Methodological, Social and Clinical Applications of the Implicit Relational Assessment Procedure (IRAP)

NUI MAYNOOTH

Ollscoil na h Éireann Má Nuad

Claire Campbell B.A.

Thesis submitted to the National University of Ireland, Maynooth in fulfillment of the requirements for a PhD. in Psychology

Supervised by Dr. Yvonne Barnes-Holmes
Department of Psychology
October, 2008

To my parents, Ann and Hugh,
To Niamh and,
To Richie
Thank you for all the love and support

Table of Contents

Acknowledgements

Acknowledgements

Thank you to Yvonne for being a great inspiration and mentor. Thank you to my family and friends for all their support and encouragement. Thank you to all the postgrads who have made this experience a pleasure and who were great for advice or a laugh. And most importantly, thank you to everyone who participated in my research and a special thank you to Caiti Keegan, Claire McGoldrick, Aoife O'Kelly, Aislinne Freeman, Louisa Murphy, Suzanne Comerford and Hilary Duffy who's research formed part of the current work.

Abstract

The aims of the current thesis were three fold. The first aim was primarily methodological and attempted to determine the stimulus presentation that would produce the strongest IRAP effect. *Experiment 1* set out to address this simple experimeental question. The key manipulation involved the presence or absence of the randomisation of the sample stimuli and/or the respponse options within the IRAP. The results indicated that the ramdomisation of the sample stimuli (with or without the randomisation of the response options) generated the strongest D-$_{IRAP}$ scores. Randomising the response options alone without randomising the samples also produced a significant IRAP effect, although this was smaller. The D-$_{IRAP}$ score that resulted from the static presentation of both samples and response options was almost negligible.

The second aim of the thesis was to determine the utility of the IRAP as a measure of implicit attitudes to race and this was the core focus of Experiments 2 to 6. In the IRAP in Experiment 2, participants were simply required to relate the phrases WHITE PERSON and BLACK PERSON to a range of simple positive and negative terms that could be readily categorised as safe and dangerous. We predicted that our Irish participants would show a pro- White and anti- Black bias. The D-$_{IRAP}$ trial-type scores were in part consistent with experimental predictions and showed a pro- White bias where White was safe, but not an anti-Black bias. In fact, participants significantly confirmed that Black was safe. This coincided with the explicit measures, on which participants indicated the absence of racial bias against Black people.

In line with existing IAT evidence, to determine whether racial pictures would generate stronger IRAP effects than words, and to examine the potential role of

attentional weapon bias, the IRAP in *Experiment 3* presented pictures of Black and White men holding guns. For correct responding, however, the presence of the guns was incidental and the discriminations among the pictures were based on race alone. We predicted that Irish participants would show a pro- White and anti- Black bias that was perhaps enhanced by the presence of the guns. Participants significantly confirmed that White was safe and Black was dangerous, although they also significantly confirmed that Black as safe. Hence, this was our first evidence of an anti- Black bias, but other effects were not entirely as expected. Again on the majority of explicit measures, participants indicated the absence of racial bias.

In order to determine whether the anti-Black bias in the previous study did reflect the influence of the guns, Experiment 4 presented pictures of Black and White men holding mundane objects. We predicted that Irish participants would show a pro-White and anti- Black bias, even though both were holding mundane objects. Participants significantly confirmed that both White and Black were safe, but had mixed views of either as dangerous. Hence, the anti- Black bias from the previous study was now lost, suggesting that the weapons had facilitated the effect. Again, on the majority of the explicit measures, participants indicated the absence of racial bias.

In line with previous IAT studies of race and to examine further the outcomes from the previous studies, *Experiment 5* presented pictures of Black and White men holding *both* guns and mundane objects. We predicted that Irish participants would show a pro- White and anti- Black bias that may be differentially influenced by the items presented in the pictures. The four trial-types examined in the previous studies were sub-divided into eight to permit comparisons of the relative influence of the different items held in the pictures. Overall, seven of the eight trial-types were non-significant and participants only significantly confirmed that White men with

mundane objects were safe. In fact, close inspection of the non-significant effects indicated unexpected evidence of pro- Black bias. Again these agreed with the explicit measures, on which participants indicated the absence of racial bias.

In Experiment 6, participants were presented with the same IRAP as the previous study, but were required to discriminate the objects in the pictures as the basis for correct responding and ignore race. That is, on consistent trials guns were always dangerous and mundane objects were always safe and correct responding on inconsistent trials was reverse (guns always safe and mundane objects dangerous). We predicted that Irish participants would categorise guns as dangerous and mundane objects as safe, but expected that these effects might be influenced by race. In this study, six of the eight were non-significant. Participants only significantly confirmed that Black with a gun was not safe and White with a gun was dangerous, but there was no evidence of clear influence of either items or race in the remaining effects. Again, the explicit measures indicated the absence of racial bias.

The third aim of the current thesis was to examine the utility of the IRAP as a measure of clinically relevant implicit cognitions. Previous preliminary research had indicated strong IRAP effects that showed a positive bias towards the self in participants with a high explicit self-esteem. In *Experiment 7*, participants were required to indicate their agreement or disagreement with a range of statements that reflected high self-esteem (i.e. were positive about the self) or reflected low self-esteem (i.e. were negative about the self). Because the study was aslo concerned with whether on not the IRAP would correlate with an explicit measure of self, the positive and negative statements presented in the IRAP were taken directly from the Roseberg Self-Esteem Scale (RSES) and particiants also completed this as an explicit measure.

We predicted that participants high on explicit self-esteem would relate the self with the positive self-statements and not relate the self with negative self-statements.

The findings indicated that participants significantly confirmed that they agreed with the positive (high self-esteem) statements and significantly disconfirmed that they disagreed with them. They also disconfirmed that they agreed with the negaative statements, although this effect was non-significant. Although these findings were largely as expected from participants who all scored as high in self-esteem on the explicit RSES, there was no significant correlation between the two measures.

In *Experiment 8,* we used the IRAP to explore implicit attitudes towards acceptance and avoidance. For this purpose, we selected undergraduate students who showed an xplicit propensity towards high acceptance/ low avoidance on the Acceptaance and Action Questionnaire (AAQ-2). In order to enhance the potential overlap between the two types of measure, as we had done in Experiment 7, we took opposing satements directly from the AAQ and inserted them into the IRAP as target stimuli. The results indicated that participants significantly confirmed that they agreed with the acceptance statements, but all other effects were non-significant. Closer inspection of the trial-types, however, were as expected and showed that particpants disconfirmed that they agreed with the avoidance statements and confirmed that they disagreed with them. In spite of these outcomes, the IRAP data did not correlate significantly with the AAQ.

The current reaearch program offered preliminary investigations of the utility of the IRAP as a measure of impicit attitudes to a range of psychological phenomena, including race, self-esteem and acceptance. Although the IRAP effects in many cases were not as significant and predicted, particularly in the context of race, there

wasgood evidence thata further investigation would be promising. The findings here make a useful contribution to the small existing IRAP research base by highlighting key issues that influence IRAP outcomes in a range of areas.

Chapter 1
General Introduction

Chapter 1

General Introduction

Relational Frame Theory (RFT) is a modern behavioural approach to human language and cognition. The basic account is built on the pivotal assumption that the behavioural units of language and higher-cognitive functioning are best defined in terms of *derived stimulus relations* (Hayes, Barnes-Holmes, & Roche, 2001). For RFT, derived relational responding is primarily *arbitrary* in nature and is governed by contextual cues via an appropriate history of multiple-exemplar training (see Barnes, 1996; Barnes, & Holmes 1991; Barnes, & Roche, 1996; Hayes, & Hayes, 1989). The distinction between arbitrary responding and the types of non-arbitrary responding commonly observed in non-humans is an important distinction for RFT. For example, animals can be trained to select the larger, smaller, or dimmer stimulus from a range of choices (see Reese, 1968), but this behaviour is almost certainly governed by the formal properties of the stimuli involved (e.g. size). The complex functioning of specific contextual cues permits language-able humans to develop repertoires of relational responding that are arbitrary, and as such can control relations among stimuli that share no formal overlap whatsoever. For example, in many monetary systems smaller coins are often worth more than larger coins. For RFT, this relational responding is referred to as *arbitrary applicable* because it is governed by contextual cues established by social convention, rather than being governed by the physical or formal properties of the stimuli.

According to RFT, contextual cues facilitate various patterns of relational responding that are collectively referred to as *relational frames*. All relational frames possess the same three properties of mutual entailment, combinatorial entailment and

the transfer or transformation of stimulus functions, but each is identified according to the core type of stimulus relations involved. Specifically, RFT describes frames of co-ordination, opposition, distinction, hierarchy and perspective-taking, (Dymond, & Barnes, 1994).

The term *mutual entailment* encompasses Sidman's (1971) symmetrical responding, but also refers to responding that is not strictly symmetrical. For example, in a frame of co-ordination, if A is the same as B, then B is the same as A; that is, the relation is symmetrical. In a comparative frame, however, if A is better than B, then B is worse than A. The relations, therefore, are not symmetrical but they are mutually entailed.

Similarly, the term *combinatorial entailment* encompasses Sidman's transitivity, but also refers to relations beyond the scope of those commonly thought of as transitive. For example, transitive responding in the context of a frame of co-ordination would entail that if A is the same as B, and B is the same as C, then A is the same as C. However, in the case of a frame of opposition, if A is the opposite of B, and B is the opposite of C, then the relation between A and C is one of sameness, not opposite. Therefore, the relations are not *strictly transitive*, but they are combinatorially entailed. That is, the relations between A and B and between B and C combine to entail the relations between A and C and between C and A. According to RFT, therefore, both transitivity and equivalence responding are instances of combinatorial entailment in which the trained relations are the same as the derived relations (Hayes et al, 2001). An alternative term that captures the property of reflexivity does not appear to be necessary for RFT. In any case, some researchers have questioned the utility of reflexivity as a defining property of derived relational

responding because such responding may be based upon either derived stimulus relations or formal similarity (Barnes, 1994).

The third defining feature of RFT is a *transfer or transformation of functions*. When stimuli are involved in a relational frame, any psychological function attached to one of the stimuli may transfer through the frame to any, or all, of the other stimuli involved. If the relevant frame is one of co-ordination, for example, then a similar function will attach to each stimulus in the frame because the relations among them are all co-ordinated (e.g. Barnes, & Keenan, 1993). For example, if a child learns to stop talking when a teacher says “quiet” and this word participates in a relational frame of co-ordination with the Irish word "ciunas," then the child may stop talking when the teacher says "ciunas." For RFT, this effect is due to the function of the word “quiet” transferring to the word "ciunas" through the derived relation between the English and Irish words. If, however, the relational frame is not one of co-ordination the functions will be *transformed* in accordance with the relations involved. For example, if two stimuli participate in a frame of comparison, such that A is “more than” B, and B is established as discriminative for a low-response rate, A may acquire a discriminative function for a high response rate based of its “more than” relation with A.

