participants received included: "The results of your score indicate that at an automatic level, you have a relative negative bias toward mental illness compared to positive on the IAT." The purpose of informing participants that they have a negative automatic bias toward mental illness was to produce feelings dissonance in those that genuinely do not want to be biased, thus providing motivation to reduce the bias. In Condition 2 (Education, Bias Feedback, and Contact), participants received the same information in Condition 1 followed by the contact portion. The Contact condition involved watching a TED Talk video online of a young male university student describing his experience with depression as well as the stigma he faced as a result (Breel, 2013). The intention of the video was to disprove negative stereotypes of mental illness and elicit concern about the effects of discrimination, both of which are factors believed to be involved in reducing stigma toward mental illness through contact (Corrigan & Penn, 1999: Couture & Penn, 2006). This particular video was also chosen because it was believed to be a good fit with the target audience in terms of promoting their level of engagement and connection with the individual with mental illness. A video was chosen as the form of contact as it was not feasible to provide face-to-face contact with an individual with mental illness for each participant as well as to keep the contact as controlled as possible. Moreover, research on the effectiveness

of varying types of interpersonal contact in reducing mental illness stigma has shown that forms of virtual contact can be just as effective as in person (e.g., Chan, Mak, & Law, 2009; Reinke, Corrigan, Leonhard, Lundin, & Kubiak, 2004). In Condition 3 (Contact only), participants only watched the video of the person with mental illness. In Condition 4 (Control), participants read an article about emotions and psychology.

After receiving the intervention or control conditions, participants completed the GNAT followed by the questionnaire measure. Once the questionnaire was completed, participants were fully debriefed and explained the true purpose of the study.

Data Analytic Strategy

The General Linear Model (GLM) procedure in SPSS (24) was used to analyze the data given that the regression model included both continuous and categorical variables and the GLM automatically dummy codes the categorical variables as part of the analysis. To examine hypothesis 1, a mixed 4 (intervention conditions) x 2 (time) ANCOVA (controlling for level of familiarity with mental illness) was conducted for each implicit attitude to first determine if there were significant differences between the conditions across time in terms of their effectiveness in reducing automatic negative attitudes. This was then followed-up with multiple comparison tests to determine which condition was most effective for each implicit attitude. To examine hypothesis 2, a 4 (implicit/explicit attitudes) x 2 (intervention vs control) ANCOVA (controlling for social desirability and the time 1 DV's) was conducted for each DV (social distance and helping behaviour) to determine if there were significant differences between the four categories of implicit/explicit attitude combinations (high implicit/high explicit, high implicit/low explicit, low implicit/high explicit, low implicit/low explicit) after receiving the intervention compared to the control condition. Multiple comparison tests were then conducted to determine for which

implicit/explicit attitude combination the intervention was most effective in reducing discrimination (based on the social distance and helping behaviour DV's). For these analyses, the four intervention conditions were collapsed into two groups (intervention versus control) in order to make the comparisons between implicit and explicit attitudes and the intervention easier to examine and interpret. In this case, the intervention group consisted of Conditions 1 (Education and Bias Awareness) and 2 (Education, Bias Awareness, and Contact) and the control group consisted of Condition 3 (Contact only) and the Control condition. As will be seen in the results below, the decision to collapse the particular conditions in this manner was based on the effectiveness of each of the conditions and to ensure that the sizes of the two groups were relatively equal (i.e., conditions 1 and 2 were most effective overall and condition 3 was not).

As in Study 1, a composite implicit attitudes predictor variable (GNAT total score) was created that combined the three implicit attitudes (negative/positive, dangerous/harmless, helpless/competent) because it was of interest to examine the relation between general implicit and explicit attitudes with discriminatory behaviour. In creating the four implicit/explicit attitude combinations, a median split was used to identify the top and bottom 50^{th} percentiles of participants considered to be high and low, respectively, on implicit and explicit stigmatizing attitudes. A median split has been used in other research (e.g., Amodio et al., 2007; Son Hing et al., 2002) to categorize the implicit/explicit attitude groups and was used in this study to examine any significant differences between the groups for the pairwise comparisons.

Prior to conducting the analyses, a preliminary diagnostic assessment of the data was completed, which determined that the assumptions of conducting linear regression were met, including normality, homogeneity of variance, and linearity. Influential outliers were found to be present when analyzing the 2-way ANCOVA's for hypothesis 2 for both the social distance and

helping behaviour DV's based on the requirements of Cook's Distance for influential observations (i.e., high leverage and high studentized residuals; Cook, 1977) and were therefore removed from the analyses. An alpha level of .05 was used for all tests of significance.

Results

Descriptive Statistics

A correlation matrix as well as the means and standard deviations of all of the variables are listed in Tables 5 and 6, respectively.

Hypothesis 1

According to hypothesis 1, the intervention was predicted to be effective in reducing negative automatic attitudes toward mental illness for each implicit attitude. Condition 2 (Education, Bias Feedback, and Contact) was expected to be most effective, followed by Condition 1 (Education and Bias Feedback), followed by Condition 3 (Contact only). As predicted, the intervention was effective in reducing automatic negative attitudes toward mental illness for each implicit attitude (see Tables 7 – 9). As well, results of the 2-way interaction contrasts illustrated significant differences between the conditions in terms of their effectiveness in reducing each implicit attitude, though not entirely as predicted (see Table 10 and Figures 3-5). For the Negative/Positive implicit attitude, Condition 1 (Education and Bias Feedback) was significantly more effective in reducing automatic associations between mental illness and negative (compared to positive) than the Control Condition (Condition 4) (see Table 7).

Table 5

Study 2 Correlation Matrix of Mental Illness Stigma Predictors, Outcome Variables, and Covariates at Time 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. MISS Anxiety	1	.15*	.73**	.64**	.41**	.42**	-.02	.92**	.13	.08	-.02	.10	.65**	-.010	-.44**	-.14
2. MISS Visibility		1	.24**	.24**	.08	.01	-.13	.21**	.10	-.06	.07	.07	.18*	.11	.07	-.02
3. MISS Rel'p Disruption			1	.70**	.40**	.41**	.04	.90**	.10	.13	.07	.07	.66**	-.11	-.36**	-.10
4. MISS Hygiene				1	.35**	.39**	.05	.81**	.12	-.02	-.10	.002	.50**	-.08	-.28**	-.12
5. MISS Recovery					1	.41**	.11	.54**	.03	.02	-.03	.01	.31**	.03	-.21**	-.10
6. MISS Treatability						1	.34**	.57**	-.01	.14*	-.09	-.02	.32**	-.14	-.23**	-.13
7. MISS Prof Efficacy							1	.07	-.05	.05	-.12	-.07	.10	-.15*	-.02	-.14*
8. MISS Total Score [a]								1	.12	.10	-.04	.10	.69**	-.10	-.42**	-.15*
9. GNAT Negative [b]									1	.05	.03	.64**	.17*	.02	-.14	.12
10. GNAT Dangerous [b]										1	.002	.54**	.11	-.11	-.08	.04
11. GNAT Helpless [b]											1	.59**	.02	.08	.03	.07
12. GNAT Total Score												1	.17*	.003	-.10	.13
13. Social Distance													1	-.14	-.38**	-.03
14. Helping Behaviour														1	.09	.01
15. Level of Familiarity															1	.005
16. Social Desirability																1

Note. MISS = Mental Illness Stigma Scale; GNAT = Go/No-Go Association Task; [a] excludes items related to the Professional Efficacy and Visibility subscales, [b] consist of difference scores of both attribute dimensions including negative/positive, dangerous/harmless, and helpless/capable; $p < .05$ (2-tailed); * $p < .01$ (2-tailed); ** $p < .001$ (2-tailed).

Table 6

Study 2 Means and Standard Deviations of all Variables at Times 1 and 2

	Time 1		Time 2	
Variable	**Mean**	**SD**	**Mean**	**SD**
MISS Anxiety	3.04	1.45	2.94	1.36
MISS Hygiene	2.64	1.21	2.56	1.20
MISS Relp Disruption	3.04	1.37	2.81	1.27
MISS Treatability	2.82	.98	2.71	1.00
MISS Recovery	2.50	1.19	2.75	1.31
MISS Visibility	3.96	1.21	3.67	1.16
MISS Prof Efficacy	2.97	1.30	3.03	1.30
MISS Total Score[a]	2.90	1.07	2.77	1.02
GNAT Negative[b]	.38	.31	.01	.49
GNAT Dangerous[b]	.23	.57	.38	.55
GNAT Helpless[b]	.51	.59	.26	.52
GNAT Total Score	.38	.55	.22	.33
Social Distance	2.91	1.45	2.85	1.38
Helping Behaviour	1.91	4.05	2.08	4.18
Social Desirability	3.99	.96		
Level of Familiarity	8.77	2.29		

Note. Social Desirability and Level of Familiarity were only measured at Time 1 given that these variables were unlikely to change over that time period. MISS = Mental Illness Stigma Scale; GNAT = Go/No-Go Association Task; [a] excludes items related to the Professional Efficacy and Visibility subscales, [b] consist of difference scores of both attribute dimensions including negative/positive, dangerous/harmless, and helpless/capable.

Table 7

Differences Between Each Intervention Condition on Negative Implicit Attitudes

Conditions	***df***	***F***	**Error**	***p***	η^2
1 vs. 2	1	1.453	100	.231	.014
1 vs. 3	1	1.962	92	.165	.021
1 vs. 4	1	8.405	92	.005*	.084
2 vs. 3	1	.291	97	.591	.003
2 vs. 4	1	2.453	97	.121	.025
3 vs. 4	1	1.383	89	.243	.015

Note. *p<.05

Table 8

Differences Between Each Intervention Condition on Dangerous Implicit Attitudes

Conditions	*df*	*F*	Error	*p*	η^2
1 vs. 2	1	1.138	100	.289	.011
1 vs. 3	1	6.569	92	.012*	.067
1 vs. 4	1	1.299	92	.257	.014
2 vs. 3	1	10.483	97	.002**	.098
2 vs. 4	1	4.787	97	.031*	.047
3 vs. 4	1	1.300	89	.257	.014

Note. **p<.01, *p<.05

Table 9

Differences Between Each Intervention Condition on Helpless Implicit Attitudes

Conditions	*df*	*F*	Error	*p*	η^2
1 vs. 2	1	1.089	100	.299	.011
1 vs. 3	1	1.114	92	.294	.012
1 vs. 4	1	.155	92	.695	.002
2 vs. 3	1	.024	97	.877	.000
2 vs. 4	1	.484	97	.488	.005
3 vs. 4	1	.474	89	.493	.005

Table 10

Effectiveness of Intervention Conditions on Reducing Each Implicit Attitude Between Time 1 and 2.

