The Malleability of Stigmatizing Attitudes: Combining Imagined Social Contact with Implicit Attitude Feedback

Charlotte R. Pennington[a], Claire Campbell[b], Rebecca L. Monk[a] & Derek Heim[a]

[a]Department of Psychology, Edge Hill University, Ormskirk, England

[b]School of Psychology, University of Ulster, Coleraine, England

Abstract

Purpose: Research is reported which examines whether imagined social contact combined with implicit attitude feedback may be an effective intervention for inducing changes in attitudes towards mental ill-health. **Design:** The Implicit Relational Assessment Procedure (IRAP) captured participants' implicit attitudes towards individuals with a mental illness and provided a measure of attitude bias. Forty-eight participants (17 male; 95.8% White British) were randomly assigned to one of four experimental conditions: (1) Imagined social contact with implicit attitude feedback; (2) imagined social contact without feedback; (3) control with feedback; (4) control without feedback. This resulted in a data set detailing 12, 288 implicit responses, with each participant completing 256 IRAP trials respectively. Participants then completed an attitude change assessment 24 hours later. **Results:** Results revealed that imagined social contact was successful in changing implicit attitudes; with the addition of implicit attitude feedback further strengthening participants' positive attitudes towards mental ill-health. Explicit attitudes remained unaffected. **Conclusion:** These findings are the first to highlight the practical importance of combining imagined social contact with implicit attitude feedback in order to improve attitudes towards stigmatized out-groups.

Keywords: Attitude change; explicit attitudes; Implicit Association Test; implicit attitudes; Implicit Relational Assessment Procedure; mental illness; public stigma

Introduction

Individuals with mental health problems not only have to cope with the symptoms of their condition but also face discrimination, stigma, and prejudiced attitudes in their everyday lives (Sacca & Ryan, 2011). Research high- lights that 70% to 80% of the general public hold the belief that individuals with a mental illness are dangerous and unpredictable (Crisp, Gelder, Rix, Meltzer, & Rowland, 2000; Dickerson, Sommerville, & Origoni, 2002; Pescosolido, Monahan, Link, Stueve, & Kikuzawa, 1999). The ubiquity of such beliefs is further evidenced by findings suggesting that by age 14 years, high school students can generate as many of 250 stigmatizing labels related to mental ill health (Rose, Thornicroft, Pinfold, & Kassam, 2007). Underpinned by a social-cognitive model, stigmatising attitudes towards mental ill health may also translate into discrimination as an individual responds to their attitudes and emotions (Corrigan, 2002). Specifically, stigmatising attitudes regarding the perceived dangerousness of an individual with a mental health diagnosis may be particularly problematic, leading to fear and avoidance of persons with mental illness (Corrigan, 2002). As such, public stigma surrounding mental ill-health is a major public health concern (Rüsch, Kanter, Angelone, & Ridley, 2008) which may further affect adversely those who experience mental illness (Corrigan, 2000; Corrigan, Larson, & Rüsch, 2009; Corrigan & Watson, 2002). Interventions aimed at assisting individuals to cope with public stigma (Link, Struening, Neese-Todd, Asmussen, & Phelan, 2002) present one solution to this pressing issue. However, targeting the root of the problem may be found in the development of interventions that change the general publics' attitudes towards mental illness.

One strategy that has garnered empirical support as an effective attitude change intervention stems from Allport's (1954) intergroup contact theory. This theory posits that direct contact between members of opposing groups should lessen hostility and lead to more

positive out-group attitudes (Turner & Crisp, 2010). Indeed, meta-analyses suggest that intergroup contact is fundamentally associated with reduced prejudice, and that the positive effects of contact typically generalize to the entire outgroup (e.g., Pettigrew & Tropp, 2006). However, the opportunity for contact with out-group members is not always feasible, nor does the public necessarily desire such contact (Stathi, Tsantila, & Crisp, 2012), particularly with highly stigmatized groups such as those with diagnosis of mental illness (Angermeyer & Dietrich, 2006). Nevertheless, the limitations of intergroup contact theory are not intractable and the power inherent to the concept of contact provides a theoretical framework for further theories of stigma reduction to be elucidated. Capitalising on this, Crisp and Turner (2009) postulate that imagined intergroup interactions may be a beneficial alternative to direct interactions (Imagined Social Contact, ISC; Crisp & Turner, 2009). Providing empirical evidence to support this claim, Stathi et al. (2012) revealed that by imagining a positive interaction with an individual diagnosed with schizophrenia, participants formed greater intentions to engage with such individuals in the future. Moreover, these feelings were also found to generalize to those with general mental illness. Such findings offer a promising research avenue that might pave the way for the development of interventions to reduce negative mental health-related attitudes. Nevertheless, by examining explicit stigmatizing attitudes in isolation, these studies are potentially vulnerable to self-presentational motives (Rydell & McConnell, 2006). In other words, as individuals are consciously aware of their self-reported explicit attitudes (Cullen, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009; Nisbett & Wilson, 1977; Wegner, 2002), it is conceivable that these may be shaped by social desirability (Schwarz, 1999).

Owing to the deliberate and controllable nature of explicit attitudes, which may be manipulated by situational demands (Davies, 1997; Rydell & McConnell, 2006; Schwarz, 1999), one major revolution in attitude change research has seen the transition from explicit

to implicit attitude measures (Gawronski & LeBel, 2008). The most established implicit attitude measure is the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). The IAT is proposed to uncover introspectively unidentified traces of past experiences that may drive feelings towards social objects (Greenwald & Banaji, 1995) and works on the pivotal assumption that it should be quicker to map two concepts onto a single response option when those two concepts are associated in memory (De Houwer, 2002). Consequently, implicit attitude measures are thought to able to better diagnose socially sensitive attitudes than explicit measures (Greenwald et al., 2002).

More recently, however, researchers have suggested that implicit attitudes may be better understood in behavioral terms, and have postulated that stimulus relations, as opposed to associations *per se*, underpin human language and cognition (Barnes-Holmes et al., 2006; Hughes & Barnes-Holmes, 2011; Lynott, Kansal, Connell, & O'Brien, 2012). One such account stems from the relational elaboration and coherence model (REC), which is underpinned by the Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001). This model posits that self-report and indirect attitude measures mirror the operations of the same behavioral system. However, implicit attitudes reflect brief and immediate relational responding whereas explicit attitudes reflect more controlled and extended relational responding (Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010a). Based on this theoretical framework, the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes et al., 2006) was developed to reflect both the directionality between attitude concepts (i.e., a relation), and the continua along which such attitudes may occur (i.e., a relational network; De Houwer, 2002; Barnes-Holmes et al., 2006). In view of this distinction, the IRAP appears to provide a direct measure of implicit beliefs (Barnes-Holmes et al., 2006).