According to RFT, any stimuli may participate in a given instance of arbitrarily applicable relational responding, given the appropriate contextual cues. Hence, stimulus equivalence and other derived relational responses are viewed as a form of generalised operant behaviour. There is now a growing body of data that provides empirical support for the assumption that responding in accordance with the arbitrarily applicable relations of ‘different’, ‘opposite’, ‘perspective’ ‘comparison’ are forms of generalised operant responding (see Dymond, & Barnes, 1995).

Empirical Evidence for Relational Responding

As well as generic support for the generalised operant nature of derived relational responding, numerous studies have also supported RFT's identification of the various types of relational frames. Specifically, there is empirical evidence for co-ordination relations (Barnes, Browne, Smeets, & Roche, 1995); comparison relations (Vitale, Barnes-Holmes, Barnes-Holmes, & Campbell, 2008); opposition relations (Dymond, & Barnes, 1996); deictic frames of perspective (McHugh, Barnes-Holmes, O'Hora, & Barnes-Holmes, 2004); and even for relational responding between relational frames (Stewart, Barnes-Holmes, Roche, & Smeets, 2001).

A variety of methodologies have been employed to examine derived relational responding. These include problem-solving procedures on a continuum for the very simple for use with young children to the very complex involving numerous overlapping relational networks. Other procedures employed in RFT research to date include: Match-to-Sample (MTS); respondent-type training procedures; and relational evaluative procedures (REPs).

Problem Solving Procedures may be table-top or automated and are commonly employed with young participants (Barnes-Holmes, Barnes-Holmes, Smeets, Strand and Friman, 2004). These researchers developed a problem-solving procedure as a means of testing and training co-ordination relations in a sample of young children. The methodology involved presenting each child with identically-sized paper coins and describing how the coins compared to one another in terms of their value. The child was then asked to pick the coin that would 'buy as many sweets as possible'. These interventions were successfully used to establish increasingly complex patterns of co-ordination responding in all participants, including generalisation to novel stimuli and novel experimenters.

Matching-to-sample (MTS) procedures have been the most commonly employed methodology in the study of stimulus relations. This generally involves training participants in a series of related conditional discriminations using arbitrary stimuli, such as nonsense syllables. For example, in the presence of the sample stimulus A1, participants must choose comparison B1 over the alternative comparison B2. In contrast, they must choose B2 (not B1) when the sample A2 is present. Hence the relations A1-B1 and A2-B2 are established. Subsequently, participants are then trained to select C1 in the presence of B1 (now a comparison rather than a sample) and C2 in the presence of B2. Hence, the additional relations B1-C1 and B2-C2 have also been established. Following the conditional discrimination training, participants are tested to determine if the expected untrained relations among the A and C stimuli have emerged as a result of training the baseline relations. For example, if presented with C1 participants should choose A1 and vice versa, and when presented with C2 participants should choose A2 and vice versa. In this way, the MTS training established the co-ordination relations (also called stimulus equivalence classes) involving the A and C stimuli. Of course, MTS procedures have been effective in establishing many more complex patterns of relational responding than those in the current example (e.g. Stewart, Barnes-Holmes, Roche, & Smeets, 2001).

The dominance of the MTS procedure as a methodology for studying stimulus relations has caused concern among some researches. Typically in behaviour analysis, core concepts are *transituational* and are thus not tied to specific experimental procedures (Barnes-Holmes, Dymond, Roche, & Grey, 1999). Consider for example, reinforcement, discrimination, primary stimulus generalisation and respondent conditioning. Each of these has been observed and manipulated in a variety of basic

and applied contexts. In contrast, equivalence relations in particular have been almost entirely tied to the MTS procedure for both experimental and applied purposes.

It has been within this drive to develop new procedures for examining the multiple stimulus relations that RFT researchers, in particular, have made considerable efforts. Put simply, if the concept of derived relational responding is a valuable one, then it should be possible to study it in a variety of methodological contexts.

The Respondent Training Procedure was first reported by Smeets, Leader and Barnes (1997). The methodology comprises two parts -- respondent-type training and an equivalence test (Leader, Barnes-Holmes, & Smeets, 2000; Leader, & Barnes-Holmes, 2001). During respondent-type training, participants are exposed to up to six stimulus pairs presented with 0.5s inter-stimulus intervals and 3s. inter-pair intervals between each pair, that is 0.5s between pair members (e.g. AB) and 3s between pairs (e.g. between AB and BC). In the subsequent equivalence test, participants are presented with a three-choice MTS procedure that assesses the derived symmetry and equivalence relations based on the training. What is novel about respondent training, relative to MTS, is that it does not provide explicit, differential reinforcement for selecting any of the stimuli (as is the case with the standard conditional discrimination training). The methodology was, in fact, the first to demonstrate reliable equivalence responding without an experimental history of explicit differential reinforcement for MTS responding or an experimental history of successful equivalence testing.

The Precursor to the Relational Evaluation Procedure (pREP) is based on the Go/No Go Procedure (e.g. D'Amato, & Colombo, 1985). Other similar preparations include go left/go right, yes/no, or same/different (D'Amato, & Worsham, 1974; Edwards, Jagielo, Zentall, & Hogan, 1982) procedures. The pREP was developed as a

methodology that is independent of the MTS paradigm in the training stage, but does rely on MTS in the testing phase. The pREP involves the presentation of two stimuli on each trial -- one conditional stimulus (CS) and one discriminative stimulus (Sd). On trials that combine a CS with a positive Sd, responding (for example, pressing the space bar) is reinforced, but not on trials that combine a CS with a negative Sd (in this case it is appropriate not to make any response and to wait for the next trial). That is, reinforcement is provided for responding to target relations (e.g. A1->B1 and A2->B2), but not for non-target relations (e.g. A1->B2 and A2->B1). Cullinan, Barnes and Smeets (1998) reported that although the pREP was as effective as MTS in producing symmetry, it was relatively ineffective in producing equivalence responding, where participants had not been previously exposed to a history of MTS training and testing (see also Cullinan, Barnes-Holmes, & Smeets, 2000).

The Relational Evaluation Procedure (REP) was an extension of the pREP work and an attempt to rectify some of the weak equivalence performances that had been observed. Specifically, pREP researchers recognised the need to identify possible sources of contextual control that could be incorporated into the procedure (Cullinan, Barnes-Holmes, & Smeets, 2001). This involved replacing the "press" and "no press" responses on the space-bar with contextual cues for relations of "same" and "different". In order for these functions to be established, participants were first trained, across a number of exemplars, to choose one arbitrary stimulus in the presence of identical stimuli (thereby establishing a "same" function for the arbitrary stimulus) and a second arbitrary stimulus in the presence of physically different stimuli (thereby establishing a "different" function for the second arbitrary stimulus).

Once the "same" and "different" functions had been established for the two arbitrary stimuli, two response keys were assigned to these stimuli for subsequent

training and testing using the same procedure as the pREP. That is, participants pressed the "same" function key on CS-positive/Sd trials and pressed the "different" function key on CS-negative/Sd trials. Cullinan et al. (2001) reported found that the use of the contextual cues facilitated both symmetry and equivalence responding across participants.

Relational Frames versus Stimulus Classes

An alternative traditional behavioural view has long proposed that stimulus relations and relational networks can be more readily accounted for by the established concept of *stimulus classes* (Donahoe, & Palmer, 1994). The core difference between this and RFT is that the former views stimulus classes as the central and most complex unit of derived relational responding, whereas multiple stimulus relations lie at the heart of RFT. This controversy has arisen, in part, from the fact that class formation can be understood as both a product and a process (Hayes et al., 2001). That is, the process of deriving equivalence relations always gives rise to the products -- reflexivity, symmetry and transitivity. Furthermore, stimulus generalisation is considered the basic behavioural process that permits the formation of stimulus classes. The classes of stimuli that emerge via the process of stimulus generalisation are thus defined as part of the product of that process (Hayes et al., 2001). However, when class formation is also taken to be a process, it is difficult to provide an explanation for its occurrence (Hayes, & Barnes, 1997).

Perhaps the strongest argument against the use of the concept of stimulus classes in this context lies in the empirical evidence, especially from RFT research. For example, numerous studies have demonstrated contextually controlled, arbitrary responding in accordance with multiple stimulus relations such as same, different,

opposite and comparison. Indeed, extraordinarily complex patterns of responding can result from the simplest form of training. Consider the research by Steele and Hayes (1991) as a case in point. It is difficult to see how the concept of stimulus class can account for the complex and interconnected array of derived stimulus relations that emerged from the very basic establishment of "same" and "opposite" relations (see also Wulfert, & Hayes, 1988).

Relational Frames versus Associations

Another behavioural approach to stimulus relations hails from the more traditional concept of the *association.* Although much work here (mainly with non-humans) pre-dates RFT and even stimulus equivalence, the concept of the association has in a sense been re-invented by social psychologists to account for matching performances that bear a strong resemblance to work on derived relations. The dominant methodology in this area is the Implicit Association Test (IAT) which has been repeatedly used to demonstrate strong and predictable patterns of stimulus matching across a whole array of psychological phenomena. These include: race (Dunham, Baron, & Banaji, 2006), ethnicity (Rudman, & Ashmore, 2007), self-esteem (Yamaguchi, Greenwald, Banaji, Murakami, Chen et al., 2007) and even eating habits (Grover, Keel, & Mitchell, 2003).

From this perspective, the concept of the association appears to describe the process by which *implicit cognitions* (also referred to as *implicit attitudes*) emerge (Greenwald, 1990). These are defined as "introspectively unidentified or inaccurately identified traces of past experience that mediate favourable or unfavourable feeling, thought, or action toward social objects" (Greenwald, & Banaji, 1995, p.8). Like the concept of stimulus relations, individuals are often not aware that the relevant

associations have been formed, or that they can manifest as judgements or overt actions. As a result, implicit cognitions do not require conscious effort and appear to be initiated by stimulus cues in the environment (Shiffrin, & Dumais, 1981). Indeed, implicit attitudes are believed to be inescapable (Neely, 1977). In contrast, explicit attitudes are intentional and require the active attention of the listener. They are more flexible than automatic processes and this makes them particularly useful for decision-making, problem-solving and the initiation of new behaviour (Devine, 1989).

Social psychologists have devoted considerable effort towards the development of robust methodologies for the study of implicit attitudes. Those available to date include: the Go/No-Go Association Task (GNAT: Nosek, & Banaji, 2001); the Extrinsic Effective Simon Task (EAST: de Houwer, 2003); word fragment completion tasks (Gilbert, & Hixon 1991); name-letter preference tasks (Koole**,** Dijksterhuis, & van Knippenberg, 2001); and the original spatial Simon task (Simon, 1990). However, by far the most well established is the Implicit Association Task (IAT: Greenwald, McGhee, & Schwartz, 1998).

The Implicit Association Test (IAT)

The IAT has been successfully employed to measure implicit attitudes toward a broad range of phenomena, including cognitions that are particularly susceptible to concealment or manipulation on explicit measures. This array includes: race (Greenwald, Oakes, & Hoffman, 2003); gender (Greenwald, & Farnham, 2000) dysfunctional beliefs (de Jong, Pasman, Kindt, & van den Hout, 2001), eating disorders (Grover, Keel, & Mitchell, 2003); and anxiety (Egloff, & Schmukle, 2002).