	Implicit Attitudes								
	Negative vs Positive			**Dangerous vs Harmless**			**Helpless vs Competent**		
Condition	**T1**	**T2**	**D**	**T1**	**T2**	**D**	**T1**	**T2**	**D**
1	.324	-.075	.400[a]***	.544	.321	.223[a]*	.411	.192	.219*
2	.244	.009	.236*	.562	.202	.360[bc]***	.272	.211	.061
3	.229	.039	.190	.361	.476	-.114[ac]	.382	.318	.064
4	.127	.106	.020[a]	.542	.491	.051[b]	.428	.269	.159

Note. Values are based on estimated marginal means of each implicit attitude controlling for level of familiarity with mental illness; Condition 1 = education and bias feedback, 2 = education, bias feedback, and contact, 3 = contact only, 4 = control; T1= pre-intervention, T2=post intervention; D = Mean difference of implicit attitude between T1 and T2; [a, b, c] represent conditions that significantly differ from one another for each implicit attitude. Values with the same letter indicate the specific conditions that significantly differ from each other; ***p<.001, * p<.05.

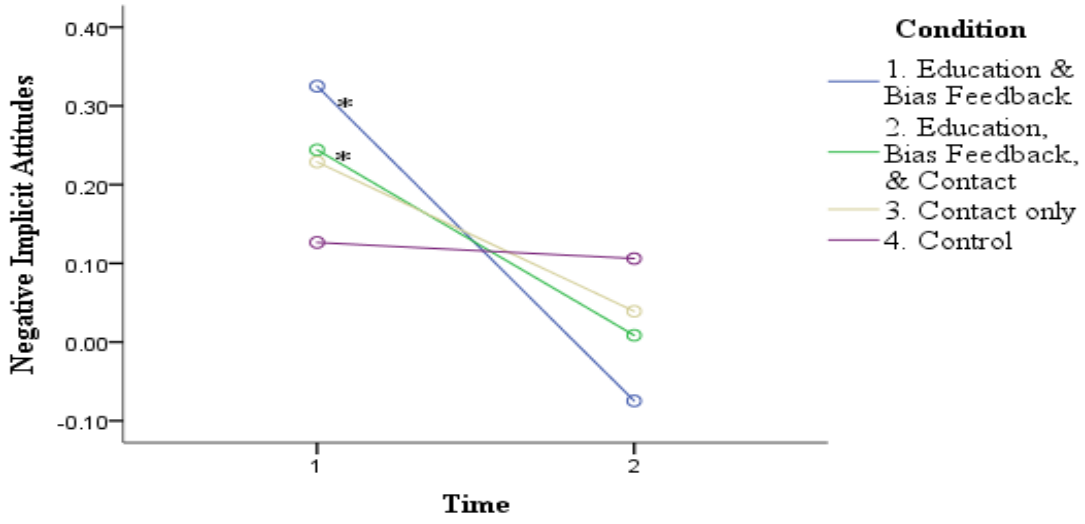

Figure 3. Effect of each condition on negative implicit attitudes pre (Time 1) and post intervention (Time 2) for Study 2 (controlling for level of familiarity).
Note. *p<.05

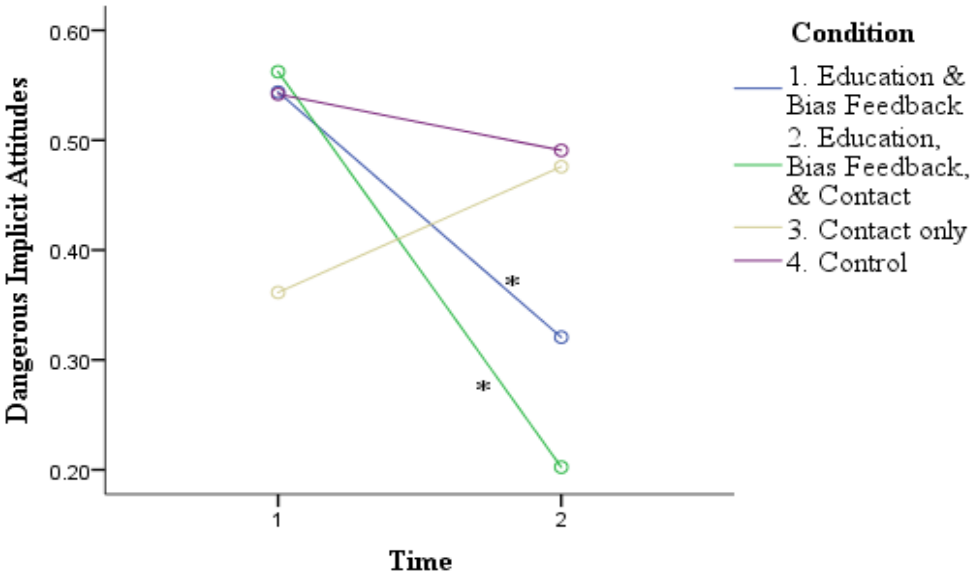

Figure 4. Effect of each condition on dangerous implicit attitudes pre (Time 1) and post intervention (Time 2) for Study 2 (controlling for level of familiarity).
Note. *p<.05

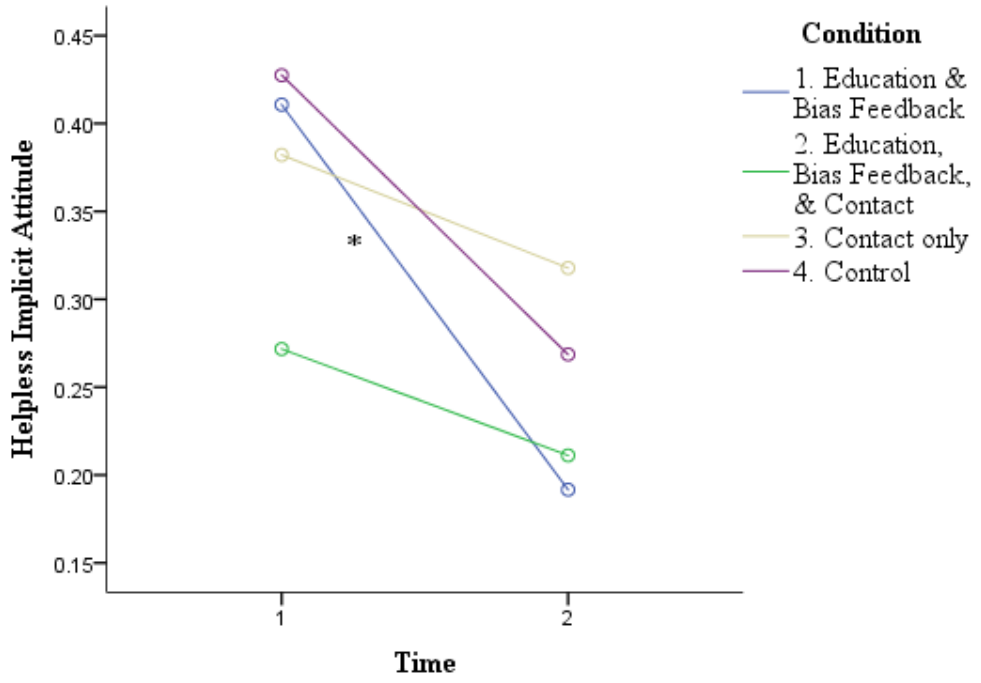

Figure 5. Effect of each condition on helpless implicit attitudes pre (Time 1) and post intervention (Time 2) for Study 2 (controlling for level of familiarity).
Note. *p<.05

Although no other conditions significantly differed from each other in terms of their effectiveness, as can be seen from Table 10 and Figure 3, Conditions 1 (Education and Bias Feedback) and 2 (Education, Bias Feedback, and Contact) caused significant reductions in automatic associations between mental illness and negative (compared to positive), with Condition 1 causing the most change (mean difference = .400). For the Dangerous/Harmless attitude, Condition 1 (Education and Bias Feedback) was significantly more effective than Condition 3 (Contact only), and Condition 2 (Education, Bias Feedback, and Contact) was significantly more effective than Conditions 3 (Contact only) and 4 (Control) (see Table 8). As can be seen from Table 10 and Figure 4, Conditions 1 (Education and Bias Feedback) and 2

(Education, Bias Feedback, and Contact) caused significant reductions in automatic associations between mental illness and dangerous (compared to harmless), with Condition 2 causing the most change (mean difference = .360). For the Helpless/Competent attitude, none of the conditions significantly differed from each other in terms of their effectiveness (Table 9). However, as can be seen from Table 10 and Figure 5, Condition 1 caused significant reductions in automatic associations between mental illness and helpless (compared to competent; mean difference = .219). Given that Conditions 1 and 2 were found to be more effective than Conditions 3 and 4 (Condition 3 did not produce any significant reductions in implicit attitudes on its own), Conditions 1 and 2 were combined to create the intervention condition and Conditions 3 and 4 were combined to create the control condition to examine hypothesis 2.

Hypothesis 2

According to hypothesis 2, it was predicted that there would be significant differences between aversive stigmatizers, high stigmatizers, low stigmatizers, and intentional stigmatizers in terms of the effect of the intervention on their level of social distance and helping behaviour. Specifically, aversive stigmatizers were expected to have the most reduction in social distance and increase in helping behaviour after receiving the intervention (compared to the control condition) compared to the other implicit/explicit attitude combinations.

Social distance DV. Results of the 2-way ANCOVA for the social distance DV illustrated a non-significant interaction between implicit/explicit attitudes and intervention/control, F (3,169) = .723, p = .539, η^2=.013. Although the 2-way interaction was not significant, it remained of interest to examine patterns among the variables given that this was the first study to examine this hypothesis with the goal of better understanding the nature and expression of implicit mental illness stigma. Moreover, it has been argued that omnibus tests are

not always sensitive enough to detect significant smaller order effects (e.g., Maxwell & Delaney, 2004, p. 236). As a result, pairwise comparisons were conducted to further examine specific differences between the (implicit/explicit attitude) groups in terms of their level of social distance after receiving the intervention compared to the control condition (please see Table 11 and Figure 6).