As well as providing a measure of individuals' automatic attitudinal relations, implicit attitude measures may also hold promise as an accessible and cost-effective intervention for

reducing explicit stigmatized attitudes. For example, by revealing discrepancies between people's explicit and implicit attitudes, mental health related-stigma may be reduced (Devine, Monteith, Zuwerink, & Elliot, 1991; Plant & Devine, 1998; 2009). Testing this assumption, Menatti, Smyth, Teachman and Nosek (2012) examined whether implicit attitude feedback or the experience of completing a Brief IAT (BIAT) could alter individuals explicit attitudes towards mental illness. Results revealed that participation in an implicit attitude measure reduced explicit stigmatising attitudes towards mental ill health. The addition of implicit attitude feedback did not appear to modify this effect. This research suggests that the mere act of participating in an implicit attitude measure may influence individuals to evaluate their level of implicit bias and engage in self-regulating behaviours to reduce ostracizing attitudes. However, to date, no research has utilised the IRAP to provide a direct measure of implicit stigmatizing attitudes towards individuals with a mental illness. Furthermore, it is not known whether implicit attitude feedback or participation in an implicit measurement test can heighten the effectiveness of ISC and reduce negative attitudes towards mental-ill health.

Overview of Research

Building on a long and rich history of studying attitudes (Allport, 1935, 1954), interventions aiming to change stigmatizing attitudes toward out-groups have proliferated in recent years (e.g., Menatti, et al., 2012; Stathi et al., 2012), many of which have focused on the malleability of explicit attitudes (Wilson, Lindsey, & Schooler, 2000). However, individuals' explicit and implicit attitudes can be divergent in nature (Power, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009). The current study therefore suggests that the development of attitude change interventions should target explicit and implicit attitudes. Accordingly, as well as examining participants' explicit responses, the IRAP is employed to provide an implicit attitude change measure. The current research has a threefold remit: First, to build on the notion that ISC can effectively reduce explicit public stigma (Stathi et al., 2012), and to

examine the effectiveness of this approach to alter implicit and explicit attitudes. Second, to test the contention that making individuals consciously aware of their implicit attitudes (Menatti et al., 2012; Power et al., 2009) may be a particularly effective means of bringing about attitude change. Third, to combine ISC (Crisp & Turner, 2009) and implicit attitude feedback (Menatti et al., 2012) to examine if this may reduce implicit and/or explicit stigmatized attitudes toward individuals with mental illness.

Method

Participants

Following ethical approval, opportunity sampling was employed to recruit 61 undergraduate students. Participants took part on a voluntary basis and were retained on the basis of two inclusion criteria: no previous social contact with an individual diagnosed with any mental health problems (such as close relatives or friends), in light of research which suggests that prior contact can increase positive attitudes toward this group (Angermeyer & Matschinger, 1996; Brown, 2012); second, they had no previous exposure to, or knowledge of, the IRAP. Upon completion of IRAP practice trials, data from 13 participants were removed from the sample for failure to respond with at least 80% accuracy and with response latencies below 2,000 milliseconds (ms) (Barnes-Holmes, Murphy, Barnes-Holmes, & Stewart, 2010; Vahey, Barnes- Holmes, Barnes-Holmes, & Stewart, 2009).[1] This yielded a final sample of 48 participants (17 male, *M*age1⁄421 years, *SD*1⁄42.81, 95.8% White British). This sample size is in line with previous IRAP studies (Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, 2008; Barnes-Holmes, Murphy, et al., 2010; Barnes-Holmes, Waldron, Barnes-Holmes, & Stewart, 2009; Power et al., 2009).

Materials and Equipment

Explicit measure. A sub-section of the Community Attitudes toward the mentally ill scale, specifically the 10-item *social restrictiveness scale* (CAMI-SR; Taylor & Dear, 1981, Cronbach's *a = .80*) gauged participants' explicit attitudes towards individuals with a mental illness. This questionnaire was administered to participants directly after the ISC manipulation in order to examine the change in stigmatizing and socially restrictive attitudes towards people with a mental illness. Participants were asked to rate the extent to which individuals with a mental illness are perceived as dangerous (i.e., "The mentally ill are far less dangerous than most people suppose" and "Most women who were once patents in a mental hospital can be trusted as babysitters") and the degree to which they would avoid an individual with a mental illness (i.e., The mentally ill should be isolated from the rest of the community" and "I would not want to live next door to someone who had been mentally ill"). Responses were marked on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). Higher scores on this measure indicated a greater propensity to avoid an individual with a mental illness and a heightened perception that they are dangerous. The internal consistency of the CAMI-SR was satisfactory in both Phase 1, (α = .74) and Phase 2 of the experiment (α = .76).

Implicit measure. In the current study, the IRAP measured participants' implicit attitudes towards individuals with a mental illness. The IRAP was presented on a computer situated in a laboratory setting. It comprised of the sample stimuli 'mentally ill person' and 'physically ill person' and twelve positive and negative evaluative target relations (see Table 1). The sample stimuli of 'mentally ill person' was paired with the comparison of 'physically ill person' due to previous research which found that participants were both implicitly and explicitly biased towards individuals with a mental illness compared to those with a physical illness (Teachman, Wilson, & Kamarovskaya, 2006). Negative evaluative target relations

were selected from a list of words that are commonly associated with mental illness (Rose et al., 2007).

[INSERT TABLE 1 HERE]

To capture the direction of participants' attitudinal relations, two response options of 'similar' and 'opposite' were directly paired with the sample stimuli. IRAP trial types were categorized in line with previous research (Crisp et al., 2000; Dickerson et al., 2002; Menatti et al., 2012; Pescosolido et al., 1999; Rose et al., 2007) and reflected participants' consistent or inconsistent verbal relations (Power et al., 2009). . Accordingly, participants were required to relate the phrases 'mentally ill person' and 'physically ill person' to positive and negative terms that could be readily categorized as safe or dangerous.

To aid participants' understanding of these four IRAP trial types, participants were presented with four visual aids before the test began. These visual aids were constructed from the configurations of the sample stimuli (MENTALLY ILL PERSON and PHYSICALLY ILL PERSON) and the two types of target stimuli (positive or negative). For experimental purposes, the four trial types were referred to as PHYSICALLY ILL PERSON/Harmless (top left), PHYSICALLY ILL PERSON/Dangerous (top right), MENTALLY ILL PERSON/Harmless (bottom left) and MENTALLY ILL PERSON/Dangerous (bottom right; Figure 1).

Participants competed two practice blocks in order to evaluate their accuracy and average response latencies. A maximum limit of four attempts was set for participants to pass the practice blocks and commence on to the critical blocks. During both the practice and critical blocks, participants were instructed as to which sample stimuli should be paired with a specific response. After each block, participants were then instructed that the previously correct and incorrect responses would be reversed in the next block, thus, eliminating any

trial-and-error learning (Barnes-Holmes et al., 2010a). A total of eight blocks were presented to participants. Each block of IRAP trials contained eight exposures to each of the four trial types, presented quasi-randomly across a block. This resulted in a total of 32 trials in each block, with a set restriction that none of the four trial types could be presented twice in succession. The positioning of the two response options were also presented quasi-randomly and could not appear in the same left-right position more than three times in succession. Participants progressed through trials by pressing the correct response option, which then removed all stimuli from the screen for 400ms before the next trial was presented. Choosing an incorrect response resulted in a red cross appearing on the screen, with the trial continuing once the correct option had been selected.