The IAT is based on the assumption that it should be easier to map two concepts onto a single response when those concepts are associated in memory than when they are unrelated (Farnham, Greenwald, & Banaji, 1999). It is a latency-based measure that compares the speed of the congruent versus incongruent associations. Specifically, it predicts that participants will require less time to categorise an attribute and concept strongly associated in memory than an attribute and concept weakly associated in memory. According to Greenwald et al. (1998), the IAT measures the strength of associations between concepts and then infers the nature or directionality of the associations on this basis. The procedure is considered to be an implicit measure because participants are required to respond in less time than is needed to make a conscious response.

Consider a typical IAT trial from Greenwald et al. (1998). The target concept discriminations were flowers (e.g. ROSE and TULIP) versus insects (e.g. BEE and WASP) that were combined with pleasant (HAPPY, PEACE) and unpleasant attributes (ROTTEN, UGLY). On a trial (referred to as *consistent*), participants were presented with INSECT and UNPLEASANT on the top left of the screen with FLOWERS and PLEASANT on the top right. In the centre the target word may have been FLEA. In this case, it was relatively easy for participants to emit a left-hand key press to indicate that FLEA was associated with the INSECT-UNPLEASANT combination, rather than FLOWERS-PLEASANT. On another trial however (referred to as *inconsistent*), the concept-attribute pairings may have been alternated such that INSECT was now combined with PLEASANT and FLOWERS with UNPLEASANT. In this alternative case, it was likely more difficult for participants to associate FLEA with either INSECT-UNPLEASANT or FLOWERS-PLEASANT. The simple logic is that the FLOWERS-PLEASANT/INSECT-UNPLEASANT associations are already

better established than FLOWERS-UNPLEASANT/INSECT-PLEASANT. The predicted superiority of responding on consistent trials over inconsistent trials is known as *the IAT effect.*

Greenwald and colleagues have published a range of studies showing that the IAT appears to reveal implicit attitudes that participants typically deny on explicit measures. For example, participants who claimed not to hold racist attitudes on explicit measures more readily categorised names typical of White persons with positive words and names typical of Black persons with negative words, than vice versa on the IAT (see Greenwald, Banaji, Rudman, Farnham, Nosek et al., 2002, for a review). The basic IAT effect has been replicated many times, and it has become an increasingly popular method for assessing implicit attitudes, particularly to socially sensitive phenomena (e.g. de Houwer, 2002).

There is little to doubt about the robustness of IAT effects or the utility of the procedure in a broad array of phenomena. But it remains the case, at a conceptual level, that the methodology relies on the basic idea that associations are central to higher cognitive functioning. But, his is problematic for RFT researchers and others. Consider the following paragraph by de Houwer (2002, pp. 117-118):

> Greenwald et al. (1998) designed the IAT to assess the strength of associations between concepts in memory. One can argue that beliefs involve more than just associations between concepts. First, beliefs reflect qualified associations. For instance, the belief "I am a bad person" implies *a special type of association* between the concept "self" and the concept "bad", namely a *directional association* which specifies that "bad" is a property or characteristic of "self". IAT effects do not reflect the *nature of directionality of an association between concepts,* they can reflect only strength of association. Second, many beliefs involve several associations and several concepts. For instance, conditional beliefs such as "if I do not perform well on a task, then I am an inferior person" involve rather complex structures of qualified associations between several concepts. The IAT cannot be used to directly capture such complex conditional beliefs (also see de Jong et al., 2001, p.111). In sum, the IAT does not provide a measure of beliefs, nor was it designed to do so. It can only provide an index of associations that are assumed to be involved in certain beliefs and thus indirect evidence for the presence of certain beliefs

Put simply, the IAT cannot readily assess the nature or directionality of associations.

In one sense, this limitation is not of central concern if one considers that human cognition is a largely associative set of processes (see Greenwald, Nosek, Banaji, & Klauer, in press for a relevant discussion). But, if this is not the case then one begins to wonder what it is that the IAT is measuring. Researchers in RFT have recently developed an alternative measure of implicit cognitions, which, not surprisingly, has stimulus relations at its core and which appears to go some way towards addressing de Houwer's concerns.

The Implicit Relational Assessment Procedure (IRAP)

The IRAP is an automated latency-based measure that also juxtaposes consistent and inconsistent responding. But there are a number of key procedural differences with the IAT. First, all IRAP trials are identical in format, so the complexity of the target relations is not increased across blocks. That is, one of two attributes is always presented as a sample stimulus and one member from two sets of target concepts is always the target stimulus. Hence, consistent and inconsistent trials do not differ in format. Second, in order to capture the directionality or specificity of the target relations, the response options are directly paired with specific relational terms (e.g. SIMILAR and OPPOSITE).

Consider a trial from the research by Barnes-Holmes, Barnes-Holmes, Power, Hayden, Milne et al. (2006). On each trial, either PLEASANT or UNPLEASANT appeared as a sample stimulus at the top of the screen. The target stimulus in the middle of the screen comprised of a word from one of two concept categories that could be readily evaluated as positive (e.g. LOVE and HOLIDAY) or negative (CANCER and JAIL). The relational response options in the case were SIMILAR and

OPPOSITE. In short, participants were required to indicate that the relationship between PLEASANT and LOVE is one of co-ordination (by choosing SIMILAR) or opposition (by choosing OPPOSITE). In line with the IAT, the IRAP predicts that already-established relations will result in shorter response times than novel relations. For example, participants should take longer to relate UNPLEASANT-LOVE-SIMILAR than PLEASANT-LOVE-SIMILAR. IRAP effects have already been successfully established across an array of psychological phenomena, including attitudes to homosexuality (Cullen, & Barnes-Holmes, 2008), meat-eating (Barnes-Holmes, Murtagh, Barnes-Holmes, & Stewart, in press), self-esteem in children (Scanlon, 2008) and even attitudes to sexual offending (Dawson, Barnes-Holmes, Gresswell, Hart, & Gore, 2009).

One of the first published IRAP studies compared the explicit and implicit attitudes of professionals with varying degrees of experience working with individuals with Autistic Spectrum Disorder (ASD -- Barnes-Holmes et al., 2006). Specifically, one group of participants had between six months and six years experience; another group had less than six months experience; and the remainder had never worked professionally with this population. The IRAP employed the phrases AUTISTIC SPECTRUM DISORDER and NORMALLY DEVELOPING as sample stimuli, with positively or negatively evaluated target stimuli (e.g. CALM or DIFFICULT). During consistent blocks, participants were required to relate, for example: NORMALLY DEVELOPING-CALM-SIMILAR; NORMALLY DEVELOPING-DIFFICULT-OPPOSITE; AUTISTIC SPECTRUM DISORDER-CALM-OPPOSITE; and AUTISTIC SPECTRUM DISORDER-DIFFICULT-SIMILAR, thus assessing bias towards normally-developing children and against children with ASD. On inconsistent blocks, the expected patterns were reversed. In

order to compare outcomes on implicit and explicit measures, participants also completed the Challenging Behaviours Attribution Scale (CHABA: Hastings, 1997) and the Attitudes to Autism Scale (AAS).

The results from the study showed IRAP effects biased in favour of normally-developing and against ASD for all three groups. Although the group with the most professional experience showed the strongest effects, there was no statistically significant main effect for group. In contrast however, the two groups with experience showed significantly more positive attitudes to ASD than the group with no experience on the explicit measures. In short, the professionals appeared to understate their bias against this population on the explicit measure, but all was revealed on the IRAP.

In another study of implicit attitudes to nationality, the same researchers presented Irish students with a likeability hierarchy among Scottish, American and African groups. The researchers predicted that Irish participants would respond in accordance with Irish as most likeable, Scottish as next most likeable (probably because they were most similar), followed by American (less similar) and then African (least similar). In other words, participants were required to choose between Irish over Scottish, Scottish over American and American over African as more or less likeable. In this case, the response options were TRUE and FALSE. Participants were also presented with a series of explicit measures of their attitudes to these nationalities.

Consistent with experimental predictions, participants showed implicit bias in terms of Irish as more likeable than Scottish, Scottish more likeable than American and American more likeable than African. In contrast however, attitudes on the explicit measure were either mixed or opposite (e.g. Irish and Scottish were preferred

equally and Africans were preferred over Americans). Once again, therefore, there was clear divergence between the implicit and explicit measures that showed more politically correct responding on the explicit measure relative to the IRAP.

The Current Thesis

The aims of the current thesis were three-fold. The first was primarily methodological and attempted to determine the stimulus presentation that would produce the strongest IRAP effect. Experiment 1 set out to address this simple experimental question. The key manipulation involved the presence or absence of the randomisation of the sample stimuli (with or without the randomisation of the response options) generated the strongest $D\text{-}_{IRAP}$ scores. Randomising the response options alone without randomising the samples also produced a significant IRAP effect, although this was smaller. The $D\text{-}_{IRAP}$ score that resulted from the static presentation of both samples and response options was almost negligible.

The second aim was to determine the utility of the IRAP as a measure of implicit attitudes to race and this was the core focus of Experiments 2 to 6. In the IRAP in Experiment 2, participants were simply required to relate the phrases WHITE PERSON and BLACK PERSON to a range of simple positive and negative terms that could be readily categorised as safe and dangerous. We predicted that out Irish participants would show a pro- White and anti- Black bias. The $D\text{-}_{IRAP}$ trial-type scores were in part consistent with experimental predictions and showed a pro- White bias where White was safe, but not an anti-Black bias. In fact, participants significantly confirmed that Black was safe. This coincided with the explicit measures on which participants indicated the absence of racial bias against Black people.

In line with existing IAT evidence, to determine whether racial pictures would generate stronger IRAP effects than words, and to examine the potential role of attentional weapon bias, the IRAP in *Experiment 3* presented pictures of Black and White men holding guns. For correct responding, however, the presence of the guns was incidental and the discriminations among the pictures were based on race alone. We predicted that Irish participants would show a pro- White and anti- Black bias that was perhaps enhanced by the presence of the guns. Participants significantly confirmed that White was safe and Black was dangerous, although they also significantly confirmed that Black as safe. Hence, this was our first evidence of an anti- Black bias, but other effects were not entirely as expected. Again on the majority of explicit measures, participants indicated the absence of racial bias.

In order to determine whether the anti-Black bias in the previous study did reflect the influence of the guns, Experiment 4 presented pictures of Black and White men holding mundane objects. We predicted that Irish participants would show a pro-White and anti- Black bias, even though both were holding mundane objects. Participants significantly confirmed that both White and Black were safe, but had mixed views of either as dangerous. Hence, the anti- Black bias from the previous study was now lost, suggesting that the weapons had facilitated the effect. Again, on the majority of the explicit measures, participants indicated the absence of racial bias.