The pairwise comparisons also revealed non-significant differences in social distance between the intervention and control group for all of the implicit/explicit attitude combinations. However, looking at trends among the variables (Figure 6), it appears as though the intervention was least effective for the low stigmatizers (low implicit, low explicit) as the amount of social distance for those who received the intervention was surprisingly higher compared to the control condition. Also contrary to what was expected, trends suggested that the intervention appears to be most effective for the intentional stigmatizers (low implicit, high explicit) who had the greatest difference in social distance between those who received the intervention compared to the control, followed by the aversive stigmatizers, and lastly the high stigmatizers. Comparisons were also made between the implicit/explicit attitude combinations within the intervention condition to determine if there were significant differences in social distance between the groups for those who received the intervention only. Although there were again no significant differences found, as can be seen from Table 11 and Figure 6, high stigmatizers appear to have the highest levels of social distance, followed by aversive stigmatizers, low stigmatizers, and intentional stigmatizers.

Table 11

Effect of Intervention versus Control on Level of Social Distance for Combinations of Implicit x Explicit Attitudes

Implicit x Explicit Attitudes	Intervention	Control	Mean Difference	p
Low implicit, low explicit	2.906	2.622	.284	.279
Low implicit, high explicit	2.787	3.028	-.241	.356
High implicit, low explicit	2.911	3.006	-.094	.714
High explicit, high implicit	2.987	3.032	-.045	.871

Note. Values are based on estimated marginal means of social distance controlling for social desirability and social distance at time 1; Intervention = combined Conditions 1 & 2, Control = combined conditions 3 & 4.

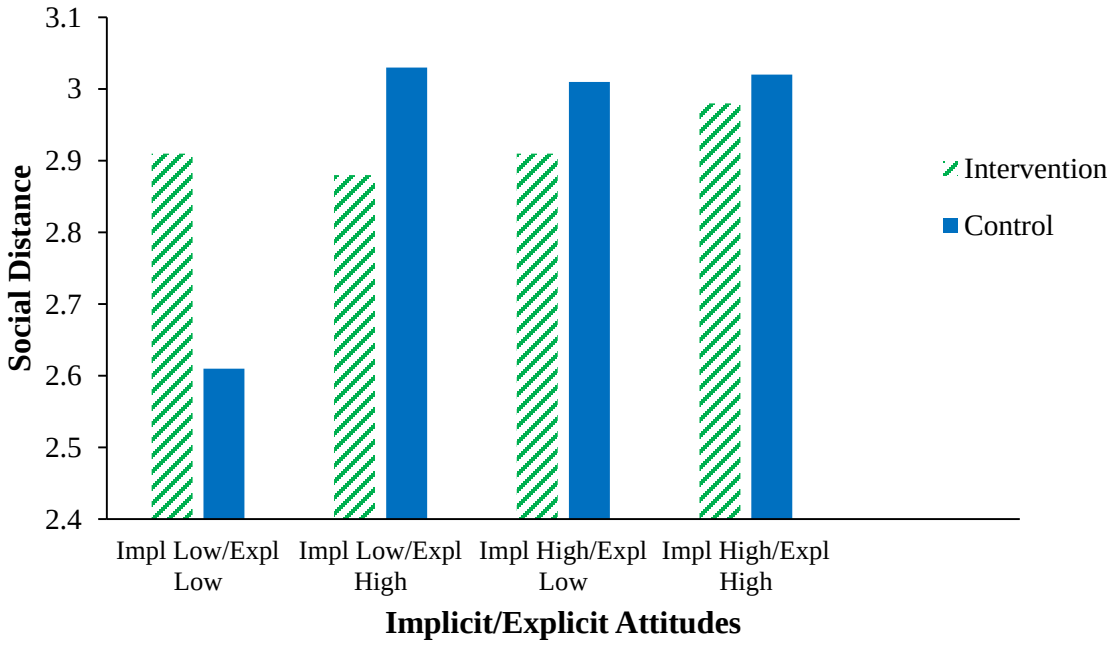

Figure 6. Effect of intervention versus control condition on social distance for each implicit/explicit attitude combination for Study 2 (controlling for social distance at time 1 and social desirability).

Helping behaviour DV. Results of the 2-way ANCOVA for the helping behaviour DV illustrated a non-significant interaction between implicit/explicit attitudes and intervention/control, F (3,169) = 1.083, p = .358, η^2=.019. Although the 2-way interaction was not significant, as was done with the social distance DV, pairwise comparisons were still conducted to examine specific differences between the (implicit/explicit attitude) groups in terms of their level of helping behaviour after receiving the intervention compared to the control condition (please see Table 12 and Figure 7). Consistent with what was predicted, the pairwise comparisons revealed a significant difference in helping behaviour between the intervention and control group for aversive stigmatizers (high implicit, low explicit) only (mean difference =2.11, p =.004), such that they had significantly more helping behaviour after receiving the intervention compared to the control.

Significant differences were also found in helping behaviour between the implicit/explicit attitude combinations for those who received the intervention only (Table 12). Specifically, after receiving the intervention, aversive stigmatizers (high implicit, low explicit) had significantly more helping behaviour compared to intentional stigmatizers (low implicit, high explicit; mean difference = 1.55, p =.046) and (marginally significantly more than) high stigmatizers (high implicit, high explicit; mean difference = 1.45, p =.053). Although no other significant differences were found, as can be seen from Table 12 and Figure 7, the pattern of helping behaviour after receiving the intervention appears to be highest for aversive stigmatziers, followed by low stigmatizers, high stigmatizers, and intentional stigmatizers.

Table 12

Effect of Intervention versus Control on Level of Helping Behaviour for Combinations of Implicit x Explicit Attitudes

Implicit x Explicit Attitudes	Intervention	Control	Mean Difference	p
Low implicit, low explicit	2.637	2.063	.574	.443
Low implicit, high explicit	2.483[a*]	1.347	1.136	.131
High implicit, low explicit	4.031[ab]	1.926	2.106**	.004
High explicit, high implicit	2.584[b*]	2.217	.367	.641

Note. Values are based on estimated marginal means of helping behaviour controlling for social desirability and helping behaviour at time 1; Intervention = combined Conditions 1 & 2, Control = combined conditions 3 & 4. [a, b] represent implicit x explicit attitude combinations that significantly differ from one another in terms of helping behaviour after receiving the intervention. Values with the same letter indicate the specific attitude combinations that significantly differ from each other; **p<.01,* p<.05.

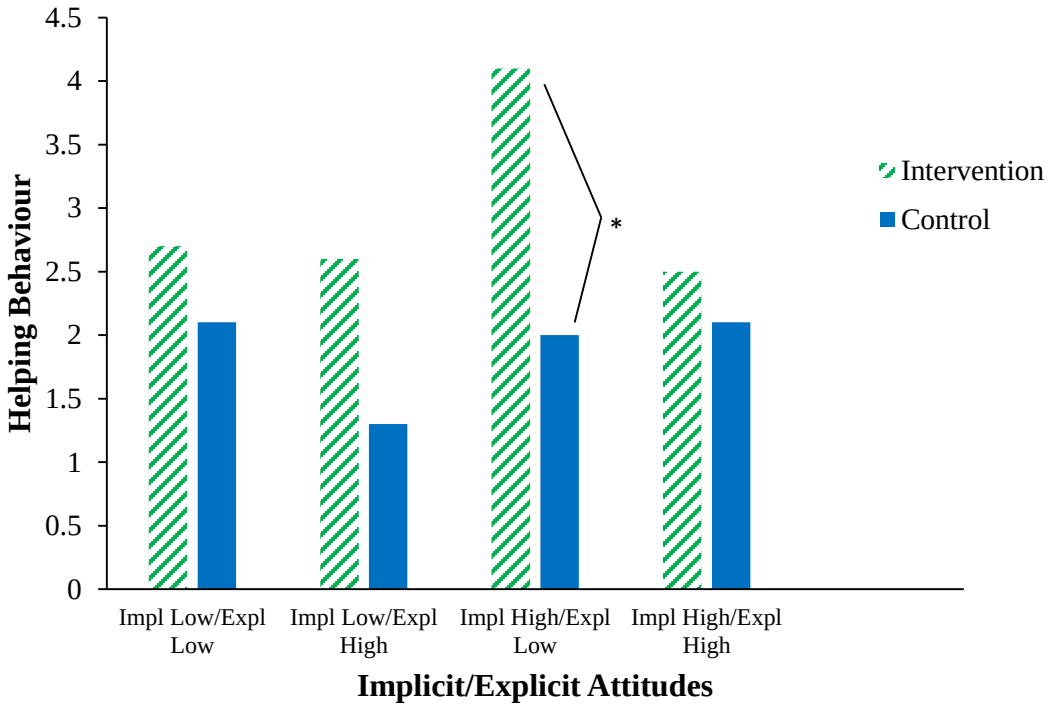

Figure 7. Effect of intervention versus control condition on helping behaviour for each implicit/explicit attitude combination for Study 2 (controlling for helping behaviour at time 1 and social desirability).
Note. Helping behaviour is represented by a difference score between donations made to a mental illness versus a physical illness charity such that higher scores indicate more donations to the mental illness charity; *p<.05

Study 2 Discussion

The results of Study 2 demonstrated an effective intervention in reducing negative implicit attitudes toward mental illness. Automatic associations between mental illness and negative, dangerous, and helpless were all significantly reduced after receiving the combined intervention components of Education, Bias Feedback, and Contact compared to the control condition. Although it was predicted that Condition 2 (Education, Bias Feedback, and Contact) would be most effective for all of the implicit attitudes, Condition 1 (Education and Bias Feedback) appeared to be most effective in reducing automatic associations between mental illness and negative and helpless while Condition 2 appeared to be most effective in reducing automatic associations between mental illness and dangerous. Furthermore, Condition 3 (Contact only) was found to have little effect on reducing automatic negative attitudes overall. Thus, although contact with mental illness has been shown to be a highly effective strategy in reducing explicit stigmatizing attitudes, it did not appear to be similarly effective in reducing implicit stigmatizing attitudes in this study. This finding has been shown in other implicit stigma research (e.g., Kopera et al., 2015) and may be due to the fact that implicit and explicit attitudes are believed to reflect independent processes and thus require different methods in order to produce change. Therefore, education about automatic bias, including its development and consequences to individuals with mental illness as well as increasing people's awareness of their own automatic biases appear to be particularly important factors in the bias reducing process. This is consistent with previous research that has identified the importance of education and personal bias awareness in reducing negative automatic bias (e.g., Devine et al., 2012; Dovidio et al., 2000; Monteith, 1993; Son Hing et al., 2002).