IRAP scoring

The primary data produced by IRAP trials are response times (RTs), defined as milliseconds between the onset of the trial and a correct response. Accuracy was also recorded in every trial, however this data was simply employed as a screening mechanism to ensure that only scores with greater than 80% accuracy in practice and test blocks were included within analyses. D-IRAP scores were computed from the RTs of critical test blocks following the improved scoring algorithm by Greenwald, Nosek, and Banaji (2003): Step (1) Only RTs from test blocks were used; (2) RTs above 10,000ms from test blocks were eliminated; (3) all data for participants were removed if they produced more than 10% of test block trials with RTs less than 300ms; (4) 12 standard deviations were computed across the six test blocks; (5) 24 mean RTs for the four trial types in each block were calculated; (6) the difference score were calculated by subtracting the mean latency score from the consistent trial (mentally ill person-dangerous, physically ill person-harmless) from the mean latency of the inconsistent trial (mentally ill person- harmless, physically ill person-dangerous); (7) each difference

score was divided by its corresponding standard deviation score created in Step 4), which yielded 12 D-IRAP scores, one for each trial type for each pair of test blocks; (8) four overall D-IRAP scores were created by averaging the scores for each trial type across the three pairs of test blocks (Barnes-Holmes, Barnes-Holmes, et al., 2010).

[INSERT FIGURE 1 HERE]

Procedure

Phase 1. Participants were randomly and equally allocated to one of four experimental conditions: (1) ISC with feedback, (2) ISC without feedback, (3) control with feedback, and (4) control without feedback. Participants in the two ISC conditions listened to a digital recording that asked them to imagine a hypothetical contact situation with an individual who was mentally ill. A positive contact situation was chosen as previous research indicates that a neutral ISC can have negative effects on attitudes toward schizophrenia (West, Holmes, & Hewstone, 2011). In line with Crisp and Turner (2009), the following audio transcript was chosen as it disconfirms the common stereotype that individuals with a mental illness are dangerous (Reinke, Corrigan, Leonhard, Lundin, & Kubiak, 2004):

> *"I would like you to take a minute to imagine yourself meeting a mentally ill individual for the first time. Imagine that the encounter is positive, relaxed and comfortable".*

Those that were in the two control conditions did not listen to this digital recording. Participants were then asked to define the terms 'dangerous' and 'harmless' to ensure that they understood the nature of the words used in the IRAP. To examine explicit and implicit attitudes towards individuals with a mental illness, all participants then completed an explicit (i.e., CAMI-SR) and implicit (i.e., IRAP) attitude measure, administered in a counterbalanced design. These measures were administered after the ISC manipulation in order to examine its

effectiveness as an explicit and implicit attitude change intervention. This also allowed the researchers to further test whether the act of completing an implicit attitude measure may have had on attitude change. Participants in the conditions of ISC with implicit feedback and the control condition with implicit feedback received personalized implicit attitude feedback which was calculated by observing participants mean D-IRAP score for the trial type of 'mentally ill person-dangerous'. This score is a measure of effect size and indexes the difference in average response latencies across the two trial-types; consistent and inconsistent, whilst accounting for the participants overall response speed. The experimental programme automatically produced this score. Using Cohen's (1992) effect size guidelines, participants received one of four pre-prepared pieces of feedback that was communicated to the participant via the researcher. Accordingly, a positive D-IRAP score in the range of .15-.34 resulted in participants being told that they had revealed a "slight implicit stigmatized attitude towards individuals with a mental illness". A D-IRAP score of .35-.64 resulted in participants being told that they had revealed a "moderate implicit stigmatized attitude in relation to individuals with a mental illness". Participants with a score of .65 or above were notified that they had displayed a "strong implicit stigmatized attitude", and a "no difference" feedback message was given to participants with D-scores within the range of 0.0-0.14 (Menatti et al., 2012). Negative D-IRAP scores indicated that participants held a positive implicit bias towards individuals with a mental illness, and this was communicated to participants using the same effect size guidelines (Cohen, 1992). Participants assigned to the conditions of ISC without feedback and control without feedback did not receive implicit attitude feedback.

Phase Two. Phase 2 acted as a follow-up to examine whether receiving implicit attitude feedback, or simply participating in the implicit attitude measurement procedure can reduce implicit and explicit attitudes. All 48 participants were re-recruited after a 24-hour gap

following initial experimental testing and re-assigned to their original experimental conditions. Participants then completed the IRAP and CAMI-SR for a second time. These measures were, again, administered in a counterbalanced design.

Results

Analytic Strategy

In relation to the statistical analytic procedure, a 2 (Experimental phase) 4 (IRAP trial type) 4 (Experimental condition) mixed factorial design was utilized. Specifically, experimental phase (1 and 2), and the four IRAP trial types (mentally ill person-dangerous, mentally ill person-harmless, physically ill person-dangerous, physically ill person-harmless) were input as within-subject variables. The four experimental conditions, as detailed above, were input as between-participants variables.

Implicit Measures

For clarity of interpretation the mean D- IRAP scores for the trial types of 'mentally ill person- dangerous' and 'mentally ill person- safe' were inverted so that positive scores indicate that participants held the belief that individuals with a mental illness were harmless, and negative scores indicate the belief that individuals with a mental illness were dangerous.

In line with previous IRAP analyses (Barnes-Holmes et al., 2010a; Barnes-Holmes et al., 2009; Cullen et al., 2009), four one-sample *t*-tests were conducted in order to examine if the mean D-IRAP score for each trial type differed significantly from zero. Across all four experimental conditions, participants were quicker to agree, rather than deny, that participants with a mental illness are harmless in both phase one (M = .23, SD = .48), $t(47) = 3.31$, p = .002, and phase two of the experiment (M = .22, SD = .60), $t(47) = 2.53$, p = .02. In a similar vein, participants were significantly quicker to deny, rather than agree, that individuals with a mental illness are dangerous in phase two of the experiment, (M = .19, SD = .55), $t(47)$ =

2.37, $p = .02$. The two trial types of 'physically ill person- dangerous' and 'physically ill person- harmless' were not significantly different to zero ($p > .05$), indicating that participants held a neutral response towards individuals with a physical illness, and were thus excluded from further interpretation.

Experimental condition. There was no main effect of experimental condition, $p > .05$. However, there was a significant two-way interaction between the four IRAP trial types and the four experimental conditions was observed, $F(5.47, 80.23) = 2.44$, $p = .04$, $\eta_p^2 = .14$. A subsequent mixed design multivariate analysis of variance (MANOVA) revealed that the interaction was in relation to the trial type 'mentally ill person- dangerous' and the four experimental conditions, Wilks $\Lambda = .738$, $F(6, 86) = 2.35$, $p = .04$, $\eta_p^2 = .14$. Pairwise comparisons further indicated that in phase two of the experiment, participants assigned to the condition of ISC with implicit feedback ($M = .52$, $SE = .15$) held a greater implicit positive bias compared to participants assigned to the control condition without implicit feedback ($M = -.17$, $SE = .15$), who held a negative implicit bias towards mental-ill health, $p = .01$ (see Figure 2). No other experimental conditions reached significance, $p > .05$. In addition, the trial type of 'mentally ill person- harmless' revealed a non-significant effect across conditions, $p > .09$.