In line with previous IAT studies of race and to examine further the outcomes from the previous studies, *Experiment 5* presented pictures of Black and White men holding *both* guns and mundane objects. We predicted that Irish participants would show a pro- White and anti- Black bias that may be differentially influenced by the items presented in the pictures. The four trial-types examined in the previous studies were sub-divided into eight to permit comparisons of the relative influence of the

different items held in the pictures. Overall, seven of the eight trial-types were non-significant and participants only significantly confirmed that White men with mundane objects were safe. In fact, close inspection of the non-significant effects indicated unexpected evidence of pro- Black bias. Again these agreed with the explicit measures, on which participants indicated the absence of racial bias.

In Experiment 6, participants were presented with the same IRAP as the previous study, but were required to discriminate the objects in the pictures as the basis for correct responding and ignore race. That is, on consistent trials guns were always dangerous and mundane objects were always safe and correct responding on inconsistent trials was reverse (guns always safe and mundane objects dangerous). We predicted that Irish participants would categorise guns as dangerous and mundane objects as safe, but expected that these effects might be influenced by race. In this study, six of the eight were non-significant. Participants only significantly confirmed that Black with a gun was not safe and White with a gun was dangerous, but there was no evidence of clear influence of either items or race in the remaining effects. Again, the explicit measures indicated the absence of racial bias.

The third aim of the current thesis was to examine the utility of the IRAP as a measure of clinically relevant implicit cognitions. Previous preliminary research had indicated strong IRAP effects that showed a positive bias towards the self in participants with a high explicit self-esteem. In *Experiment 7*, participants were required to indicate their agreement or disagreement with a range of statements that reflected high self-esteem (i.e. were positive about the self) or reflected low self-esteem (i.e. were negative about the self). Because the study was aslo concerned with whether on not the IRAP would correlate with an explicit measure of self, the positive and negative statements presented in the IRAP were taken directly from the Roseberg

Self-Esteem Scale (RSES) and particiants also completed this as an explicit measure. We predicted that participants high on explicit self-esteem would relate the self with the positive self-statements and not relate the self with negative self-statements.

The findings indicated that participants significantly confirmed that they agreed with the positive (high self-esteem) statements and significantly disconfirmed that they disagreed with them. They also disconfirmed that they agreed with the negaative statements, although this effect was non-significant. Although these findings were largely as expected from participants who all scored as high in self-esteem on the explicit RSES, there was no significant correlation between the two measures.

In *Experiment 8,* we used the IRAP to explore implicit attitudes towards acceptance and avoidance. For this purpose, we selected undergraduate students who showed an xplicit propensity towards high acceptance/ low avoidance on the Acceptaance and Action Questionnaire (AAQ-2). In order to enhance the potential overlap between the two types of measure, as we had done in Experiment 7, we took opposing satements directly from the AAQ and inserted them into the IRAP as target stimuli. The results indicated that participants significantly confirmed that they agreed with the acceptance statements, but all other effects were non-significant. Closer inspection of the trial-types, however, were as expected and showed that particpants disconfirmed that they agreed with the avoidance statements and confirmed that they disagreed with them. In spite of these outcomes, the IRAP data did not correlate significantly with the AAQ.

The current reaearch program offered preliminary investigations of the utility of the IRAP as a measure of impicit attitudes to a range of psychological phenomena, including race, self-esteem and acceptance. Although the IRAP effects in many cases

were not as significant and predicted, particularly in the context of race, there wasgood evidence thata further investigation would be promising. The findings here make a useful contribution to the small existing IRAP research base by highlighting key issues that influence IRAP outcomes in a range of areas.

Chapter 2

Exploring IRAP Screen Presentations

Chapter 2

Exploring IRAP Screen Presentations

Experiment 1

There are numerous experimental features of the IAT, and even the IRAP, that raise questions for behavioural psychologists. For example, would it be better to have the sample stimuli in the top centre of the screen, rather than at either corner? Or are simple left/right key presses the most effective means of facilitating the emission of appropriate responding? Because it is important for implicit measures to ensure that there is not adequate time for conscious responding, participants are all the more dependent on specific features of the stimulus presentation. Indeed, these questions are perhaps all the more important for the IRAP, where all trials are presented in the same format. Practically all available IRAP studies to date have employed a presentation format in which both the sample stimuli and response options randomly switch positions across trials within each block. However, there is no empirical evidence to suggest the superiority of this format over any others (e.g. keeping either of these features static).

Prior to examining the utility of the IRAP in an ever-increasing area of psychological phenomena, Experiment 1 of the current research set out to address this simple experimental question. Its main aim was to systematically manipulate the presentation of stimuli on-screen to determine what influence this might have on responding. Naturally, a simple IRAP, rather than one on race or nationality for example, was chosen for this purpose. Although we made no firm predictions about which format would produce the best IRAP effect, we assumed that the commonly

used format in which both the individual sample stimuli and response options switch positions would be amongst the most effective.

Participants

A total of sixty undergraduate students of the National University of Ireland, Maynooth (NUIM) participated in Experiment 1. All were aged between 18 and 26 years old (*M*: 22 years and 11 months) and were recruited though notice boards and classroom announcements in the Department of Psychology. All participants presented with high levels of fluency in English and with normal or corrected to normal vision. Each participant was randomly allocated to one of four experimental conditions (12 participants in each condition), across which the on-screen stimulus arrangements were manipulated. No participants had previous experience of the IRAP and none received any remuneration for participating in the study. The data from twelve participants was removed from the analyses because these individuals failed to reach criterion during the practice blocks, leaving data from forty-eight participants.

Settings

The current study was conducted in the Computer Laboratory in the Department of Psychology at NUIM. The room was generally free from noise and distraction. All participants conducted the study on an individual basis and only one was present in the Laboratory at any one time. In all cases, the Experimenter remained present in the Laboratory throughout participation and interacted directly with participants only during instructional, but not test, phases.

Apparatus and Materials

Experiment 1 involved two basic sets of materials. The *explicit* measures comprised of Likert scales and the *implicit* measure comprised of the IRAP.

In the *Likert scales,* participants were presented with twelve 13-point scales, one for each of the target words that would subsequently be presented in the IRAP. Participants rated each word from +6/Pleasant to -6/Unpleasant -- see Appendix A.

The IRAP. Participants completed all IRAP trials on a DELL desktop computer with a Pentium 4 processor. The procedure was delivered via a program written in Visual Basic (Version 6.0.) that controlled all aspects of stimulus presentation and the recording of all participant responses. There were minor software modifications that were necessary to distinguish the IRAP presentations that appeared during each of the four conditions (see below).

The stimuli presented during the IRAP comprised of three groups of words. The sample stimuli contained the words PLEASANT and UNPLEASANT. The target stimuli contained twelve evaluative words -- six positive (CARESS, FREEDOM, HEALTH, LOVE, PEACE and CHEER) and six negative (ABUSE, CRASH, FILTH, MURDER, SICKNESS and ACCIDENT). Both of these word sets were taken from Greenwald et al. (1998). The response options contained the relational terms SIMILAR and OPPOSITE. The stimulus arrangements employed in Experiment 1 are presented in Table 1.

Table 1
The Stimulus Arrangements Employed in Experiment 1

Sample Stimulus 1 PLEASANT	**Sample Stimulus 2** UNPLEASEANT
Response Option Consistent with Sample 1 SIMILAR	**Response Option consistent with Sample 2** OPPOSITE
Targets Stimuli consistent with Sample 1	**Targets Stimuli consistent with Sample 2**
CARESS	ABUSE
FREEDOM	CRASH
HEALTH	FILTH
LOVE	MURDER
PEACE	SICKNESS
CHEER	ACCIDENT

Experimental Conditions

The four conditions in Experiment 1 varied only in terms of the randomisation algorithm that controlled the presentation of the sample stimuli and the response options (the target stimuli were always presented in the middle of the screen).

Moving-Moving Condition. The Moving-Moving Condition contained two features that were alternated across trials (i.e. they were moving). Specifically, the order of the presentation of the sample stimuli (PLEASANT and UNPLEASANT) at the top of the screen alternated in a quasi-random manner throughout each block of trials. In addition, the left-right positions of the response options (SIMILAR and OPPOSITE) at the bottom of the screen also alternated in a quasi-random manner. Hence, the label Moving-Moving refers to the simultaneous randomisation of both samples (Moving) and response options (Moving, respectively).

Moving-Static Condition. In this condition, the presentation of the samples was randomised as above, but the locations of the response options remained static or identical across all trials (hence the term Moving-Static). The actual presentation of each response option on either side of the screen was counterbalanced across

participants (i.e. for half, SIMILAR was always on the left and OPPOSITE on the right, with the reverse presented to the remaining half).

Static-Moving Condition. In this condition, the sample stimuli remained static such that one appeared on the first twelve trials of each block and the second sample was then presented in the remaining twelve trials. The sequencing of the two samples was also counterbalanced across participants (i.e. half saw PLEASANT first, then UNPLEASANT, and this was reversed for the rest). The response options alternated in a quasi-random manner within blocks as before.

Static-Static Condition. In this condition, the sample stimuli were static (i.e. their sequence was not random) and the response options were always in the same fixed locations (hence, the label Static-Static). Again, the sequencing of the individual sample stimuli and the response options were counterbalanced across participants.

General Procedure

All participants completed the study in a single experimental session that lasted between 30 and 40mins. in total. Although made available, none of the participants opted for short breaks throughout the study.

Procedure

Verbal Instructions and Visual Aids. Because pilot studies involving the IRAP had previously noted difficulties in ensuring that participants understood clearly how to respond appropriately to IRAP trials (Barnes-Holmes et al., 2006), the current experiment involved the presentation of a series of visual aids or screen shots of each of the four basic trial-types (see Figure 1 for an example).

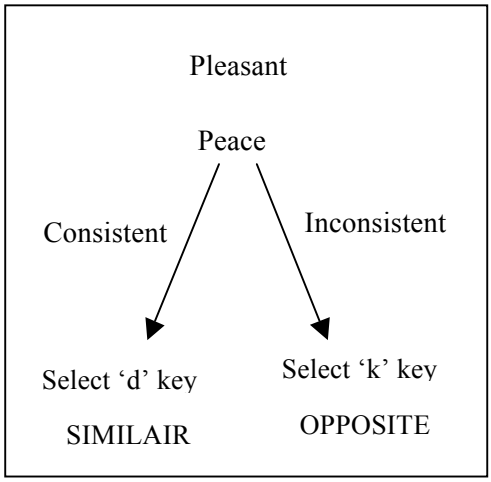


Figure 1. A visual aid containing a schematic representation of an IRAP trial.

In all cases, the screen shots were accompanied by additional verbal instructions along the following lines:

This is a word-matching task that involves you having to match certain words together. Sometimes the computer will want you to match words in a way you agree with and sometimes it will want you to match words in a way you do not agree with. This is part of the experiment. For example, sometimes the computer might ask you to match the word PLEASANT with this set of words (participant is shown the six positive target stimuli), and sometimes it will want you to match the word UNPLEASANT to this set of words (participant is shown the six negative target stimuli). So, you might have to match PLEASANT with PEACE and UNPLEASANT with CRASH. Then you will have a break and the computer will let you know how you are doing.