In addition, these results provide support for the application of a contemporary prejudice framework to mental illness stigma regarding the behaviour of aversive stigmatizers. Specifically, although there were no significant effects shown for social distance, the intervention was found to have a significant impact on actual behaviour change (helping) for aversive stigmatizers. Consistent with what was predicted based on the behaviour of aversive racists from the contemporary prejudice literature, aversive stigmatizers (high implicit, low explicit) showed the most increase in their willingness to help individuals with mental illness (by donating to a mental illness rather than physical illness charity) after receiving the intervention compared to the control condition and compared to any of the other implicit/explicit attitude combinations. Although not directly measured in this study, based on the aversive racism literature, this effect is believed to be due to aversive stigmatizers' motivation to reduce their negative bias after being made aware of it given that it is inconsistent with their personal standards and beliefs about themselves. Therefore, the intervention likely created some form of internal dissonance for the aversive stigmatizers who were more motivated to reduce their negative bias toward mental illness and therefore became more willing to help individuals with mental illness, which is more consistent with their personal standards. This possibility should be explored in future research. Lastly, although not traditionally measured in the contemporary prejudice research, the paradoxical implicit/explicit attitude combination described in this study as "intentional stigmatizers" (low implicit, high explicit) produced some interesting patterns of results (though not significant) as this particular group appeared to receive the most benefit from the intervention compared to the other implicit/explicit attitude combinations in terms of reducing social distance from individuals with mental illness. Thus, it may be of interest to continue to examine these intentional stigmatizers who may be intentionally choosing to act in

discordance of their values, but may also be more willing to shift their behaviour and act in non-stigmatizing ways that match their implicit attitudes when given the opportunity.

Taken together, the results of Study 2 demonstrated the effectiveness of an intervention designed to reduce negative implicit attitudes toward mental illness and highlighted aspects of the intervention that are important in the bias reducing process (education and bias awareness). The results also identified that the intervention may be particularly effective for the group of aversive stigmatizers at least when it comes to changing prosocial behaviour, which may be more amenable to change than fear/avoidance based responses related to social distance. These results are encouraging in that they provide evidence that it is possible to change automatic stigmatizing beliefs and behaviours in a relatively short time period that can ultimately improve the lives of individuals with mental illness and also benefit the aversive stigmatizers by helping them choose to act in ways that are based on their values and personal standards.

General Discussion

Research Summary

The overall goal of this program of research was to obtain an improved understanding of the complex nature of implicit mental illness stigma and determine an effective intervention to reduce it. The results of Study 1 confirmed previous research on implicit mental illness stigma by illustrating that people tend to have negative implicit attitudes toward individuals with mental illness (e.g., Monteith & Pettit, 2011; Peris et al., 2008; Teachman et al., 2006) and automatically associate mental illness with dangerous and helpless attributes. Consistent with a contemporary prejudice framework (e.g., Gaertner & Dovidio, 1986; Kovel, 1970; Pearson, et al., 2009), Study 1 also provided support for the existence of a contemporary form of stigma that tends to be expressed subtly, indirectly and without complete awareness of the prejudiced individual. These

individuals, known as aversive stigmatizers, report explicit non-stigmatizing attitudes toward mental illness, yet harbour negative automatic attitudes that can result in subtle, but harmful forms of discrimination. In this study, such discrimination was shown through various degrees of self-reported social distancing from someone with a mental illness and a reduced willingness to help individuals with mental illness by choosing to donate to a physical illness versus mental illness charity.

In the contemporary racial prejudice literature, the subtle behaviour of aversive racists has been shown in a variety of settings, such as helping in emergency situations, selection decisions in employment and college admission, interpersonal judgments, and policy and legal decisions (see Dovidio & Gaertner, 2004). In these cases, aversive racists tend to only discriminate in situations when their behaviour can be personally and explicitly justified so that it does not appear prejudice (Gaertner & Dovidio, 1986). It is likely that these same findings apply to implicit mental illness stigma regarding the behaviour of aversive stigmatizers who in Study 1 tended to show less discrimination than high stigmatizers, but more than low stigmatizers, indicating that they express stigma more subtly when it is able to be justified and remain unnoticed.

Research conducted on health care providers' actual behaviour in relation to their explicit and implicit attitudes toward mental illness can likely be explained by aversive stigmatization. For example, health care providers with positive explicit, but negative implicit attitudes toward mental illness (i.e., aversive stigmatizers) have been shown to over-pathologize patients (Peris et al., 2008) and provide more restrictive treatment (Stull et al., 2012). Given that aversive stigmatizers tend to be unaware of their negative automatic biases and genuinely do not consider themselves stigmatizing, their behaviour is presumed to be unintentional. However, these subtle

forms of discrimination, particularly in health care settings can be especially harmful to individuals with mental illness who already have difficulty seeking help in the first place. Not only may it compromise their actual clinical care, it may also serve to reinforce their own self-stigmatizing beliefs and prevent them from seeking further treatment.

The goal of Study 2 was to develop and test an intervention designed to reduce negative automatic attitudes toward mental illness, which had not yet been addressed in the literature. This was the first study to demonstrate that it is possible to change negative implicit attitudes toward mental illness through a laboratory intervention as well as identify individuals for whom the intervention is most likely to be effective based on the combination of implicit and explicit attitudes. Specifically, although negative implicit attitudes toward mental illness reduced for all participants, it was the most effective for the aversive stigmatizers in terms of changing prosocial behaviour (i.e., donating to mental illness charity) compared to other types of stigmatizers (i.e., high, low, and intentional). However, this same result was not found regarding social distance from mental illness as there were no significant differences in the effectiveness of the intervention between the different types of stigmatizers. This could be because avoidance behaviours are typically formed through automatic associations with fear, which may be more resistant to change due to their evolutionary self-protective nature (e.g., Ohman & Mineka, 2001) compared to choosing to help someone with a mental illness.

The study also confirmed the importance of conscious awareness in negative implicit attitude reduction (e.g., Devine et al., 2012) as education and awareness of one's own biases were shown to be key factors in reducing the three types of negative attitudes targeted in this study (negative, dangerous, helpless). Interestingly, although interpersonal contact with mental illness has shown to be effective in reducing explicit stigmatizing attitudes, it did not, on its own,

show a similar positive effect on negative implicit attitudes in this study. This finding is not entirely surprising given that stigmatizing attitudes persist among mental health professionals (Lauber, et al., 2006; Nordt, Rossler, & Lauber, 2006) despite more frequent contact with individuals with mental illness from working with them on a regular basis (Schulze, 2007; Wahl, Aroesty-Cohen, 2010). The reasoning behind this somewhat paradoxical finding is likely due to the complexities underlying contact as a stigma-reduction strategy. Research has shown that the way in which contact works to reduce stigma is not straightforward and moderated by different factors, such as the quality and type of contact (face-to-face, video, virtual), whether the contact is a deliberate choice (Couture & Penn, 2003; Kolodziej & Johnson, 1996), and the degree to which stereotypes about mental illness are challenged (with moderate disconfirmation being most effective; Corrigan et al., 2012; Reinke et al., 2004). Moreover, given that implicit and explicit stigma are believed to reflect independent processes, it may be the case that contact is effective in reducing explicit stigmatizing attitudes, but is not similarly effective for targeting the automatic, subtle nuances of implicit stigma. Thus, although contact with mental illness is an important factor in reducing stigmatizing beliefs and behaviour, the role it plays in reducing implicit attitudes specifically is not completely clear and should be further examined.

Research Implications

The findings from this research have implications for identifying and changing automatic stigmatizing attitudes at the broader societal level. Given the pervasiveness of mental illness stigma and the difficulty in detecting and changing the automatic, subtle, and often unintentional expressions of stigma, it is important to implement stigma-reduction strategies in a variety of settings that aim to target stigmatizing attitudes at both explicit and implicit levels of processing. Preventing negative automatic attitudes and beliefs about mental illness from being formed in the

first place is one critical point of intervention. This could be accomplished by implementing anti-stigma strategies related to bias education, awareness, and positive role modeling for children and youth in elementary and high schools. This is a particularly important time for intervention as belief systems are in the process of being developed and these strategies can counteract the automatic associative conditioning about mental illness that occurs naturally through socialization. Many interventions have been developed and implemented for school age children and youth that have been beneficial in targeting explicit stigma (e.g., Pinfold, Stuart, Thornicroft, & Arboleda-Flórez, 2005; Pinfold, Toulmin, Thornicroft, Huxley, Farmer, & Graham, 2003; Rickwood, Cavanagh, Curtis, & Sakrouge, 2004), but there have not been any interventions created for implicit stigma.

In addition to prevention strategies, implicit stigma reduction strategies should be implemented in health care settings where negative automatic bias toward individuals with mental illness directly impacts clinical care (e.g., Peris et al., 2008; Stull et al., 2013) as well as college/university campuses and various workplace settings where discrimination can result in exclusion and mistreatment of individuals with mental illness (e.g., Corrigan & Lundin, 2001; Corrigan & Watson, 2002). As with stigma prevention strategies, the stigma reduction interventions that have been implemented to date have been successful at reducing explicit levels of stigma, at least in the short-term (see Dalky, 2012, for a review). The effectiveness of these explicit stigma reduction campaigns can likely be explained by dual-processing theories in psychology (e.g., Epstein, 1994; Smith & DeCoster, 2000; Strack & Deutsch, 2004) that state that the explicit processing system is context-independent and is thus able to change relatively quickly. In contrast, the implicit processing system is highly contextual and is likely to only achieve long-lasting change with considerable time, effort, and experience. Therefore, because

these large scale intervention strategies that are typically implemented at only one point in time must counteract a long history of automatic associative learning, they are unlikely to produce enduring change in the implicit system. Rather, such change is likely to occur over time as individuals proceed through a series of stages related to belief and behaviour change that includes an initial lack of awareness of bias, awareness of bias, the ability to detect bias, consideration of change, deciding to change, initiating strategies to accomplish change, and finally, maintaining the change (Prochaska, DiClemente, & Norcross, 1992).

As this research has confirmed, awareness of one's own bias is a necessary first step in reducing negative automatic attitudes toward mental illness and should thus be incorporated into intervention strategies aimed at reducing implicit stigmatizing attitudes. Given that most people who harbour negative automatic attitudes toward mental illness genuinely do not want to be stigmatizing (aversive stigmatizers), making them aware of their unwanted negative bias in a non-judgmental way can begin the bias-reducing process of allowing them to start to change these beliefs and act in ways consistent with their true values and personal standards. Education about the origins of automatic bias in terms of how it is developed, maintained, and expressed, as well as the consequences for individuals with mental illness can help normalize automatic bias and at the same time elicit guilt and concern that motivates individuals to begin to take action to change their behaviour (Devine et al., 2012).