Experimental Phase. A significant two-way interaction was also found between the four IRAP trial types, the four conditions, and experimental phase, $F(7.12, 104.42) = 2.98$, $p = .01$, $\eta_p^2 = .17$. A paired samples *t*-test indicated that participants in the condition of ISC with implicit attitude feedback were quicker to agree, than disagree, that individuals with a mental illness are harmless in phase one ($M = .21$, $SD = .49$), with this positive implicit attitude further increasing after a 24-hour gap in testing ($M = .63$, $SD = .62$), $t(11) = -2.19$, $p = .05$. A significant main effect also showed that participants in the ISC without feedback

condition first revealed a neutral attitude towards individuals with a mental illness in phase one (M = -.03, SD = .42) which then changed to a slight positive implicit attitude after a 24-hour gap in experimental testing (M = .28, SD = .42), $t(11) = -2.53$, $p = .03$. There were no significant main effects in relation to the other experimental conditions between experimental phases one and two, $p > .05$. Figure 3 displays the direction of this interaction. Again, positive scores indicate that participants held a positive belief that individuals with a mental illness are harmless, whereas negative scores indicate that participants held a stigmatizing belief that individuals with a mental illness are dangerous.

In sum, these results suggest that imagined social contact might be an effective attitude change intervention. Furthermore the addition of implicit attitude feedback further heightened participants' positive attitudes towards mental illness over a 24-hour period.

[INSERT FIGURE 3 HERE]

Explicit Measures

A 2 (experimental phase) × 4 (experimental condition) mixed factorial ANOVA revealed no significant difference in participants' explicit attitudes toward individuals with a mental illness as a function of the four experimental conditions, $p > .05$. There was also no significant difference between participants' stigmatizing explicit attitudes from phase one (M = 20.54, SD = 4.56) to Phase 2 (M = 19.56, SD = 4.97) of the experiment, $p > .05$. There was no interaction between experimental condition and phase, $p > .05$. Table 2 presents descriptive statistics for both the explicit and implicit attitude measures.

[INSERT TABLE 2 HERE]

Relationship Between Implicit and Explicit Attitudes

A Spearman's rank-order correlation revealed a moderate negative relationship

between the IRAP trial type of 'mentally ill person - dangerous' and the CAMI-SR completed in phase one, r = -.36, n = 48, p = .01. Additionally, a moderate negative correlation was revealed for the IRAP trial type of 'mentally ill person - harmless', and the CAMI-SR completed in phase one, r = -.39, n = 48, p = .01. There were no significant correlations between explicit and implicit attitudes in phase two of experimental testing (p > .05). Thus, although participants agreed that individuals with a mental illness were harmless on implicit attitude measures, they revealed a higher level of endorsed stigma on measures of explicit attitudes towards mental ill-health.

Discussion

In the first of its kind, the current study examined the effectiveness of ISC combined with implicit attitude feedback as a method for reducing public stigma toward individuals with a mental illness. The results suggest that ISC may successfully change implicit stigmatizing attitudes toward mental ill health, with the addition of implicit attitude feedback further heightening participants' positive implicit attitudes. Such results potentially illustrate novel ways of reducing barriers of stigmatization between the general public and individuals with a mental illness. This research suggests that the IRAP (Barnes-Holmes et al., 2006) may be a feasible tool for diagnosing stigmatizing implicit attitudes. As such, the findings indicate that combining methodological tools (e.g., the IRAP), with theoretical frameworks (e.g., ISC) can contribute to the theoretical and applied efficacy of attitude change interventions.

In Phase 1 of experimental testing, participants who were assigned to ISC without implicit attitude feedback held a neutral implicit attitude toward individuals with a mental illness, neither believing that they were harmless or dangerous. However, following a 24-hour break from experimental testing, participants revealed an implicit positive bias toward individuals with a mental illness. In Phase 2, participants assigned to the condition of ISC

combined with implicit attitude feedback displayed an implicit positive bias toward individuals with a mental illness compared to participants who did not undergo this intervention. Furthermore, the addition of implicit feedback in the control condition did not decrease participants stigmatizing implicit attitudes toward individuals with a mental illness. In other words, our findings suggest that ISC (Crisp & Turner, 2009) may be a direct and effective implicit attitude change intervention, with positive attitudes toward indivi- duals with a mental illness further enhanced by the addition of implicit attitude feedback.

These results appear to be in line with the assertion that when an individual recognizes the difference between their perceived behavior and the way they actually behave they may undergo a global feeling of discomfort, which leads to the individual engaging in efforts to decrease this discrepancy (Festinger, 1957; Monteith, Ashburn-Nardo, Voils, & Czopp, 2002). As such, in combination with ISC, simply making an individual consciously aware of their implicit attitudes may act as a 'cue for control' − increasing self-regulatory behavior and reducing stereotypical attitudes (ibid). From this perspective, combining implicit attitude measurement with implicit attitude feedback may influence individuals to reflect more deeply on their attitudes than they might during explicit procedures alone.

The theory of ISC is also consistent with the theoretical framework of mental simulation (Crisp & Birtel, 2014; Miles & Crisp, 2014), which posits that the effects of mental imagery on behavior operate via the availability of mental scripts and cognitive representations (Schank & Albelson, 1977). Drawing on this literature may provide a greater explanation as to why ISC is most effective when combined with implicit attitude feedback. Perhaps the very conceptualization of a positive contact situation contradicts the implicit attitudes held by participants pre-experiment, thus creating new cognitive representations and mental scripts. Such an explanation would be in contrast to primary cognitive theorizing (Hughes, Barnes-Holmes, & De Houwer, 2011) which postulate that implicit attitudes − as

mental representations − should be stable across time and contexts (Bargh, 1999; Devine, 1989; Greenwald & Banaji, 1995; Smith & De Coster, 1999; Wilson et al., 2000). Rather, results from the current study may support the recent assertion that implicit attitudes may be influenced by social and contextual cues (Barden, Maddux, Petty, & Brewer, 2004; Barnes-Holmes et al., 2010b; Cullen et al., 2009; Ferguson & Bargh, 2004; Hahn, Judd, Hirsch, & Blair, 2013). Viewing implicit attitudes as adaptive rather than fixed representations may be an important premise for the development of interventions that strive for stigma reduction by addressing both explicitly and implicitly held attitudes.