Then the computer will change and it will want you to match the word PLEASANT with this set of words (participant is shown the six negative target stimuli) and the word UNPLEASANT with this set of words (participant is shown the six positive target stimuli). So, now you might have to match PLEASANT with CRASH and UNPLEASANT with PEACE. A message will appear on-screen when the computer wants you to change the way you are responding.

In order to match the words in this game you have to press either the ‘d’ or ‘k’ button on the keyboard. They are the only two keys you have to press.

Now, there are two other words involved in this task -- OPPOSITE and SIMILAR. So if I tell you that the computer wants you to match the word PLEASANT to PEACE, FREEDOM, HEALTH, CARESS, LOVE and CHEER and you see this coming up on the screen (participant is presented with schematic representation of a consistent trial with PLEASANT) which key will you press? Will you press ‘d’ or ‘k’?

Ok, so what happens if UNPLEASANT and PEACE come up on the screen but I have told you that I want you to match PLEASANT to PEACE? What key would you press? Would you press ‘d’ or ‘k’?

If you press the wrong key during the experiment, a red X will appear on-screen. To continue, press the correct key. It is important that you go as fast as you can while still getting as many answers as possible correct.

At this point, minor variations were adopted in the verbal instructions provided to participants in each of the four conditions. Consider the following employed in the Moving-Moving Condition:

> In this task the word at the top of the screen will randomly change between PLEASANT and UNPLEASANT. The words OPPOSITE and SIMILAR will also switch places randomly during the experiment. Sometimes the word OPPOSITE is on the right-hand side of the screen and SIMILAR is on the left-hand side of the screen (participant is shown schematic representation with OPPOSITE on the right and SIMILAR on the left) and sometimes it is the reverse (participant is shown schematic representation with SIMILAR on the right and OPPOSITE on the left). So in this task you have to keep your eye on the word at the top (i.e. PLEASANT or UPLEASANT), the word in the middle (e.g. PEACE) and the places in which OPPOSITE and SIMILAR appear on the screen. Do you understand?

Automated Instructions. In addition to the verbal instructions and the screen shots, the automated procedure commenced with a series of more detailed instructions for understanding and completing IRAP trials as follows:

> Our research investigates cognitive processes that are used in decisions that involve memory. We are seeking to develop and test theories of cognitive processes that occur inside and outside of awareness in the routine use of memory.
>
> Stimuli will be presented on this display screen and your responses will be entered on the keyboard.
>
> The research assumes that you can read English fluently, and that your vision is normal or corrected to normal. If you do not consider yourself fluent in English, or if your vision is not normal or corrected to normal, and ESPECIALLY IF YOU ARE HAVING SOME DIFFICULTY READING THIS DESCRIPTION, PLEASE ask the Experimenter now whether or not you should continue.
>
> Your identity as a participant is confidential. Further, you are free to discontinue participation at any time, without penalty.
>
> In keeping with standard practice, your data may be retained for 5 years or so, during which time only the investigators on this or successor projects will have access to them.
>
> PLEASE NOW READ THE STATEMENT BELOW, WHERE YOU WILL BE ASKED TO RESPOND TO A STANDARD INFORMED CONSENT QUESTION.
>
> **CONSENT STATEMENT**
>
> I have read the description of the procedure. I understand that the questions I may have about this research will be answered by Professor Barnes-Holmes or one of the other researchers working on this project.
>
> If you consent to participate in the research that has been described on the preceding display pages you should now read the Instructions for the sorting tasks below.

[INSTRUCTION: If you wish to ask any questions first, alert the experimenter now. IF YOU WISH NOT TO PROCEED, you should inform the experimenter].

INSTRUCTIONS

Shown below are illustrations of the four different types of task *[only one illustration is presented here]* that will be presented repeatedly in this part of the experiment. To help you understand the tasks each of the four illustrations is explained immediately underneath. Please examine each illustration and then read carefully the explanation attached to it. Please make sure that you understand each task before continuing with the experiment.

IMPORTANT: FROM TRIAL TO TRIAL THE POSITIONING OF THE RESPONSE OPTIONS (SIMILAR AND OPPOSITE) WILL VARY RANDOMLY BETWEEN LEFT AND RIGHT *[THIS WAS ONLY PRESENTED, AS APPROPRIATE]*.

Illustration 1

Pleasant

Crash

Select 'd' for **Similar** | Select 'k' for **Opposite**

Explanation for Illustration 1

If you select "Similar" by pressing the 'd' key, you are stating that "Crash IS Pleasant."

If you select "Opposite" by pressing the 'k' key, you are stating that "Crash IS NOT Pleasant."

NOTE: DURING THE EXPERIMENT A RANGE OF OTHER WORDS WILL BE PRESENTED, AS WELL AS THE ONES USED IN THE EXAMPLES.

REMEMBER: FROM TRIAL TO TRIAL THE POSITIONING OF THE RESPONSE OPTIONS (SIMILAR AND OPPOSITE) WILL VARY RANDOMLY BETWEEN LEFT AND RIGHT *[AS APPROPRIATE]*.

FINAL INSTRUCTIONS

During the experiment you will be asked to respond as **quickly and accurately** as you can across all trials.

The relating tasks will be presented in short ***sessions*** that are separated by the appearance of instructions on the computer screen. You can take a short break if you like while the instructions are on-screen.

During each short session the relating task follows one general rule. An incorrect response on any trial is signalled by the appearance of a red 'X' in the centre of the

> screen. To remove the red 'X' and move on to the next trial, please press the correct response key quickly.
>
> After each session, further instructions will appear and they will tell you that the general rule that applied in the previous session is now completely reversed. Please pay close attention to these instructions and do your best to follow them.
>
> So, just to clarify, there will be only ***two*** general relating rules, and so the first thing you should do at the beginning of each session is to discover the rule by using the feedback you get in the form of the red 'X'.
>
> It is very important to understand that sometimes you will be required to respond to the tasks in a way that ***agrees*** with what you believe and at other times you will be required to respond in a way that ***disagrees*** with what you believe. **This is part of the experiment**.
>
> The first two sessions are for ***practice*** only and these are repeated until you respond accurately on at least 70% of the relating trials, and respond faster, on average, than 3000 milliseconds (i.e. 3 seconds). When you complete the practice phase, the ***test-phase*** will then start. Remember, you should try to make your responses as **accurately and quickly** as possible.
>
> Good Luck. If you do not understand something about the foregoing instructions or have any further questions please talk to the researcher before clicking on the blue button.

Feedback. Although IRAP trials are generally referred to as test trials (as with the IAT), all incorrect responses are consequated with automated written corrective feedback and a correction procedure. That is, after an incorrect response was emitted the stimuli remained on screen and a red 'X' appeared directly below the target word. The 'X' remained until the participant emitted a correct response and the next trial then appeared automatically. No feedback consequated responding.

IRAP Trials. All blocks of IRAP trials (i.e. both practice and test blocks) were identical in format, and practice trials were always completed first. The number of practice blocks that participants received ranged from two to eight and was contingent upon their performances therein. Once participants had reached the mastery criteria (in terms of both accuracy and response latency) they preceded immediately to the first test block. All participants were exposed to a total of six test blocks. At the end of each block of trials, the IRAP presented participants with automated feedback on

the percentage of trials correct and the median response time (in ms.) achieved during that block.

Trial-types. The IRAP comprised of four basic trial-types that were constructed from the configurations of the sample stimuli (PLEASANT and UNPLEASANT) and the two types of target stimuli (positive and negative words). For experimental purposes, these were referred to as: PLEASANT/Positive (top left of Figure 2); PLEASANT/Negative (top right); UNPLEASANT/Positive (bottom left); and UNPLEASANT/Negative (bottom right). Each block of IRAP trials contained six exposures to each of the four trial-types randomly presented across the block (i.e. a total of 24 trials in each block). On each of the six exposures to a trial-type, each of the three target stimuli appeared twice with the same sample.

Consistent and Inconsistent Responding. Each block of IRAP trials was designated, for experimental purposes, as consistent or inconsistent. Participants were exposed to a minimum of two practice blocks (one consistent and one inconsistent) and six test blocks (three consistent and three inconsistent). The IRAP sequence was always presented as alternating blocks of consistent and inconsistent trials. As a result, participants were required to switch their patterns of correct responding across blocks (i.e. the contingencies were reversed). In order to control for potential order effects, the sequencing of the blocks was counterbalanced across participants. That is, half of the participants were presented with a consistent practice block first, followed by the inconsistent practice block, followed by a consistent test block and so on. In contrast, the other half were presented with an inconsistent practice block first, and so on.

The recording of a response on any trial as correct or incorrect depended on whether the trial had been categorised as consistent or inconsistent. Trials were

referred to as *consistent* when the relations among the sample and target stimuli were consistent with the views believed to be held by participants prior to the study. That is, on these trials participants should more readily relate positive evaluations with PLEASANT, and the negative (rather than positive) evaluations with UNPLEASANT. Consider the trials presented in Figure 2. Correct responses here required participants to respond PLEASANT/PEACE/SIMILAR; PLEASANT/CRASH/OPPOSITE; UNPLEASANT/PEACE/SIMILAR; and UNPLEASANT/CRASH/OPPOSITE.