In terms of the ongoing training and practice involved in maintaining attitude and behaviour change, the strategies involved and how it will be implemented need to be further addressed as different types of intervention strategies will likely be more or less effective depending on the particular individual and setting (Keren, 1990). Devine and colleagues' (2012) intervention to reduce long-term implicit racial bias incorporated a number of different self-

directed strategies as part of their bias-reduction training approach that was effective in reducing implicit racial bias for up to two months. These strategies included stereotype replacement (replacing a stereotypical response with a non-stereotypical response), individuation (evaluating individuals based on personal rather than stereotypical attributes), counter-stereotypic imaging (imagining counter-stereotypic others that challenge the stereotype), taking the perspective of the outgroup member, and increasing opportunities for contact with the outgroup member. While it is not known which of these strategies were more or less effective for each person, the combination of the one-time bias-reduction training as well as continued self-directed bias-reducing strategies was effective in reducing implicit racial bias after two months. Therefore, it is clear that awareness of and education about automatic bias are necessary foundations for the beginning of implicit attitude change and that additional strategies (either self-directed or formal training) are important to create long-lasting effects.

Stigma and Culture

As this program of research has demonstrated, stigma does not operate in the same way for everyone. Not only does the expression of stigma vary as a function of the combination of individuals' explicit and implicit attitudes about mental illness, stigma has also been shown to vary across cultures (Rao, Feinglass, & Corrigan, 2007). Given that individuals' values and personal standards are typically informed by their cultural background and that mental illness diagnoses are made based on deviations from sociocultural norms, it is expected that the way in which mental illness stigma is developed and expressed should differ across cultures (Rao et al., 2007). Despite this understanding, much of the research conducted on mental illness stigma has not examined the role of culture (Abdullah & Brown, 2011). The studies that have examined stigma cross-culturally have indicated that culture is critically important in

understanding the variation in how stigmatizing attitudes, beliefs, and behaviour are developed and maintained (e.g., Anglin, Link, & Phelan, 2006; Cooper-Patrick, Powe, Jenckes, Gonzales, Levine, & Ford, 1997; Rao et al., 2007; Whaley, 1997). These studies have generally concluded that ethnic minorities express more stigmatizing attitudes than European Americans (e.g., Angermeyer, Buyantugs, Kenzine, & Matschinger, 2004; Littlewood, Jadhav, & Ryder, 2007) and are less likely to utilize mental health services (DHHS, 2001). Furthermore, it is argued that although there are racial and ethnic differences in stigmatizing attitudes toward mental illness, it is unlikely that these differences are actually due to racial or ethnic demographics themselves (Sue, 1999). Rather, it is argued that differences in cultural values, beliefs, and norms related to mental illness are what influence stigma (Abdullah & Brown, 2011; Rao, et al., 2007).

As such, Abdullah and Brown (2011) have recommended utilizing a cultural anthropology approach when attempting to understand mental illness stigma, which views psychological constructs as "part of a culturally specific system of beliefs and practices" (Price Shea, Murray, & Hilditch, 1995, p. 10) and involves examining the values, norms, social, political, and economic contexts in which individuals operate (Abdullah & Brown, 2011; Price, et al., 1995; Van Dongen, 2000). From this perspective, cultural norms are believed to be the first step of the stigma process as they initially determine what constitutes a mental illness within a particular culture (e.g., behaviour that deviates from cultural norms and expectations). One's cultural history and values then influence related beliefs that people hold in that culture about mental illness. Cultural history, socialization practices, and culturally-informed attitudes about mental illness then determine whether stereotypes about mental illness are actually endorsed (e.g., people with mental illness are dangerous). Whether or not a person ultimately becomes stigmatizing toward individuals with mental illnesses likely depends on a combination of their

cultural values and their own personal standards regarding the acceptance of discrimination toward individuals with mental illness (Abdullah & Brown, 2011).

Therefore, mental illness and culture are inextricably linked and in order to form a complete and accurate understanding of the stigma associated with mental illness, it is necessary to integrate the important role of culture in its research and examination. This will allow for a deeper understanding of the experience of stigma for all individuals within and across cultures, which can inform interventions that aim to reduce and change stigmatizing attitudes about mental illness.

Limitations and Directions for Future Research

This research has a number of strengths including being the first to provide a more nuanced understanding of the way in which stigma varies as a function of explicit and implicit attitudes as well as developing an intervention to reduce implicit stigmatizing attitudes toward mental illness. However, there are several limitations that should be noted. First, although the intervention was found to be effective in reducing negative automatic attitudes toward mental illness, it is not known whether these effects lasted beyond the time frame that the study was conducted. Given the difficulty involved in producing long-term change in automatic associations that occur at the implicit level of processing, it is important for future research to assess the long-term effectiveness of interventions designed to reduce negative automatic bias toward mental illness to determine if such changes endure over time. Further, this research only examined the effectiveness of three factors (education, awareness of bias, and contact) known to impact mental illness stigma. Whereas education and awareness of bias were found to be key factors in reducing implicit stigma, contact was not found to be similarly effective on its own despite its known negative relation to stigma (e.g., Couture & Penn, 2003; Holmes et al., 1999).

Thus, it would be beneficial for future research to tease apart the particular mechanisms that make contact more or less effective in reducing implicit stigmatizing attitudes as well as additional strategies that may be effective in the stigma reduction process. This point leads to a related limitation of Study 2 that involved combining contact into the control condition to test for differences in discrimination between the different types of stigmatizers after receiving the intervention or the control conditions. This analysis was done post-hoc rather than a-priori given the results of contact being relatively ineffective in reducing negative implicit attitudes and should be noted as a limitation of the study.

Another limitation is that although this research incorporated an actual measure of prosocial behaviour toward mental illness (i.e., donation to charity), the behavioural discrimination variable consisted of a questionnaire measure that asked participants about their behaviour toward mental illness (i.e., social distance) rather than measuring actual behavioural discrimination. Although the social distance scale (Link et al., 1999) has been used extensively in the stigma literature as a proxy measure of behavioural discrimination, it is important for future research to assess how individuals may actually respond to someone with a mental illness, which is likely to be different than how one may think they would respond in a given situation. A related measurement limitation is that the study examined stigmatizing attitudes and behaviours toward mental illness in general rather than assessing stigma in relation to specific types of mental illnesses. Research has shown that certain types of mental disorders, such as those that are more severe (i.e., schizophrenia) tend to be more stigmatized (Day et al., 2007; Link et al. 1999, Phelan et al., 2000) and thus it would be beneficial for future research to assess how implicit stigmatizing attitudes about specific types of mental illnesses are impacted after receiving an intervention.

Lastly, although this research identified the cultural background of participants, it did not examine how cultural factors may influence the expression of implicit stigmatizing attitudes and behaviour or how it plays a role in the stigma reduction process. Moreover, the sample of participants was limited to undergraduate psychology students and thus it is unclear how these results will generalize across cultures and age groups. Although little research has been conducted on culture and implicit mental illness stigma, one study demonstrated that Asian Americans showed stronger negative implicit attitudes toward mental illness and explicitly endorsed greater desire for social distance from mental illness relative to Caucasian Americans (Cheon & Chiao, 2012). Given the strong influence of cultural norms, beliefs, and values on the development and experience of stigma (Abdullah & Brown, 2011), future research should consider culture as an important factor in understanding how stigma operates for individuals of different cultural backgrounds and thus how it impacts stigma change.

Conclusion

The stigma of mental illness continues to be a pervasive social problem despite widespread efforts to reduce it. This research identified a contemporary form of stigma that is indirect and subtle, yet causes stigma to persist and negatively impacts individuals with mental illness in harmful ways. In order to obtain a more comprehensive understanding of stigma, it is necessary to examine the way in which stigma is developed, maintained, and expressed at both explicit and implicit levels of processing. In turn, this will allow for more effective and targeted intervention strategies that aim to reduce all aspects of stigma. This research is promising in that it was the first to demonstrate an effective intervention to reduce implicit stigmatizing attitudes toward mental illness as well as determine for whom the intervention is most likely to be effective. Although these findings can be considered a first step in better understanding the

nature of implicit stigma from a contemporary prejudice framework, it is necessary for future research to continue to examine the various factors that influence the relation between stigmatizing beliefs and behaviour in order to ultimately produce enduring stigma change and improve the lives of individuals affected by mental illness.

References

Abdullah, T. & Brown, T. L. (2011). Mental illness stigma and ethnocultural beliefs, values, and norms: An integrative review. *Clinical Psychology Review, 31*, 934–948.

Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2007). A dynamic model of guilt: Implications for motivation and self-regulation in the context of prejudice. *Psychological Science, 18*, 524–530. doi.org/10.1111/j.1467-9280.2007.01933.x.

Angermeyer, M. C., Buyantugs, L., Kenzine, D. V., & Matschinger, H. (2004). Effects of labelling on public attitudes towards people with schizophrenia: Are there cultural differences? *Acta Psychiatrica Scandinavica, 109*, 420–425.

Anglin, D. M., Link, B. G., & Phelan, J. C. (2006). Racial differences in stigmatizing attitudes toward people with mental illness. *Psychiatric Services, 57*(6), 857–862.

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behaviour. *Journal of Personality and Social Psychology, 83*(2), 380-393. doi.org/10.1037/0022-3514.83.2.380

Bar-Anan, Y. De Houwer, J., & Nosek, B. A. (2010). Evaluative conditioning and conscious knowledge of contingencies: A correlational investigation with large samples. *Quarterly Journal of Experimental Psychology, 63*, 2313–2335.

Bargh, J., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology 62*, 893 – 912.

Breel, K. (2013, May). *Kevin Breel: Confessions of a depressed comic* [Video file]. Retrieved from https://www.ted.com/talks/kevin_breel_confessions_of_a_depressed_comic

Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G. & Tetlock, P. E. (2009). Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *Journal of Applied Psychology, 94*, 567–582.

Chan, J. Y. N., Mak, W. W. S., Lawrence, L. S. C. (2009). Combining education and video-based contact to reduce stigma of mental illness: "The Same or Not the Same" anti-stigma program for secondary schools in Hong Kong. *Social Science and Medicine, 68*(8), 1521-1526.

Carlsson, R. & Agerstrom, J. (2016). A closer look at the discrimination outcomes in the IAT literature. *Scandinavian Journal of Psychology, 57*, 278–287.

Cheon, B. K., & Chiao, J. Y. (2012). Cultural variation in implicit mental illness stigma. *Journal of Cross-Cultural Psychology, 43*(7), 1058-1062.

Cohen, J., & Struening, E. L. (1962). Opinions about mental illness in the personnel of two large mental hospitals. *Journal of Abnormal Social Psychology, 64*, 349-360.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 15-18.