Contrary to previous research (Menatti et al., 2012; Stathi et al., 2012), participants' explicit attitudes towards individuals with a mental illness were not found to be malleable through the procedures employed. Similarly, participants' explicit attitudes appeared to diverge from their implicit attitudes. These results are in accordance with similar research into the malleability of ageist (Cullen et al., 2009) and political attitudes (Power et al., 2009), and may be explained by re-visiting the REC model (Hayes et al., 2001). This model posits that divergent implicit and explicit attitudes occur when brief and immediate relational responses (i.e., implicit attitudes) are contaminated to a lesser extent by extended and elaborated relational responses (i.e., explicit attitudes; Hughes, Barnes-Holmes, & Vahey, 2012). This is thought to occur when the behavioral response system is put under time pressure (Barnes-Holmes et al. 2010a), as is the case in the IRAP. In the current study, participants' explicit attitudes revealed a slight stigmatizing attitude towards individuals with a mental illness, regardless of experimental condition. Through the lens of the REC model (Hayes et al., 2001) it is conceivable that the strength of participants' attitudes, captured by explicit measurements, may reflect the generalized societal or meritocratic view that individuals with a mental illness are dangerous (Dickerson et al., 2002; Menatti et al., 2012; Rose et al., 2007; Rüsch, Todd, Bodenhausen, & Corrigan, 2010), rather than participants'

own beliefs. This viewpoint provides additional support for the notion that explicit attitude measures might be vulnerable to contextual control (Corrigan & Shapiro, 2010; Davies, 1997; Rydell & McConnell, 2006; Schwarz, 1999; Power et al., 2009). Therefore, as individuals may adhere to socially acceptable expectations when attitudes are gauged explicitly, there would appear to be a great potential utility in employing implicit methodologies in the measurement of socially sensitive attitudes. Accordingly, researchers should look to employ both explicit and implicit measures to reliably observe a reduction in stigmatizing attitudes.

A contradictory theory, however, argues that both explicit and implicit attitudes may co-exist in memory (Wilson et al., 2002), and influence behavior in diverse ways (Fazio & Towles-Schwen, 1999). The model of dual attitudes (Wilson et al., 2000) posits that the attitudes that individuals endorse at any given time depend on whether the explicit attitude is able to override an implicitly held one. In line with this assertion, participants in the current study may not have had the cognitive capacity to report reliably a change in their explicit attitudes. This interpretation is supported by Gawronski and LeBel (2008) who suggest implicit and explicit attitudes may diverge in a number of instances: First, when categorization of an attitude object with positive or negative evaluative terms creates new automatic associations in memory. Second, when additional consideration of the attitude object eliminates the impact of automatic relations on self-reported evaluative judgments. In the process of considering a hypothetical contact situation, therefore, participants may have acquired a new implicit attitude but may not have been able to reliably report this via explicit self-report measures. This notion contributes to recent research which has provided evidence for the existence of two, distinct cognitive categorization systems – one explicit and one implicit – which are adaptive in nature (Smith et al., 2014). Past interventions that solely target explicit attitudes may thus be neglecting the presence of unintentional, but yet stigmatizing, implicit attitudes (Greenwald & Banaji, 1995).

Limitations and Future Recommendations

There are a number of noteworthy limitations to the current research. First, the ISC scenario (Crisp & Turner, 2009) asked participants to imagine an individual with a 'mental illness'. Although this intervention was found to be effective in reducing implicit attitudes towards mental ill-health, the generic term of 'mental illness' was open to participants' own interpretations, and it is not known specifically *what* mental illness they may have been imagining. For example, research indicates that the general public commonly perceives people with schizophrenia, alcoholism and drug addiction as unpredictable and dangerous (Crisp et al., 2000). However, other mental health diagnoses, such as mild depression, may not be subject to the same stigmatizing evaluations. It is therefore plausible that the ISC experience may vary depending on the mental health diagnosis included in the manipulation. Nevertheless, participants were asked to define the term 'dangerous' before tests of implicit and explicit attitudes were administered and this may counteract this limitation to some extent. Specifically, this prime may have influenced participants to subsequently imagine a person with mental illness who would be typically classified as 'dangerous' in a positive encounter.

Furthermore, some participants were unable to complete the implicit attitude measure in a quick and accurate manner and were thus disregarded from the final analyses. Nevertheless, the use of such stringent criteria ensures that participants understand and comply with the IRAP instructions and therefore increase the validity and reliability of IRAP performance (Barnes-Holmes et al., 2010b). The complexity of the task also limits contaminating sources of contextual control (Barnes-Holmes et al., 2010a) and faking (McKenna, Barnes-Holmes, Barnes-Holmes, & Stewart, 2007). In this respect, the IRAP may provide researchers with an efficacious tool to diagnose and reduce implicit stigmatizing attitudes, which is freely accessible and available for wide dissemination. Furthermore, given

that the IRAP provides a direct measure of implicit attitudes and produces automatically the attitude score, it may be utilized to provide a measure of implicit attitude feedback. It should be noted, however, that the current research did not examine the degree of implicit attitude feedback that is required to elicit a change in individuals' implicit attitudes. Future research may therefore benefit from ascertaining whether the degree of implicit attitude feedback (i.e., positive, weak, moderate or negative feedback) moderates attitude change. It is plausible, for example, that receiving positive feedback – that the participant does not hold stigmatizing mental health attitudes – may elicit a different cognitive operation in comparison to receiving negative feedback.

In line with procedures by Crisp and Turner (2009), participants' implicit attitudes were measured immediately after the ISC manipulation. The effectiveness of this manipulation was then measured after a 24-hour gap in experimental testing. It is therefore possible that the reduction in individuals' implicit ostracizing attitudes towards individuals with a mental illness were the result of positive information priming. Indeed, research suggests that implicit attitudes may be shaped by exemplars that support 'thinking the opposite' (Cullen et al., 2009). Future research should therefore explore the longer-term effectiveness of ISC as a means to reduce stigmatizing attitudes towards individuals with a mental illness, especially with respect to the malleability of implicit attitudes.

Furthering this, participants were asked to categorize the terms 'mentally ill person' and 'physically ill person' to the IRAP target stimuli of 'dangerous' and 'harmless'. It is plausible that the language used in the implicit test could reflect benevolent paternalism rather than an increase in positive attitudes (Link, 2004). As such, researchers should be aware that IRAP effects might depend on the stimuli employed to instantiate target and attribute concepts (De Houwer, 2002).

In conclusion, implicit attitudes have predominantly been viewed as fixed and immutable (Bargh, 1999; Devine, 1989; Greenwald & Banaji, 1995; Smith & Decoster, 1999; Wilson et al., 2000). Consequently, the development of interventions aiming to change implicit attitudes has been neglected (Wilson et al., 2000). The current study, however, points to the potential effectiveness of ISC (Crisp & Turner, 2009) as a means of changing socially sensitive and stigmatizing implicit attitudes towards mental ill-health. It is also the first empirical demonstration to suggest that ISC (Crisp & Turner, 2009) combined with implicit attitude feedback (Menatti et al., 2012) may be a particularly effective means of bringing about implicit attitude change. Such findings may help to underpin effective public awareness campaigns aimed at lessening stigmatizing attitudes towards mental ill-health and improving the psychological wellbeing of those adversely affected by mental illness. Researchers might benefit from utilizing paradigms that target both implicit and explicit attitudes in the development of attitude change interventions.

References

Allport, G. W. (1935). *A handbook of social psychology.* Worcester: University Press.

Allport, G. W. (1954). *The nature of prejudice.* Massachusetts: Addison Wesley.