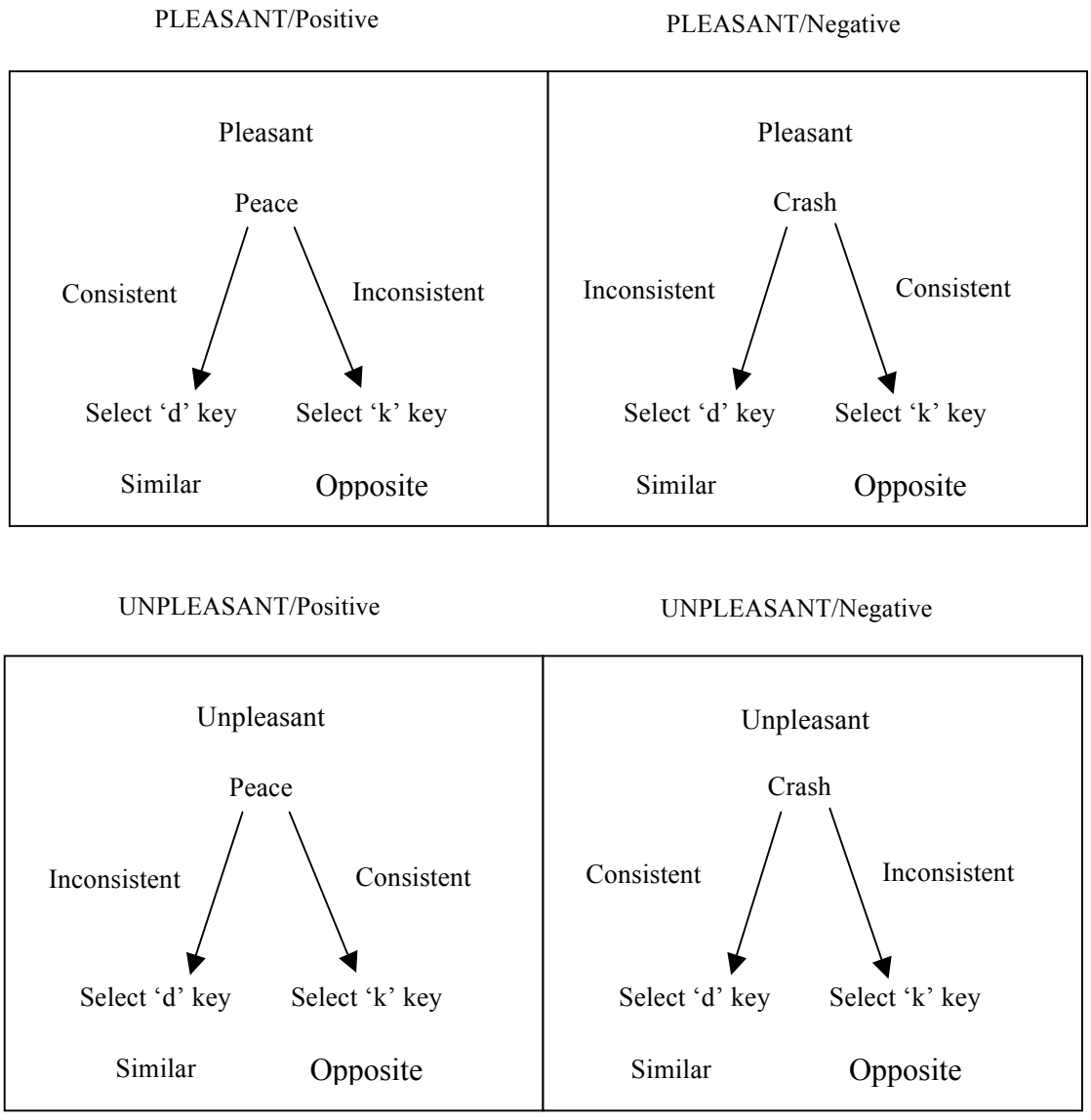


Figure 2. A schematic representation of the four basic trial-types in Experiment 1.

Trials were referred to as *inconsistent* when the relations among the sample and target stimuli were *not* consistent with views attributed to the participants. Consider again the trials in Figure 2. On inconsistent trials correct responses involved selecting PLEASANT/PEACE/OPPOSITE; PLEASANT/CRASH/SIMILAR; UNPLEASANT/PEACE/OPPOSITE; and UNPLEASANT/CRASH/SIMILAR.

The end of the sixth test block marked the end of the experiment for all participants. At this point, all were debriefed and thanked for their participation.

RESULTS

The primary datum in the IRAP was response latency, defined as the time (in ms.) that elapsed between the onset of the trial and a correct response. Although accuracy was also recorded on every trial, the accuracy data were simply employed as a screening mechanism to ensure that all data contained within the subsequent analyses comprised of scores greater than 70% accurate. All others were removed from the analyses.

In line with previous analyses of IRAP data, the response latency data for each participant were transformed into D-$_{IRAP}$ scores (see Cullen, & Barnes-Holmes, in press) using an adaptation of the D-algorithm by Greenwald, Nosek and Banaji (2003). The steps involved in calculating the D-$_{IRAP}$ scores were as follows: (i) only response-latency data from test-blocks were used; (ii) latencies above 10,000 ms. were eliminated from the dataset; (iii) all data for participants were removed if they produced more than 10% of test block trials with latencies less than 300ms; (iv) 12 standard deviations for the four trial-types were computed: four for the response latencies from Test Blocks 1 and 2, four from the latencies from Test Blocks 3 and 4, and a further four from Test Blocks 5 and 6; (v) 24 mean latencies for the four trial-

types in each test block were calculated; (vi) difference scores were calculated for each of the four trial-types, for each pair of test blocks, by subtracting the mean latency of the consistent block from the mean latency of the corresponding inconsistent block; (vii) each difference score was divided by its corresponding standard deviation from step 4, yielding 12 D-$_{IRAP}$ scores; one score for each trial-type for each pair of test blocks; (viii) four overall trial-type D-$_{IRAP}$ scores were calculated by averaging the three scores for each trial-type across the three pairs of test blocks; (ix) an overall D-$_{IRAP}$ score was calculated by averaging all 12 trial-type D-$_{IRAP}$ scores from step vii.

Responding in the predicted direction was indicated in the graphs by bars above a D-$_{IRAP}$ score of zero (positive D-$_{IRAP}$ scores) and responding that was contrary to prediction was indicated by bars below a D-$_{IRAP}$ score of zero (negative D-$_{IRAP}$ scores). For example, in the current study we predicted that participants would respond that positive words such as peace were pleasant and negative words like crash were unpleasant. As a result, in Figure 3, bars above zero, positive D-$_{IRAP}$ scores indicated that participants were responding in the predicted direction while bars below zero indicated that participants were responding contrary to prediction, for example responding that positive words such as peace were unpleasant and negative words like crash were pleasant.

Figure 3 presents the mean overall D-$_{IRAP}$ scores for each condition. All conditions generated D-$_{IRAP}$ scores that were in the predicted direction. The Moving-Static Condition showed the largest D-$_{IRAP}$ score, followed closely by Moving-Moving. The D-$_{IRAP}$ score for Static-Moving was considerably smaller and Static-Static was the lowest of all. These findings suggest that the largest D-$_{IRAP}$ scores are obtained when the sample stimulus is randomised; the D-$_{IRAP}$ score is smaller when

only the response options are randomised, and the effect becomes very small when both are static.

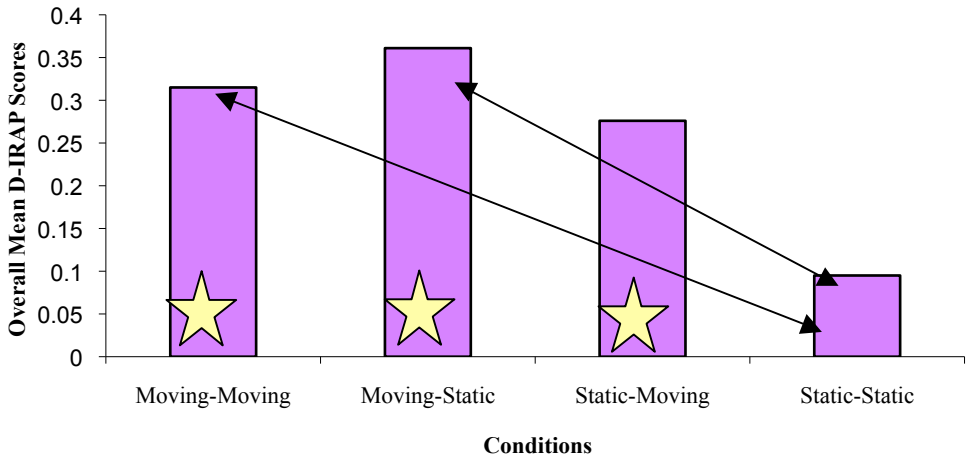


Figure 3. Mean overall D_{IRAP} score for each condition in Experiment 1.* Indicates that the D score was significant relative to zero. Arrow indicates conditions that were significantly different from each other.

A mixed repeated measures 4x2 Analysis of Variance (ANOVA) was conducted with condition and sequence (consistent-first vs. inconsistent-first) as between-participant variables. The analysis revealed a significant main effect for condition [$F (3, 40) = 2.839$, $p = .05$, $\eta_p^2 = .175$], but not for sequence ($p = .439$) and there was no significant interaction effect ($p = .634$). Post-hoc analyses (Fisher's) indicated that the significant differences among conditions were between Moving-Moving and Static-Static ($p = .023$) and between Moving-Static and Static-Static ($p = .011$), all other p's $> .11$.

Four one-sample t-tests were conducted to identify if the overall D_{IRAP} score for each condition differed significantly from zero. For three conditions, this was the case: Moving-Moving [$t (11)= 4.101$, $p = .002$]; Moving-Static [$t (11)= 4.307$, $p =$

.001]; and Static-Moving [t (11)= 3.035, p = .011]. The D-$_{IRAP}$ score for Static-Static did not differ significantly from zero (p = .126).

Split-half Correlations

To assess the internal consistency of the IRAP, a split-half reliability score was calculated. This was based on two D-$_{IRAP}$ scores, one for odd trials and one for even trials, obtained in the same way as the four trial-type D-$_{IRAP}$ scores, except that the algorithm described previously was applied separately to all odd trials and to all even trials. The split-half correlation between odd and even scores, applying Spearman-Brown corrections, proved to be strong and significant, r =.644, n = 47, p < .001. This indicated a reasonably strong level of internal consistency for the IRAP.

Explicit Measures

The Likert scales were designed to assess whether participants explicitly agreed with the experimental categorisations of the IRAP stimuli as positive or negative. The mean Likert ratings for the positive and negative words are presented in Table 2. Participants' evaluations of the two word groups were entirely consistent with the experimental categorisations.

Table 2
The Mean Likert Rating for Each IRAP Stimulus Presented in Experiment 1.

Positive Words	**Mean Rating**	**Negative Words**	**Mean Rating**
CARESS	3.9	ABUSE	-5.1
FREEDOM	5.2	CRASH	-4.6
HEALTH	4.8	FILTH	-4
LOVE	5.5	MURDER	-5.3
PEACE	4.9	SICKNESS	-4.3
CHEER	4.4	ACCIDENT	-3.6

DISCUSSION

The basic aim of Experiment 1 was to determine the type of screen presentation that would yield the strongest IRAP effect. The key manipulation involved the presence or absence of the randomisation of the sample stimuli and/or the response options. The results indicated that the randomisation of the sample stimuli (with or without the randomisation of the response options) generated the strongest $D\text{-}_{IRAP}$ scores. Randomising the response options alone without randomising the samples produced a much smaller effect. The $D\text{-}_{IRAP}$ score that resulted from the static presentation of both samples and response options was almost negligible.

Practically all available IRAP studies to date have employed the Moving-Moving presentation format, in spite of the lack of empirical evidence to suggest the superiority of this format over the others. Although the current work indicated that the strongest IRAP effects were observed with the randomisation of the sample stimuli but not response options (the Moving-Static Condition), the difference between this and the Moving-Moving condition was small and non-significant. This suggests, therefore, that the consistent use of the Moving-Moving format in IRAP research to date has been wise and has at least to some extent facilitated the strong IRAP effects reported therein. In line with these findings and the existing literature's use of this format, all future studies in the current thesis will employ the Moving-Moving presentation. In any case, the current results also support those reported previously in the observation of strong and predicted IRAP effects.

With this relatively simple methodological question addressed, we turned our attentions in the rest of the thesis to broadening out the range of applications of the IRAP to more social and clinical phenomena. Towards this aim, Experiments 2-6 presented in Chapter 3 employed the IRAP as a measure of implicit attitudes to race.