Cooper-Patrick, L., Powe, N. R., Jenckes, M. W., Gonzales, J. J., Levine, D. M., & Ford, D. E. (1997). Identification of patient attitudes and preferences regarding treatment of depression. *Journal of General Internal Medicine, 12*, 431–438.

Corrigan, P. W. (Ed.) (2005). *On the stigma of mental illness: Practical strategies for research and social change*. Washington, DC: American Psychological Association.

Corrigan, P., & Cooper, A. (2005). Mental illness and dangerousness: Fact or misperception, and implications for stigma. In P. W. Corrigan (Ed.), *On the stigma of mental illness:*

Practical strategies for research and social change (pp. 165 – 179). Washington, DC: American Psychological Association Press.

Corrigan, P. W., Edwards, A., Green, A., Diwan, S., & Penn, D. (2001). Prejudice, social distance, and familiarity with mental illness. *Schizophrenia Bulletin, 27*, 219 – 225.

Corrigan, P. W., Kleinlein, P. (2005). The impact of mental illness stigma. In P. W. Corrigan (Ed.), *On the stigma of mental illness: Practical strategies for research and social change* (pp. 343). Washington, DC: American Psychological Association Press.

Corrigan, P.W. & Lundin, R.K. (2001). Don't Call Me Nuts! *Coping with the Stigma of Mental Illness.* Tinley Park, IL: Recovery Press.

Corrigan, P. W., Lurie, B. D., Goldman, H. H., Slopen, N., Medasani, K., & Phelan, S. (2005). How adolescents perceive the stigma of mental illness and alcohol abuse. *Psychiatric Services, 56,* 544-550. doi:10.1176/appi.ps.56.5.544

Corrigan, P. W., Morris, S. B., Michaels, P. J., Rafacz, J. D., & Rüsch, N. (2012). Challenging the public stigma of mental illness: a meta-analysis of outcome studies. *Psychiatric Services, 63*(10), 963-73. doi: 10.1176/appi.ps.201100529.

Corrigan, P. W., & Penn, D. L. (1999). Lessons from social psychology on discrediting psychiatric stigma. *American Psychologist, 54,* 765-776. doi:10.1037/0003-066X.54.9.765

Corrigan, P. W., Rafacz, J., & Rusch, N. (2011). Examining a progressive model of self-stigma and its impact on people with serious mental illness. *Psychiatry Research, 189,* 339-343.

Corrigan, P. W., & Watson, A. C. (2002). The paradox of self-stigma and mental illness. *Clinical Psychology: Science and Practice, 9,* 35-53. doi:10.1093/clipsy/9.1.35

Corrigan, P. W., Watson, A. C., & Barr, L. (2006). The self–stigma of mental illness: Implications for self–esteem and self–efficacy. *Journal of Social and Clinical Psychology, 25*(8), 875-884. doi: 10.1521/jscp.2006.25.8.87

Couture, S. M., & Penn, D. L. (2006.) The effects of prospective naturalistic contact on the stigma of mental illness. *Journal of Community Psychology, 34(5),* 635-645.

Crocker, J. & Major, B. (2003). The self-protective properties of stigma: Evolution of a modern classic. *Psychological Inquiry, 14,* 232-237.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354.

Dalky, H. F. (2012). Mental illness stigma reduction interventions review of intervention trials. *Western Journal of Nursing Research, 34*(4), 520–547.

Day, E. N., Edgren, K., & Eshleman, A. (2007). Measuring stigma toward mental illness. Development and application of the Mental Illness Stigma Scale. *Journal of Applied Social Psychology, 37*(10), 2191–2219.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, *44*(1), 113-126.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5 – 18.

Devine, P. G., & Monteith, M. J. (1993). The role of discrepancy-associated affect in prejudice reduction. *Affect, cognition, and stereotyping: Interactive processes in group perception* (pp. 317–344). San Diego, CA: Academic Press.

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology, 48*, 1267–1278.

Devine, P. G., Monteith, M. J., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology, 60*, 817–830. http://dx.doi.org/10.1037/0022-3514.60.6.817.

Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 36, pp. 1–51). San Diego, CA: Academic Press.

Dovidio, J., Kawakami, K., & Beach, K. (2001). Implicit and explicit attitudes: Examination of the relationship between measures of intergroup bias. In R. Brown & S. L. Gaertner (Eds.), *Blackwell handbook of social psychology* (Vol. 4: Intergroup relations, pp. 175–197). Oxford, UK: Blackwell.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2000). Reducing contemporary prejudice: Combating explicit and implicit bias at the individual and intergroup level. In S. Oskamp (Ed.), *Reducing Prejudice and Discrimination* (pp. 137–163). Hillsdale, NJ: Erlbaum.

Dovidio, J., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). The nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33*, 510–540.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*, 709–724.

Faulkner, G., Irving, H., Paglia-Boak, A., & Adlaf, E. (2010). Adolescent knowledge of schizophrenia and social distancing: A province-wide survey. *Journal of Community Psychology*, *38*, 933-942. doi:10.1002/jcop.20406

Festinger, L. (1957). A theory of cognitive dissonance. Evanston, 111.: Row, Peterson.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Abnormal Social Psychology, 58,* 203-210.

Fiske, S. (1998). Stereotypes, prejudice, and discrimination. In D. Gilbert, S. Fiske, &. G. Lindzey (Eds.), *Handbook of social psychology* (4th ed, pp. 357-411). Oxford, England: Oxford University Press.

Gaertner, S., & Dovidio, J. (1986). The aversive form of racism. In J. Dovidio, & S. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61 – 89). Orlando, FL: Academic Press.

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. http://dx.doi.org/10.1037/0022-3514.78.4.708.

Graves, R., Cassisi, J., & Penn, D. (2005). Psychophysiological evaluation of stigma toward schizophrenia. *Schizophrenia Research, 76*, 317 – 327.

Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., & Banaji, M. R. (2007). Implicit bias among physicians and its predictions of thrombolysis decisions for Black and White patients. *Journal of General Internal Medicine*, 22, 1231–1238.

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

Greenwald, A., & Farnham, S. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*, 1022 – 1038.

Greenwald, A., McGhee, D., & Schwartz, J. (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. & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Metaanalysis of predictive validity. *Journal of Personality and Social Psychology, 97*, 17-41.

Guimon, J., Fischer, W., & Sartorius, N. (Eds.) (1999). *The image of madness: The public facing mental illness and psychiatric treatment*. Basel: Karger.

Hanisch, S. E., Twomey, C. D., Szeto, A. C. H., Birner, U. W., Nowak,, D., & Sabariego, C. (2016). The effectiveness of interventions targeting the stigma of mental illness at the workplace: A systematic review. *BMC Psychiatry, 16*(1).

Hartman, L.I., Michel, N.M., Winter, A., Young, R.E., Flett, G.L., & Goldberg, J.O. (2013). Self-stigma of mental illness in high school youth. *Canadian Journal of School Psychology 28*(1), 28–42. DOI: 10.1177/0829573512468846

Heatherton, T., Kleck, R., Hebl, M., & Hull, J. (2000). The social psychology of stigma. New York, NY: Guilford Press.

Henderson, C., Evans-Lacko, S., Flach, C., & Thornicroft, G. (2012). Responses to mental health stigma questions: The importance of social desirability and data collection method. *Canadian Journal of Psychiatry, 57(3)*, 152-60.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*, 1369–1385.

Holmes, E. P., Corrigan, P. W., Williams, P., Canar, J. & Kubiak, M. A. (1999). Changing attitudes about schizophrenia. *Schizophrenia Bulletin 25*(3*)*, 447–456.

Hinshaw, S. (2005). The stigmatization of mental illness in children and parents: Developmental issues, family concerns, and research needs. *Journal of Child Psychology and Psychiatry, 46*, 714 – 734.

Hinshaw, S.P. (2007). *The Mark of Shame: Stigma of Mental Illness and an Agenda for Change*. New York: Oxford Univ. Press

Hinshaw, S., & Cicchetti, D. (2000). Stigma and mental disorder: Conceptions of illness, public attitudes, personal disclosure, and social policy. *Development and Psychopathology, 12*, 555 –598.

Jorm, A., Christensen, H., & Griffiths, K. (2006). The public's ability to recognize mental disorders and their beliefs about treatment: Changes in Australia over 8 years. *Australian and New Zealand Journal of Psychiatry, 40*, 36 – 41.

Keren, G. (1990). Cognitive aids and debiasing methods: can cognitive pills cure cognitive ills? *Advances in Psychology, 68*, 523–52.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81*, 774–788.

Kawakami, K. Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder: Improving implicit racial attitudes and interracial interactions through approach behaviours. *Journal of Personality and Social Psychology, 92*, 957–971.

Kaufman, G.F., & Libby, L.K. (2012). Changing beliefs and behaviour through experience taking. *Journal of Personality and Social Psychology, 103*(1), 1-19. doi: 10.1037/a0027525

Kessler, R., Chiu, W., Demler, O., & Walters, E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey replication. *Archives of General Psychiatry, 62*, 617 – 627.

Kolodziej, M. E., & Johnson, B. T. (1996). Interpersonal contact and acceptance of persons with psychiatric disorders: a research synthesis. *Journal of Consulting and Clinical Psychology, 64*(6), 1387-1396.

Kopera, M., Suszek, H., Bonar, E., Myszka, M., Gmaj, B., Ilgen, M., & Marcin, M. (2014). Evaluating explicit and implicit stigma of mental illness in mental health professionals and medical students. *Community Mental Health Journal, 51*, 628–634. DOI 10.1007/s10597-014-9796-6

Kovel, J. (1970). *White Racism: A Psychohistory*. New York, NY: Pantheon.

Lai, C. K., Hoffman, K. M., & Nosek, B. A. (2013). Reducing implicit prejudice. *Social and Personality Psychology Compass, 7*(5), 315–330. DOI: 10.1111/spc3.12023

Link, B. G. (1987). Understanding labeling effects in the area of mental disorders: an assessment of the effects of expectations of rejection. *American Sociological Review, 52*, 96–112.

Link, B., & Cullen, F. (1983). Reconsidering the social rejection of ex-mental patients: Levels of attitudinal response. *American Journal of Community Psychology*, *11*, 261 – 273.

Link, B. G., Cullen, F. T., Frank, J., & Wozniak, F. (1987). The social rejection of former mental patients: Understanding why labels matter. *American Journal of Sociology*, *92*, 1461-1500.