Angermeyer, M. C., & Matschinger, H. (1996). The effect of personal experience with mental illness on the attitude towards individuals suffering from mental disorders. *Social Psychiatry and Psychiatric Epidemiology, 31,* 321-326.

Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. E. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology, 87,* 5-22.

Bargh, J. A. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361-382). New York: Guilford.

Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, R., & Stewart, I. (2006). Do you really know what you believe? Developing the Implicit Relational Assessment Procedure (IRAP) as a direct measure of implicit beliefs. *The Irish Psychologist*, *32*, 169-177.

Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I., & Boles, S. (2010a). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the relational elaboration and coherence (REC) model. *The Psychological Record, 60,* 527-542.

Barnes-Holmes, D., Murphy, A., Barnes-Holmes, Y. & Stewart, I. (2010b). The Implicit Relational Assessment Procedure (IRAP): Exploring the impact of private versus public contexts and the response latency criterion on pro-white and anti-black stereotyping among white Irish individuals. *The Psychological Record, 60,* 57-80.

Barnes-Holmes, D., Waldron, D., Barnes-Holmes, Y., & Stewart, I. (2009). Testing the validity of the Implicit Relational Assessment Procedure (IRAP) and the Implicit Association Test (IAT): Measuring attitudes towards Dublin and country life in Ireland. *The Psychological Record, 59*, 389-406.

Brown, S. A. (2012). The contribution of previous contact and personality traits to severe mental illness stigma. *American Journal of Psychiatric Rehabilitation, 15,* 274-289.

Cullen, C., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). The Implicit Relational Assessment Procedure (IRAP) and the malleability of ageist attitudes. *The Psychological Record*, *59*, 591-620.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Corrigan, P. W. (2000). Mental health stigma as social attribution: Implications for research methods and attitude change. *Clinical Psychology-Science & Practice, 7,* 48-67.

Corrigan, P. W. (2002). Testing social cognitive models of mental illness stigma: The prairie state stigma studies. *Psychiatric Rehabilitation Skills, 6,* 232-254.

Corrigan, P. W. (2004). How stigma interferes with mental health care. *American Psychologist*, *59*, 614-625.

Corrigan, P. W., Larson, J. E., & Rüsch, N. (2009). Self-stigma and the "why try" effect: Impact on life goals and evidence-based practices. *World Psychiatry*, *8*, 75-81.

Corrigan, P. W., & Rüsch, N. (2002). Mental illness stereotypes and clinical care: Do people avoid treatment because of stigma? *Psychiatric Rehabilitation Skills, 6*, 312-334.

Corrigan, P. W, & Shapiro, J. R. (2010). Measuring the impact of programs that challenge the stigma of mental illness. *Clinical Psychology Review, 30,* 907-922.

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

Crisp, R. J., & Birtel, M. D. (2014). Reducing prejudice through mental imagery: Notes on replication, interpretation and generalization. *Psychological Science, 25,* 840-841.

Crisp, A. H., Gelder, M. G., Rix, S., Meltzer, H. I., & Rowlands, O. J. (2000). Stigmatisation of people with mental illnesses. *British Journal of Psychiatry*, *177*, 4-7.

Crisp, R. J., & Turner, R. N. (2009). Can imagined interactions produce positive perceptions? Reducing prejudice through simulated social contact. *American Psychological Association, 64*, 231-240.

Davies, J. B. (1997). The Myth of Addiction. Amsterdam: Psychology Press.

De Houwer, J. (2002). The Implicit Association Test as a tool for measuring dysfunctional associations in psychopathology: Strengths and limitations. *Journal of Behavior Therapy and Experimental Psychiatry, 33,* 115-133.

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

Dickerson, F. B., Sommerville, J. L., Origoni, A. E. (2002). Mental illness stigma: An impediment to psychiatric rehabilitation. *Psychiatric Rehabilitation Skills, 6,* 186-200.

Dijksterhuis, A., Bargh, J. A., & Mark, P. Z. (2001). The perception-behaviour expressway: Automatic effects of social perception on social behaviour. *Advances in Experimental Social Psychology, 33,* 1-40.

Dinos, S., Stevens, S., Serfaty, S., Weich, S., & King, M. (2004). Stigma: The feelings and experiences of 46 people with mental illness. *British Journal of Psychiatry*, *184*, 176-181.

Fazio, R. H., & Towles-Schwen, T. (1999). The MODE model of attitude-behavior processes. In S. Chaiken & Y. Troope (Eds.), *Dual-process theories in social psychology* (pp. 97-116). New York: Guilford.

Ferguson, M. J., & Bargh, J. A. (2004). Liking is for doing: The effects of goal pursuit on automatic evaluation. *Journal of Personality and Social Psychology, 87,* 557-572.

Festinger, L. (1957). A theory of cognitive dissonance. Stanford, CA: Stanford University Press.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage Publications.

Gawronski, B., & LeBel, E. P. (2008). Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Psychology, 44*, 1355-1361.

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

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109,* 3-25.

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

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197-216.

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General, 143,* 1369-1392.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational Frame Theory: A post-Skinnerian account of human language and cognition.* New York: Plenum.

Hughes, S., & Barnes-Holmes, D. (2011). On the formation and persistence of implicit attitudes: New evidence from the Implicit Relational Assessment Procedure (IRAP). *The Psychological Record, 61,* 391-410.

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorizing in implicit attitude research: Propositional and behavioural alternatives. *The Psychological Record, 61,* 465-496.

Hughes, S., Barnes-Holmes, D., & Vahey, N. (2012). Holding on to our functional roots when exploring new intellectual islands: A voyage through implicit cognition research. *Journal of Contextual Behavioural Science, 1,* 17-38.

Kester, J. D. (2001). A revolution in social psychology. *The American Psychological Society Observer*, *14*, 9-26.

Kim, D. (2003). Voluntary controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly, 66,* 83-96.

Link, B. G., Yang, L. H., Phelan, J. C., & Collins, P. Y. (2004). Measuring mental illness stigma. *Schizophrenia Bulletin, 30,* 511-541.

Link, B. G., Struening, E. L., Neese-Todd, S., Asmussen, S., & Phelan, J. C. (2002). On describing and seeking to change the experience of stigma. *Psychiatric Rehabilitation Skills, 6,* 201-231.

Lynott, D., Kansal, H., Connell, L., & O'Brien, K. (2012). Modelling the IAT: Implicit association test reflects shallow linguistic environment and not deep personal attitudes. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society, pp. 1948-19.53.

McKenna, I. M., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2007). Testing the fake-ability of the Implicit Relational Assessment Procedure (IRAP): The first study. *International Journal of Psychology and Psychological Therapy*, *7*, 123-138.

Menatti, A., Smyth, F., Teachman, B. A., & Nosek, B. A. (2012). Reducing stigma toward individuals with mental illnesses: A brief, online intervention. *Stigma Research and Action*. http://stigmaj.org/. Available from: http://projectimplicit.net/nosek/papers/MSTN2012.pdf

Miles, E., & Crisp, R. J. (2014). A meta-analytic test of the imagined contact hypothesis. *Group Processes & Intergroup Relations, 17,* 3-26.

Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology*, *83*, 1029-1050.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259.

Pescosolido, B. A., Monahan, J., Link, B. G., Stueve, A., & Kikuzawa, S. (1999). The public's view of the competence, dangerousness, and need for legal coercion of persons with mental health problems. *American Journal of Public Health*, *89*, 1339-1345.

Power, P., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). The Implicit Relational Assessment Procedure (IRAP) as a measure of implicit relative preferences: A first study. *The Psychological Record, 59,* 621-640.

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,* 377-389.

Rüsch, N., Angermeyer, M. C., & Corrigan, P. W. (2005). Mental illness stigma: Concepts, consequences and initiatives to reduce stigma. *European Psychiatry*, *20*, 529-539.

Rüsch, N., Todd, A. R., Bodenhausen, G. V., Corrigan, P. W. (2010). Do people with mental illness deserve what they get? Links between meritocratic worldviews and implicit versus explicit stigma. *European Archives of Psychiatry and Clinical Neuroscience, 260,* 617-625.

Rüsch, L. C., Kanter, J. W., Angelone, A. F., & Ridley, R. C. (2008). The impact *in our own voice* one stigma. *American Journal of Psychiatric Rehabilitation, 11,* 373-389.

Rose, D., Thornicroft, G., Pinfold, V., & Kassam, A. (2007). 250 labels used to stigmatise people with mental illness. *BMC Health Services Research*, *7*, 1-7.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology, 91*, 995-1008.

Sacca, R., & Ryan, C. (2011). Relationships between interpersonal contact as a volunteer companion and stigma. *Australasian Psychiatry*, *19*, 439-443.

Schwarz, N. (1999). Self-report; how the questions shape the answers. *American Psychologist, 54*, 93–105.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding.* Hillsdale: NJ, Erlbaum.

Smith, J. D., Boomer, J., Zakrzewski, A. C., Roeder, J. L., Church, B. A., & Ashby, F. G. (2014). Deferred feedback sharply dissociates implicit and explicit category learning. *Psychological Science, 25,* 447-457.

Smith, E. R., & De Coster, J. (1999). Associative and rule-based processing: A connectionist interpretation of dual process models. In S. Chaiken & Y. Ropes (Eds.), *Dual-process theories in social psychology* (pp. 323-336). New York: Guilford.

Stathi, S., Tsantila, K., & Crisp, R. J. (2012). Imagining intergroup contact can combat mental health stigma by reducing anxiety, avoidance and negative stereotyping. *The Journal of Social Psychology*, *152*, 746-757.

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

Teachman, B. A., Wilson, J. G., & Komarovskaya, I. (2006). Implicit and explicit stigma of mental illness in diagnosed and health samples. *Journal of Social and Clinical Psychology, 24,* 75-95.

Vahey, N. A., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). A first test of the implicit relational assessment procedure as a measure of self-esteem: Irish prison groups and university students. *The Psychological Record, 59,* 371-388.

Wegner, D. M. (2002). *The illusion of conscious will.* Bradford Books: Cambridge, MA.

West, K., Holmes, E., & Hewstone, M. (2011). Enhancing imagined contact to reduce prejudice against people with schizophrenia. *Group Processes and Intergroup Relations, 14,* 407-428.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*, 101-126.

Footnotes

[1] The IRAP experimental script produces a summary of participants' accuracy and response latency data after each trial. This summary allowed us to evaluate participants' ability in the practice trials.

Table 1. *The Stimulus Arrangements employed in the IRAP.*

Sample Stimulus 1	**Sample Stimulus 2**
MENTALLY ILL PERSON	PHYSICALLY ILL PERSON
Response Option consistent with Sample 1	**Response Option consistent with Sample 2**
SIMILAR	OPPOSITE
Target Stimuli consistent with Sample 1	**Target Stimuli consistent with Sample 2**
DANGEROUS	HARMLESS
IRRATIONAL	RATIONAL
VIOLENT	GENTLE
UNPREDICTABLE	PREDICTABLE
INSANE	SANE
CRAZY	SENSIBLE