Chapter 3

Using the IRAP as a Measure of Implicit Attitudes to Race

Chapter 3
Using the IRAP as a Measure of Implicit Attitudes to Race
Experiments 2-6

Discrimination and segregation, on the basis of race, religion, etc., were formerly openly expressed and commonplace in Western society. But, the expression of such views and open actions based upon them are now socially unacceptable. As a result, there has been a considerable decline in the expression of discriminatory views (Schuman, Steeth, & Bobo, 1997). However, this trend may be more indicative of a change in the verbal culture, rather than decreases in the occurrence of overtly discriminatory practices. McConaghy (1986) attempted to address this issue with the development of an explicit self-report measure, known as the Modern Racism Scale (MRS), which attempted to tap into beliefs that perhaps societies now treat various social sub-groups with positive discrimination. One item, for example, implies that Black people 'have it too easy' or 'make life hard for themselves'. This scale has been shown to predict prejudiced behaviour in decision-making even when no overt racism was expressed (McConaghy, 1983). Other similar explicit measures include: the Motivation to Respond Without Prejudice Scale (MRWPS) which has been successfully used to identify the factors that may either promote or thwart prejudice reduction (Plant, & Devine, 1998) and the Discrimination and Diversity Scale (DDS) has been shown to correlate with racial prejudice on a priming task (Wittenbrink, Judd, & Park, 1997).

In spite of researchers' best efforts, however, the change in social climate has possibly meant that it is even more difficult to get questionnaire respondents to honestly report their opinions, particularly on sensitive issues such as race

(Greenwald, & Breckler, 1985; Lindeman, & Verkasalo, 1995). This has naturally set the stage for a potentially strong role for implicit measures, such as the IAT, in the study of socially sensitive phenomena. Indeed, many studies of race, for example, have employed a combination of both explicit and implicit methodologies, and have used this as a forum for determining the nature of correlation between the two.

Using Implicit Measures to Study Race

Numerous studies have used implicit measures to examine the implicit in-group bias of White participants (Amodio, & Devine, 2006; McConnell, Rydell, Strain, & Mackie, 2008). At its simplest, participants have shown a significant implicit preference for White people over Black people even when the effect for implicit colour preference was controlled (Smith-McAllen, Johnson, Dovidio, & Pearson, 2006). More seriously, Eberherdt, Goff, Purdie and Davies (2004) demonstrated that White college students and White police officers implicitly associated criminality more with Black people than with White people, and this pro-White/anti-Black bias was enhanced for Black persons with darker skin. Indeed, Maddox and Gray (2002) demonstrated that *both* Black and White participants more readily associated very dark–skinned Black persons with negative racial stereotypic characteristics (e.g. criminal, poor, aggressive) than Black persons of paler skin colour.

Racial Bias as Safe and Dangerous

Research on attitudes to race has employed a range of attributes against which White and Black comparisons can be drawn and one commonly used dimension is safe vs. dangerous (e.g. Devine, 1989). For instance, Duncan (1976) reported that

mildly aggressive behaviour is perceived as more threatening when performed by a Black person than by a White person (see also Sagar, & Schofield, 1980). In fact, safe/dangerous appears to be a particularly useful attribute for drawing such comparisons because attentional bias is believed to be enhanced in a context of perceived threat. Indeed, an attentional bias for weapons is commonly observed in the context of eye-witness testimony. That is, when a weapon is present an individual's attention is focused on the weapon and many of the other contextual variables (e.g. the person holding the weapon) are ignored. For example, Yarmey and Jones (1983) reported that 88% of expert psychologists they interviewed believed that the presence of a weapon interferes with a witness's ability to remember the perpetrator's face. In fact, priming researchers have argued that weapon bias is such a powerful effect that participants report seeing a weapon in the context of racial stereotyping, even when none is present (Payne, 2006). This effect is more likely when participants are asked to respond at speed, rather than when the task is subject to conscious control (Payne, Lambert, & Jacoby, 2001). Hence, it is not clear in he literature to what extent attentional weapon bias interacts with racial discriminations.

Of perhaps greater concern, researchers of implicit cognitions have also reported strong links between attentional bias and overt behaviour. In a video-game context, Correll, Wittenbrink and Judd (2002) explored the impact of ethnicity on participants' decisions to 'shoot' and thus addressed the issue of the interaction of weapon bias and race. In this particular procedure, participants could see the ethnicity of the target persons and the objects held by them, but ethnicity was irrelevant to whether or not the targets were armed. The results indicated that armed targets who were African-American were shot more readily than armed-targets who were White and White unarmed targets were shot less than unarmed African-American (see

Greenwald et al., 2003 for a similar study). In this case race enhanced weapon bias. Implicit measures, therefore, appear to have been successful at highlighting strong and even dangerous implicit racial prejudice, even in a culture where overt racial practices are politically incorrect.

The Current Research

At the time of writing, there were no published studies that had employed the IRAP as a measure of implicit attitudes to race. The research reported in the current chapter (Experiments 2-6) attempted to address this issue directly. Experiment 2 was a preliminary study that attempted to explore pro-White and anti-Black bias by presenting an IRAP with the phrases WHITE PERSON and BLACK PERSON in conjunction with sets of target words that were readily categorised as safe (e.g. FRIENDLY) or dangerous (e.g. HOSTILE). In Experiments 3 and 4, new IRAPs containing White vs. Black pictures were presented and the items in the pictures were manipulated in order to determine what influence this might have on attitudes to race. Specifically, in Experiment 3, the IRAP contained pictures of White and Black men holding guns, while in Experiment 4 the White and Black men in the pictures all held mundane objects. In Experiments 5 and 6, both the White and Black men held guns *and* mundane objects. In Experiments 2 to 5, correct responding on consistent trials *always* involved categorising White as safe and Black as dangerous and responding on inconsistent trials required the reverse (Black as safe and White as dangerous). In other words, the item held in the pictures was actually irrelevant to correct responding. However in Experiment 6, correct responding was determined directly by the items held by the men in the pictures. In this case, guns were always categorised as dangerous and mundane objects as safe in consistent trials, but the reverse was true

for inconsistent trials. Hence, in Experiment 6 participants were not responding on the basis of race directly but we wanted to determine whether this would influence their reactions to the objects held.

EXPERIMENT 2

The primary aim of the preliminary study in Experiment 2 was to investigate the utility of the IRAP as an implicit measure of race. In this IRAP, participants were simply required to relate the phrases WHITE PERSON and BLACK PERSON to a range of simple positive and negative terms that could be readily categorised as safe and dangerous.

Participants

A sample of thirty participants was recruited for Experiment 2. All were Caucasian Irish undergraduate students at NUIM and were aged between 20 and 22 years old (M: 21 years and 6 months). All participants were recruited through class announcements and notice boards in the Psychology Department. No participants had previous experience of the IRAP and they received no remuneration for their part in the research. All were required to have high fluency in English, as well as normal or corrected to normal vision. In this study, all participants reached criterion in practice trials for the IRAP, hence, all are included in the data analyses.

Setting

All aspects of the setting were identical to Experiment 1.

Materials and Apparatus

Experiment 2 involved two basic sets of materials. The *explicit* measures comprised of a modified version of the Modern Racism Scale (MRS); the Discrimination and Diversity Scale (DDS); the Motive to Respond Without Prejudice Scale (MRWPS); a Racial Experience Questionnaire (REQ); a Feeling Thermometer; and Likert scales. The *implicit* measure comprised of the IRAP.

Explicit Measures. The *Modified Modern Racism Scale (MMRS)* was based on the original by McConaghy (1986) that was adapted for use with Irish participants. The measure comprised of a 5-point scale ranging from –2/Strongly Disagree to +2/Strongly Agree, along which participants indicated their level of agreement or disagreement with each of six statements (e.g. "It is easy to understand the anger of black people", see Appendix B). There was a maximum score of 10 and a minimum score of -10. A low score indicated an absence of racist attitudes.

The *Discrimination and Diversity Scale* (DDS) is a measure of racial discrimination. It comprised of two separate sub-scales -- The Discrimination Scale and the Diversity Scale. The Discrimination Scale comprised of ten statements (e.g. "More and more, Blacks use accusations of racism for their own advantage") that were scored from 1/Strongly Agree to 5/Strongly Disagree -- see Appendix C). There was a maximum score of 50 and a minimum score of 10. A high score indicated an absence of racist attitudes. The Diversity Scale was designed to measure explicit attitudes to ethnic diversity. It consisted of four statements (e.g. "There is real danger that too much emphasis on cultural diversity will tear Ireland apart") and was scored from 1/Strongly Agree to 5/Strongly Disagree, see Appendix C. There was a maximum score of 20 and a minimum score of 4. A high score indicated an absence

of racist attitudes. In three instances the words United States were substituted with Ireland to ensure this measure would be suitable for Irish participants.

The Racial Experience Questionnaire (REQ) was a closed questionnaire designed for current experimental purposes. It contained seven questions (e.g. "How many Black acquaintances do you have?"), five of which were scored on a 4-point scale from 1/None to 4/Many, see Appendix D. On one additional question, participants were asked to rate the positivity or negativity of their experiences with Black people on a 5-point scale from 1/Very Negative to 5/Very Positive.

The Feeling Thermometer simply required participants to rate their explicit feelings towards Irish people and Black people on a scale of 0/Cold to 100/Warm, see Appendix E.

In the *Likert scales*, participants were presented with a series of opposing terms (e.g. SAFE vs. DANGEROUS and FRIENDLY vs. HOSTILE) that were the juxtaposed target stimuli within the IRAP. Participants were asked to rate which term of each pair they considered *Black People* to be like using a 7-point scale, where -3 was most associated with one term (e.g. DANGEROUS) and +3 was most associated with the opposing term (e.g. SAFE, see Appendix F). There was a maximum score of 18 and a minimum score of -18. A high score indicated that participants considered Black People to be more positive (e.g. safe) than negative (e.g. dangerous).

Implicit Measure. The IRAP employed in Experiment 2 was identical in format to the previous study, except for the target stimuli (see Table 3). The sample stimuli comprised the two phrases: WHITE PERSON and BLACK PERSON. The target stimuli contained twelve evaluative terms, six readily categorised as safe (SAFE, FRIENDLY, PEACEFUL, PASSIVE, HARMLESS and TRUSTWORTHY) and six readily categorised as dangerous (DANGEROUS, HOSTILE, VIOLENT,

AGGRESSIVE, THREATENING and MALICIOUS). The response options in the current IRAP were TRUE and FALSE. The screen shots employed as visual aids in the previous experiment were adapted for current purposes.

Table 3
The Stimulus Arrangements Employed in Experiment 2

Sample 1 WHITE PERSON	**Sample 2** BLACK PERSON
Response Option 1 TRUE	**Response Option 2** FALSE
Target Deemed Consistent with Sample 1	**Target Deemed Consistent with Sample 2**
SAFE	DANGEROUS
FRIENDLY	HOSTILE
PEACEFUL	VIOLENT
PASSIVE	AGGRESSIVE
HARMLESS	THREATENING
TRUSTWORTHY	MALICIOUS

Procedure

All participants completed the study in a single experimental session that lasted between 30 and 40mins. All were presented with the same experimental sequence that involved exposure to the explicit measures followed after a short break by the IRAP.