Littlewood, R., Jadhav, S., & Ryder, A. G. (2007). A cross-national study of the stigmatization of severe psychiatric illness: Historical review, methodological considerations and development of the questionnaire. *Transcultural Psychiatry, 44*(2), 171–202.

Livingston, J. D., Tugwell, A., Korf-Uzan, K., Cianfrone, M., & Coniglio, C. (2013). Evaluation of a campaign to improve awareness and attitudes of young people towards mental health issues. *Social Psychiatry and Psychiatric Epidemiology, 48*(6), 965-973.

Loo, R., & Loewen, P. (2004). Confirmatory factor analyses of scores from full and short versions of the Marlowe-Crowne Social Desirability Scale. *Journal of Applied Social Psychology, 34*(11), 2343-2352.

Loo, R., & Thorpe, K. (2000). Confirmatory factor analyses of the full and short versions of the Marlowe–Crowne Social Desirability Scale. *The Journal of Social Psychology, 140(5),* 628-635.

Marjanovic, Z., Struthers, C. W., Cribbie, R.,start & Greenglass, E. R. (2014). The Conscientious Responders Scale: A new tool for discriminating between conscientious and random responders. *SAGE Open, 4*, 1-10.

Markowitz, F. (1998). The effects of stigma on the psychological well-being and life satisfaction of persons with mental illness. *Journal of Health and Social Behaviour, 39*, 335 – 347.

Masuda, A., Anderson, P. L., Twohig, M. P., Feinstein, A. B., Chou, Y., Wendell, J. W., & Stormo, A. R. (2009). Help-seeking experiences and attitudes among African American, Asian American, and European American college students. *International Journal for the Advancement of Counselling, 31(3),* 168-180.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data.* Mahwah, NJ: Lawrence Erlbaum.

McGuire, W. J. (1985). Attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., Vol. 2, pp. 233-346). New York: Random House.

Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology, 65*, 469–485. http://dx.doi.org/10.1037/0022-3514.65.3.469.

Monteith, M. J., & Voils, C. I. (1998). Proneness to prejudiced responses: Toward understanding the authenticity of self-reported discrepancies. *Journal of Personality and Social Psychology, 75*, 901–916. http://dx.doi.org/10.1037/0022-3514.75.4.901.

Monteith, L. L., & Pettit, J. W. (2011). Implicit and explicit stigmatizing attitudes and stereotypes about depression. *Journal of Social and Clinical Psychology, 30*, 484–505.

Moses, T. (2010). Adolescent mental health consumers' self-stigma: Associations with parents' and adolescents' illness perceptions and parental stigma. *Journal of Community Psychology, 38*, 781-798. doi:0.1002/jcop.20395

Nail, P. R., Harton, H. C., & Decker, B. P. (2003). Political orientation and modern versus aversive racism: Tests of Dovidio and Gaertner's (1998) integrated model. *Journal of Personality and Social Psychology 84*, 754–770.

Nordt, C., Rossler, W., & Lauber, C. (2006). Attitudes of mental health professionals toward people with schizophrenia and major depression. *Schizophrenia Bulletin, 32*(4), 709–714.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19*, 625-666.

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*(3), 483-522.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically-activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin, 32*, 421-433.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J. & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology, 105*, 171-192.

Ottati, V., Bodenhausen, G. V., & Newman, L. S. (2005). Social Psychological Models of Mental Illness Stigma. In P. W. Corrigan (Ed.), *On the Stigma of Mental Illness: Practical Strategies for Research and Social Change.* Washington DC: American Psychological Association.

Page, S. (1995). Effects of the mental illness label in 1993: Acceptance and rejection in the community. *Journal of Health and Social Policy, 7*, 61 – 68.

Pearson, A. R., Dovidio, J. F., & Gaertner, S. L. (2009). The nature of contemporary prejudice: Insights from aversive racism. *Social and Personality Psychology Compass, 3*, doi: 10.1111/j.1751-9004.2009.00183.x

Penn, D. L., Chamberlin, C., & Mueser, K. T. (2003). Effects of a documentary film about schizophrenia on psychiatric stigma. *Schizophrenia Bulletin, 29(2),* 383-391.

Penn, D. L., Guynan, K., Daily, T., Spaulding, W. D., Garbin, C. P., & Sullivan, M. (1994). Dispelling the stigma of schizophrenia: What sort of information is best? *Schizophrenia Bulletin*, *20*(3), 567-578.

Penn, D. L., Kommana, S., Mansfield, M., & Link, B. G. (1999). Dispelling the stigma of schizophrenia: II. The impact of information on dangerousness. *Schizophrenia Bulletin, 25(3), 437-46.*

Peris, T. S., Teachman, B. A., & Nosek, B. A. (2008). Implicit and explicit stigma of mental illness: Links to clinical care. *Journal of Nervous and Mental Disease, 196*, 752–760.

Phelan, J. C., Link, B. G., Stueve, A., & Pescosolido, B. A. (2000). Public conceptions of mental illness in 1950 and 1996: what is mental illness and is it to be feared? *Journal of Health and Social Behaviour, 41*(2), 188-207.

Pinfold, V., Stuart, H., Thornicroft, G., & Arboleda-Flórez, J. (2005). Working with young people: The impact of mental health awareness programmes in schools in the UK and Canada. *World Psychiatry, 4*(S1), 48-52.

Pinfold, V., Toulmin, H., Thornicroft, G., Huxley, P., Farmer, P., & Graham, T. (2003). Reducing psychiatric stigma and discrimination: Evaluation of educational interventions in UK secondary schools. *British Journal of Psychiatry, 182,* 342-346. doi:10.1192/bjp.182.4.342

Plant, E. A., & Devine, P. G. (2009). The active control of prejudice: Unpacking the intentions guiding control efforts. *Journal of Personality and Social Psychology, 96,* 640–652, http://dx.doi.org/10.1037/a0012960.

Pescosolido, B. A., Martin, J. K., Long, J. S., Medina, T. R., Phelan, J. C., & Link, B. G. (2010). "A disease like any other?" A decade of change in public reactions to schizophrenia, depression, and alcohol dependence. *American Journal of Psychiatry, 167,* 1321-1330.

Price, R. K., Shea, B. M., Murray, K. S., & Hilditch, D. J. (1995). Culture, stratification, and social psychiatry. In R. K. Price, B. M. Shea, & H. N. Mookherjee (Eds.), *Social psychiatry across cultures.* New York: Plenum Press.

Prochaska, J. O., DiClemente, C. C., & Norcross, J. C., (1992). In search of how people change. *American Psychologist, 47*(9), 1102–1114.

Rao, D., Feinglass, J., & Corrigan, P. (2007). Racial and ethnic disparities in mental illness stigma. *The Journal of Nervous and Mental Disease, 195*(12), 1020–1023.

Reinke, R. R., Corrigan, P. W., Leonhard, C., Lundin, R. K., & Kubiak, M. A. (2004). Examining two aspects of contact on the stigma of mental illness. *Journal of Social and Clinical Psychology, 23*(3), 377-389. doi.org/10.1521/jscp.23.3.377.35457

Rickwood, D., Cavanagh, S., Curtis, L., & Sakrouge, R. (2004). Educating young people about mental health and mental illness: Evaluating a school-based programme. *International Journal of Mental Health Promotion, 6*(4), 23-32.

Saucier, D. A., Miller, C. T., & Doucet, N. (2005). Differences in helping whites and blacks: A meta-analysis. *Personality and Social Psychology Review, 9*(1), 2-16.

Schulze, B. (2007). Stigma and mental health professionals: a review of the evidence on an intricate relationship. *International Review of Psychiatry, 19(2),* 137–155.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*, 108–131.

Son Hing, L. S., Li, W., & Zanna, M. P. (2002). Inducing hypocrisy to reduce prejudicial responses among aversive racists. *Journal of Experimental Social Psychology, 38,* 71–78.

Stier, A. & Hinshaw, S.P. (2007). Explicit and implicit stigma against individuals with mental illness. *Australian Psychologist, 42*(2), 106-117. DOI: 10.1080/00050060701280599

Stone, A. M., & Merlo, L. J. (2011). Attitudes of college students toward mental illness stigma and the misuse of psychiatric medications. *Journal of Clinical Psychiatry, 72(2),* 134-139.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behaviour. *Personality and Social Psychology Review, 8*, 220–247.

Stull, L. G., McGrew, J. H., Salyers, M. P., & Ashburn-Nardo, L. (2013). Implicit and explicit stigma of mental illness: Attitudes in an evidence-based practice. *The Journal of Nervous and Mental Disease*, *201*(12), 1072-1079. DOI: 10.1097/NMD.0000000000000056

Sue, S. (1999). Science, ethnicity, and bias: Where have we gone wrong? *American Psychologist, 54*, 1070–1077.

Taylor, S. M., & Dear, M. (1981). Scaling community attitudes toward the mentally ill. *Schizophrenia Bulletin, 7*, 225-240.

Teachman, B., Wilson, J., & Komarovskaya, I. (2006). Implicit and explicit stigma of mental illness in diagnosed and healthy samples. *Journal of Social and Clinical Psychology, 25*, 75 – 95.

Todd, A. R., Bodenhausen, G. V., Richeson, J. A., Galinksy, A. D. (2011). Perspective taking combats automatic expressions of racial bias. *Journal of Personality and Social Psychology, 100*(6), 1027-1042. DOI: 10.1037/a0022308

Tsang, H., Tam, P., Chan, F., & Cheung, W. (2003). Stigmatizing attitudes towards individuals with mental illness in Hong Kong: Implications for their recovery. *Journal of Community Psychology, 31*, 338 – 396.

U.S. Department of Health and Human Services (2001). *Mental health: Culture, race, and ethnicity—A supplement to Mental health: A report of the Surgeon General.* Rockville, MD: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services.

Van Dongen, E. (2000). Anthropology and psychiatry: Two of a kind but where is the other? In V. Skultans & J. Cox (Eds.), *Anthropological approaches to psychological medicine: Crossing bridges.* Philadelphia: Jessica Kingsley Publishers.

Wahl, O. (1999). Mental health consumers' experience of stigma. *Schizophrenia Bulletin, 25,* 467 – 478.

Weiss, M.F. (1994). Children's attitudes toward the mentally ill: A developmental analysis. *Psychological Reports,* 58, 11-20. DOI: 10.2466/pr0.1986.58.1.11

Weiss, M. F. (1994). Children's attitudes toward the mentally ill: An eight-year longitudinal follow-up. *Psychological Reports, 74*, 51-56. DOI: 10.2466/pr0.1994.74.1.51

Wahl, O., & Aroesty-Cohen, E. (2010). Attitudes of mental health professionals about mental illness: a review of the recent literature. *Journal of Community Psychology, 38*(1), 49–62.