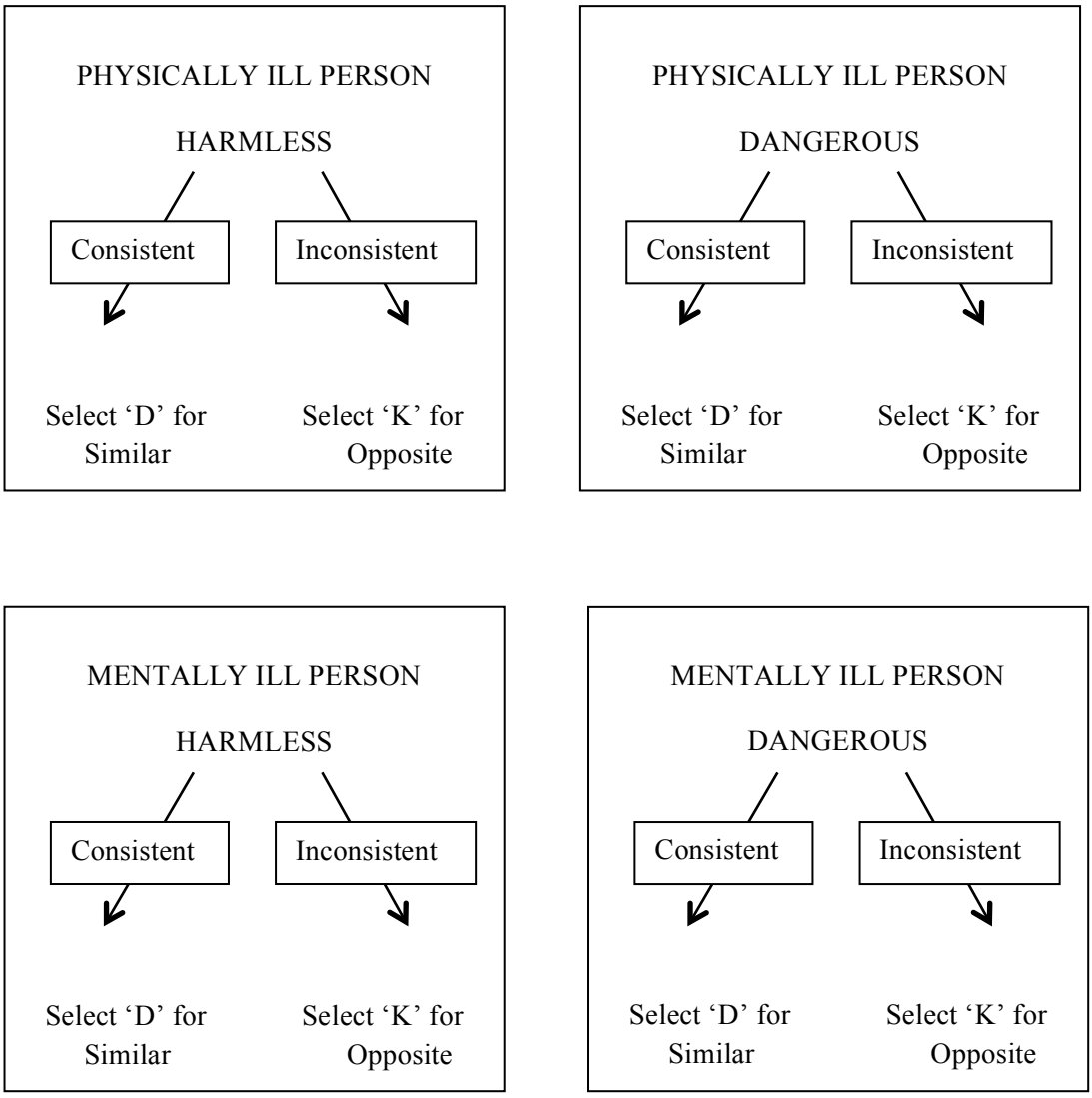

Figure 1. Schematic representations of the four IRAP trial types employed in the current experiment. The arrows and words 'consistent' and 'inconsistent' did not appear in the IRAP, but were present in the instructions and are included here for illustrative purposes.

Table 2.

Means and standard deviations for implicit and explicit attitudes across the four experimental conditions and two experimental phases.

	ISC with feedback		ISC without feedback		Control with feedback		Control without feedback	
	Phase							
	P1	P2	P1	P2	P1	P2	P1	P2
Mentally ill – Dangerous	.21 (.49)	.52 (.59)	-.03 (.42)	.28 (.42)	-.21 (.68)	.13 (.63)	-.002 (.45)	-.17 (.35)
Mentally ill – Harmless	.21 (.49)	.63 (.62)	.25 (.39)	.04 (.46)	.05 (.59)	.08 (.65)	.40 (.39)	.13 (.53)
CAMI-SR	20.33 (3.34)	18.50 (4.58)	18.25 (4.09)	18.08 (4.52)	22.00 (5.20)	21.75 (5.55)	21.58 (4.94)	19.92 (4.94)

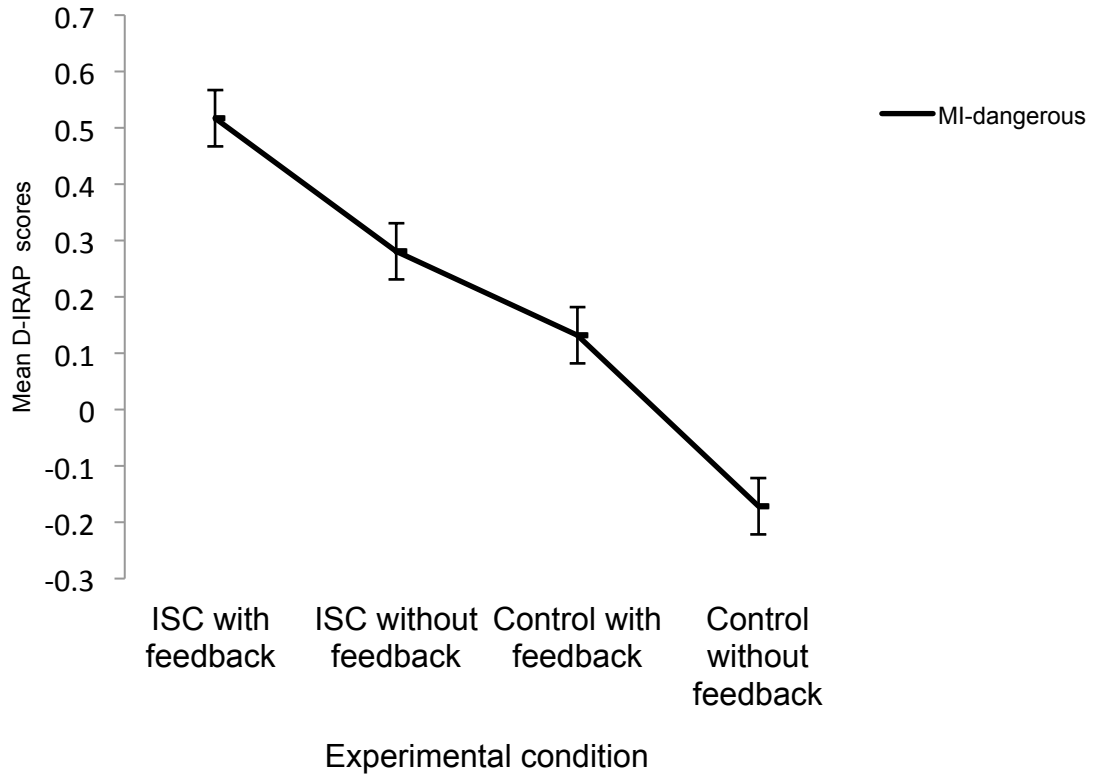

Figure. 2. Mean D-IRAP scores for the trial type of mentally ill person-dangerous across the four conditions in phase two of experimental testing.

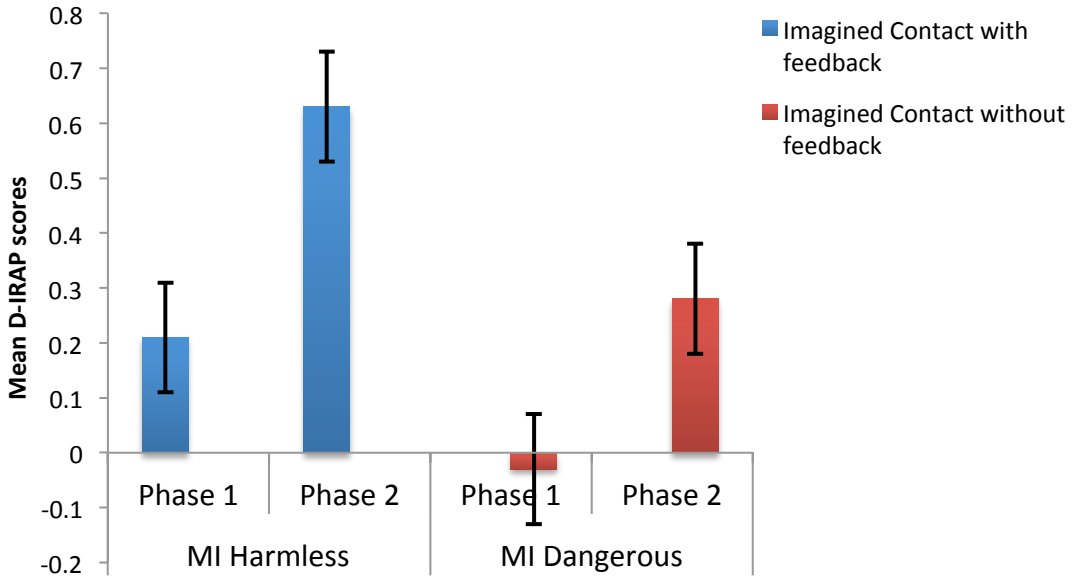

Figure. 3. Mean D-IRAP scores for the significant trial types across the conditions of ISC with and without feedback in the two experimental phases.