Whaley, A. L. (1997). Ethnic and racial differences in perceptions of dangerousness of persons with mental illness. *Psychiatric Services, 48*(10), 1328–1330.

World Health Organization. (2001a). *Mental health: New understanding, new hope*. New York: Author.

World Health Organization. (2001b). *Stigma and global health: An international perspective*. New York: Author.

Appendices

Appendix A: GNAT Stimuli

Mental illness was the target category and consisted of the following stimuli: schizophrenia, bipolar disorder, depression, anxiety, and obsessive compulsive disorder (OCD). These particular stimuli were chosen given that they tend to be commonly recognized as mental illness. The attributes that participants were asked to associate with mental illness included negative/positive, dangerous/harmless and helpless/competent. The negative/positive attribute was chosen to assess participants' overall automatic attitudes toward mental illness and the dangerous/harmless and helpless/competent attributes were chosen based on stereotypical beliefs typically associated with mental illness. The stimuli for the negative/positive attribute consisted of the words "Awful", "Terrible", "Horrible", "Nasty", and "Evil" contrasted with "Joy", "Love", "Pleasant", "Wonderful", and "Pleasure." The stimuli for the dangerous/harmless attribute consisted of the words "Dangerous", "Unsafe", "Violent", "Aggressive", and "Threatening" contrasted with "Harmless", "Safe", "Peaceful", "Gentle", and "Soft." The stimuli for the helpless/competent attribute consisted of the words "Incompetent", "Helpless", "Incapable", "Unable", and "Unskilled" contrasted with "Capable", "Qualified", "Competent", "Able", and "Skilled."

Appendix B: Education and Bias Feedback Intervention (adapted from Devine et al., 2012)

Background: Automatic Bias Toward Mental Illness

Sometimes in everyday life, people experience spontaneous thoughts, feelings, and behaviours that are different from what they desire. Like bad habits, these spontaneous reactions can be extremely difficult to control because they occur automatically, before a person even notices them or has time to reflect on them.

These spontaneous reactions vary from person to person, depending on the specifics of where one grew up. However, because people are exposed to many of the same environmental factors, such as the same movies and television shows, many people experience the same unwanted reactions.

Imagine that it's past midnight on a Friday night. Pete and Joe are walking down the street on their way home from a friend's house. The light is poor, and it's late enough that they only meet an occasional person on the darkened sidewalk.

Pete and Joe see a person in the distance walking towards them. As the person gets closer, Pete sees through the dim light that the person is a middle-aged man wearing a baseball cap, over-sized coat, and carrying several bags. Pete also hears him talking to himself in a loud voice, but can't make out the words.

Pete thinks, "This guy must be mentally ill. Maybe I should cross the street in case he might be dangerous. Whatever, Joe's with me, and we can protect ourselves if we have to."

They keep walking and, as the man approaches, Pete starts to feel a little tense. He shifts his position so he will be further away from the man when they pass by him and avoids eye contact.

As the man passes, he says, "good evening" to Pete and Joe and continues walking down the street. Pete thinks, "That was weird. He must not have wanted to hurt us after all."

As Pete and Joe walk away, Joe says to Pete, "Did you see that guy's headphones? Those were the ones that I saw in the store the other day, but were too expensive to buy."

Pete feels confused and wondered why Joe wasn't surprised that the man didn't try to hurt them. He then realizes that the man must have been listening to music and was probably reciting the words out loud.

Pete thinks, "Wow, I guess I just assumed he was mentally ill and that he was dangerous. Why did I do that? Now that I think about it, he did seem pretty harmless."

In this example, Pete was quick to expect that the man had a mental illness and because of this, that he might be dangerous. His initial expectation led to spontaneous, inaccurate thoughts and unwarranted tense feelings. Why did Pete jump to the conclusion that the man was mentally ill and dangerous?

The Origins of Automatic Bias Toward Mental Illness

Stereotypes that people with mental illness are crazy and dangerous likely influenced Pete's initial expectations about the man. Stereotypes are everywhere in our society, so we all learn about them whether we want to or not. Television shows and movies often portray individuals with mental illness as violent, delusional, or incompetent with little hope of recovery. This is especially true with more serious mental illnesses, such as schizophrenia.

Because stereotypes are all around us, we can't help but learn them and come to associate individuals with mental illness with these negative stereotypes.

The news media is also known for depicting mental illness in a negative light where mental illness is frequently associated with violent crime sometimes even before it is has been confirmed whether the accused did in fact have a mental illness. For example, consider these recent news headlines from the Boston Herald:

1) "A mentally ill knife-wielding man killed an elderly woman and a teacher in a rolling rampage in Taunton." (May 2016)
2) "Theater gunman's family called him mentally ill, violent." (July 2015)

The prevalence of stereotypic representations of individuals with mental illness is likely greater than you imagine, and can sometimes be quite subtle. For example, another news article about introducing stricter gun laws in the United States reported, "The three-part plan, which will be unveiled Monday, would make it harder for violent criminals and the mentally ill to obtain guns by focusing on background checks and mental health funding."

Notice how the report indirectly criminalizes mental illness by putting it in the same category as violent criminals who should not be allowed to own guns.

Can you think of specific times when you noticed the media portraying individuals with mental illness stereotypically? Please briefly describe any instances that come to mind.

It is hard to avoid negative reactions to individuals with mental illness when we are so frequently exposed to stereotypes in everyday life. We see them so often that they become firmly ingrained in our minds. Without intending or realizing it, we learn to associate individuals with mental illness with negative stereotypes.

As a result, when people think about or interact with individuals with mental illness, the negative stereotypes spring to mind, even among people who disagree with the stereotypes. Once in mind, these stereotypes can influence people's thoughts feelings, expectations, and behaviour toward individuals with mental illness.

In this way, stereotypes are like bad habits in that they can occur without thought or intention. Because people often don't realize when stereotypes influence their reactions to individuals with mental illness, avoiding the influence of stereotypes can be very difficult.

Now think back to the situation with Pete and Joe that we described in the beginning of this presentation. Why do you think that Pete automatically assumed that the man was mentally ill and dangerous, despite a lack of clear evidence? Perhaps Pete's judgment was influenced by the stereotype that people with mental illness are delusional (e.g., talking to someone who is not there) and violent.

How would you have reacted in Pete's situation? Is it possible that you, too, would have been quick to think that the man was mentally ill and dangerous?

The example with Pete demonstrates how one's reactions can be negatively biased toward individuals with mental illness without any awareness of the bias. If not for Joe's comments about the man's headphones that made Pete think twice, it is very likely that Pete would not have realized that his assumption that the man was mentally ill and dangerous was inaccurate. These types of automatic associations can lead to expressions of bias that are so subtle that people often fail to detect the bias in their thoughts, feelings, and behaviour.

Can you think of any times in the past where you had an automatic response that was influenced by stereotypes? Please briefly describe any instances that come to mind.

Consequences of Automatic Stereotypes

Employment Decisions

Imagine how the activation of automatic stereotypes might influence an employer's initial evaluation of a job applicant who has a mental illness and subsequent thoughts and feelings toward the applicant. Negative stereotypes (e.g., incompetent, unpredictable) might color the first impression of the applicant, leading to lower evaluations of applicants who have or have had a history of mental illness compared to applicants without a mental illness.

Treatment Decisions

Recent studies have shown that even mental health professionals show automatic bias toward individuals with mental illness, which has consequences for the quality and type of treatment they receive. For example, mental health professionals with a negative automatic bias toward mental illness have a tendency to over-diagnose their patients as well as provide more restrictive types of interventions that do not allow patients much control over their own recovery.

Everyday Interactions

In addition to employment and treatment settings, automatic negative biases toward individuals with mental illness occur frequently across all different types of settings and interactions. For example, university students show biases in the way they interact with individuals with mental illness in everyday situations. These students may show an avoidant interaction style by making less eye contact, sitting further away, showing more nervous behaviours, or cutting interactions short.

An important aspect of all these studies is that people are often unaware that they have acted with bias. In fact, many of the participants in these studies report that they did not want to treat individuals with mental illness differently from those without a mental illness and that they believed that acting with bias is wrong. Yet, despite the best intentions, the biases still occur, which end up contributing to mental illness stigma and negatively affects the lives of individuals with mental illness in various ways.

Recall that you completed an IAT before coming to the lab today. Your score has been calculated and the results are shown on the next page.

The results of your IAT score indicate that at an automatic level, you have a relative negative bias toward mental illness compared to positive on the IAT.

Breaking the Prejudice Habit

In many ways, the research that we just reviewed is discouraging because it suggests that even people who want to treat individuals with mental illness fairly can act in biased ways. This has led some researchers to explore whether it is possible to reduce biases resulting from automatic stereotypes. Here is some good news. If a few conditions are met, it is possible to reduce automatic bias.

Specifically, people can reduce automatic bias if they:

(1) are motivated to overcome the bias
(2) become aware of their bias and why it exists
(3) are able to detect the subtle influence of stereotypes
(4) learn and practice strategies that help reduce automatic bias

Being motivated to reduce prejudice and automatic biases is a necessary first step. Without motivation, people will be unlikely to expend the effort needed to eliminate the effects of automatic biases. Being motivated is a personal decision that people must make for themselves.

Even if people are motivated to reduce their bias, they still need to become aware of it and why it exists. Much of what we have discussed early in the presentation explains why so many people are affected by automatic bias, even when they believe that prejudice is wrong.

Detecting the Influence of Stereotypes

Before we can overcome the negative effects of automatic stereotypes, we must be able to detect stereotypical depictions of mental illness in our environment and when our own responses are affected by these depictions. Detecting these biases creates the opportunity to do something about them.

Because our social environment plays such a large role in perpetuating stereotypes, we must first learn to detect biased portrayals, whether they occur in the media or in interactions with others. While we may not be able to stop how others portray mental illness, we can choose how we react to those portrayals by recognizing when a biased portrayal occurs and expressing disapproval of it. As we have shown, sometimes bias can be quite subtle, so we must be vigilant to detect this bias.

Equally important as the detection of bias in our external environment is the detection of bias within ourselves. This involves figuring out the situations in which we are most likely to be subtly influenced by automatic stereotypes and monitoring our responses in these situations. We must take similar steps to break other kinds of habits, like biting nails; to stop biting nails, we must figure out the situations that trigger nail-biting behavior.

After we have figured out how stereotypes are reinforced by our environment and when stereotypes are likely to pop to mind, we can work to prevent the influence of stereotypes by training ourselves to behave in different, unbiased ways.