Explicit Measures. Participants were provided with the following verbal instructions regarding completion of the explicit measures:

> These questionnaires are trying to find out how more about your attitudes towards race. Please read all questions carefully and please answer all of the questions as honestly as you possibly can. Your answers are completely confidential and anonymous.

The IRAP. All procedural aspects of the IRAP employed in Experiment 2 were identical to the previous study, in the context of the new stimuli. In this case, the four basic trial-types were referred to as: WHITE/Safe (top left of Figure 4); WHITE/Dangerous (top right); BLACK/Safe (bottom left); and BLACK/Dangerous

(bottom right). Each block of IRAP trials comprised of six exposures to each of the four trial-types – hence one exposure to each of the six target words.

In the IRAP in Experiment 2, correct responses on consistent trials involved positivity towards WHITE PERSON and negativity towards BLACK PERSON. That is, participants, for example, were required to respond as follows: WHITE PERSON/SAFE/TRUE; WHITE PERSON/DANGEROUS/FALSE; BLACK PERSON/SAFE/FALSE; and BLACK PERSON/DANGEROUS/TRUE. In contrast, correct responses on inconsistent trials involved negativity towards WHITE PERSON and positivity towards BLACK PERSON (i.e. WHITE PERSON/SAFE/FALSE; WHITE PERSON/DANGEROUS/TRUE; BLACK PERSON/SAFE/TRUE; and BLACK PERSON/DANGEROUS/FALSE).

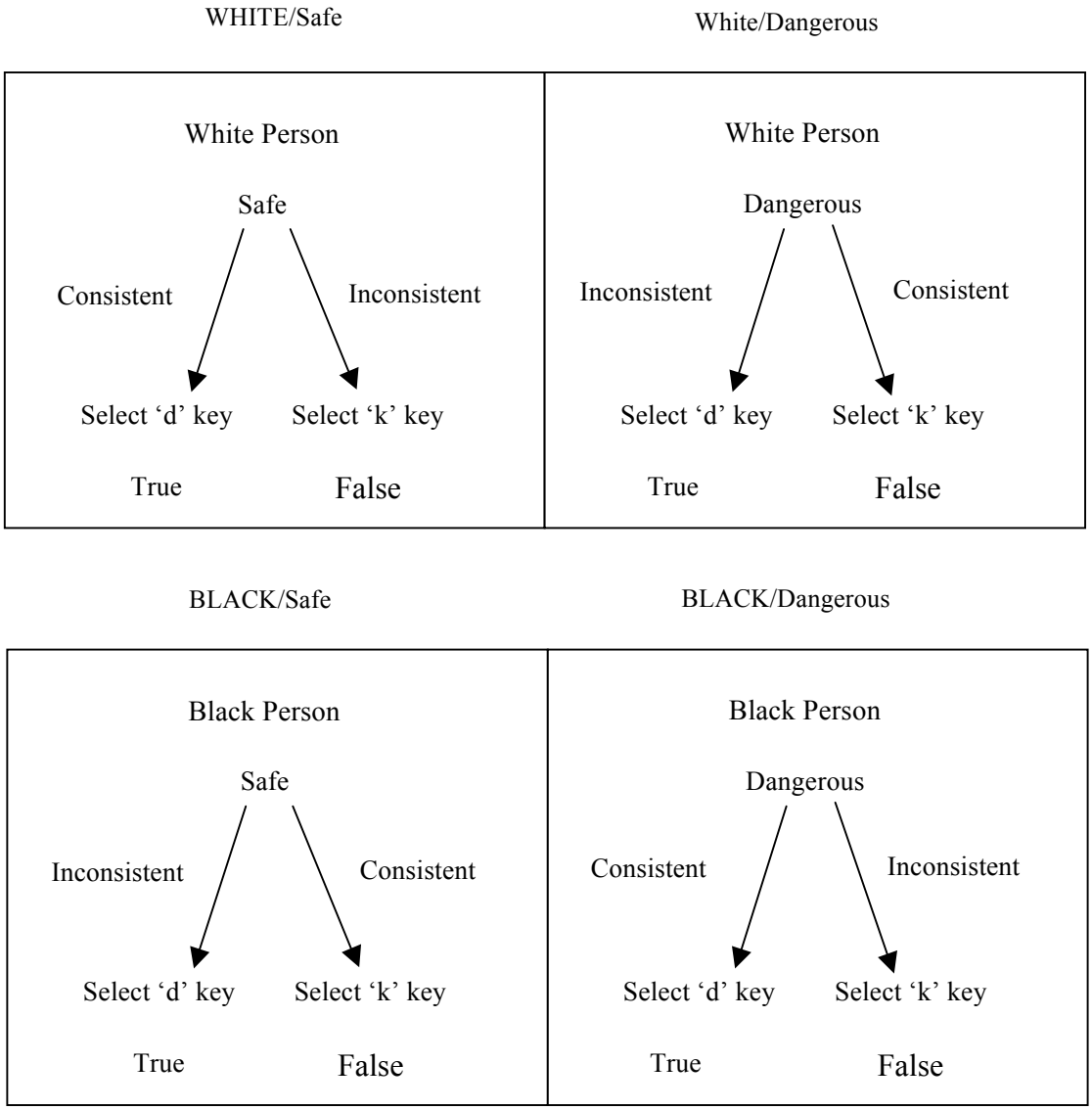


Figure 4. A schematic representation of the four basic trial-types in Experiment 2.

RESULTS

The primary analyses of the D-$_{IRAP}$ scores conducted here involved the D-$_{IRAP}$ trial-type data (rather than using overall D scores as had been the case in the previous study). The D-$_{IRAP}$ trial-type data are particularly beneficial for analysing race because they permit distinctions between responding, for example, to Black versus White and would, in subsequent experiments, be used to compare responding to Black and White pictures involving different types of held objects.

Figure 5 shows the mean D-$_{IRAP}$ score for each of the four trial-types in Experiment 2 and highlights considerable differences in this regard. Similar to Experiment 1, responding in the predicted direction (that was, responding that White was safe and Black was dangerous) was indicated in the graphs by bars above a D-$_{IRAP}$ score of zero (positive D-$_{IRAP}$ scores) and responding that was contrary to prediction (responding that Black was safe and White was dangerous) was indicated by bars below a D-$_{IRAP}$ score of zero (negative D-$_{IRAP}$ scores). The largest D-$_{IRAP}$ score was recorded for BLACK/Safe, but this was not in the predicted direction. That is, contrary to predictions, participants more readily confirmed than denied that Black was safe. WHITE/Safe produced a smaller but sizeable D-$_{IRAP}$ score and this was in the predicted direction. Hence, participants more readily confirmed than denied that White was safe. The D-$_{IRAP}$ scores for the two remaining trial-types were considerably smaller. WHITE/Dangerous was only marginally above zero, but was in the predicted direction. This indicated that participants marginally more readily denied than confirmed that White was dangerous. The smallest D-$_{IRAP}$ score recorded was for BLACK/Dangerous which was negligibly below zero, thus indicating that participants neither confirmed nor denied that Black was dangerous.

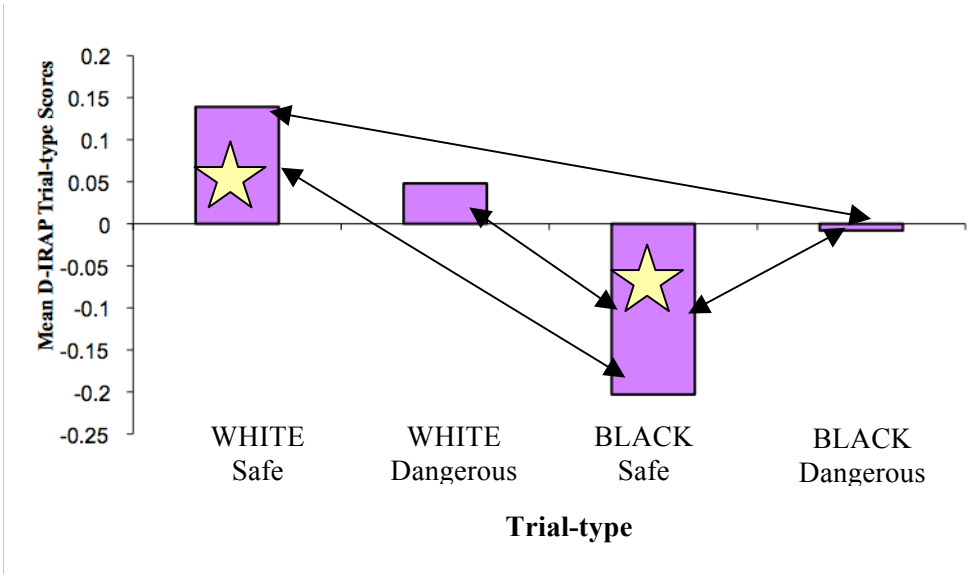


Figure 5. The mean D-$_{IRAP}$ trial-type scores recorded in Experiment 2. * Indicates that the trial-type was significant relative to zero. Arrow indicates trial-types that were significantly different from each other.

A 2x3x4 mixed repeated measures ANOVA was conducted with sequence (consistent-first vs. inconsistent-first) as the between participant variable and test order (Test 1 vs. 2, Test 3 vs. 4 and Test 5 vs. 6) and trial-type as within participant variables. There was a highly significant main effect for trial-type [$F(3, 84) = 7.302$, $p = .001$, $\eta_p^2 = 0.206$], but not for test order or sequence (both p's > .211), all other p's > .14. Post-hoc analyses (Fisher's PLSD) indicated, as expected, that BLACK/Safe was significantly different from the other three trial-types: WHITE/Safe ($p < .001$); WHITE/Dangerous ($p = .001$); and BLACK/Dangerous ($p = .012$). WHITE/Safe and BLACK/Dangerous were marginally significantly different from each other ($p = .059$, all other p's > .237).

Four one sample t-tests were conducted to determine if the D-$_{IRAP}$ score for each trial-type differed significantly from zero. As expected, only WHITE/Safe [t (29)

= 2.179, p = .038] and BLACK/Safe [t (29) =-4.504, p < .001] were significant. WHITE/Dangerous and BLACK/Dangerous were not (both p's > .416).

IRAP Summary

The D-$_{IRAP}$ trial-type scores were in part consistent with experimental predictions. Specifically, responding on both of the White trial-types indicated the expected pro-White bias, although responding on WHITE/Dangerous was not significant. Hence, participants significantly confirmed that White was safe and to a lesser extent they denied that White was dangerous. However, responding on the two Black trial-types was contrary to predictions and failed to show an anti-Black bias. Specifically, participants significantly confirmed that Black was safe and neither confirmed nor denied that Black was dangerous.

Explicit Measures

A summary of the outcomes (i.e. means, percentages etc.) recorded on the explicit measures and a brief interpretation of these outcomes is presented in Table 4.

Table 4
A Summary and Interpretations of Outcomes on the Explicit Measures in Experiment 2.

Explicit Measures	**Mean/Score**	**Participants Range**	**Interpretation of Attitudes**
The Feeling Thermometer-Irish	83	50-100	High warmth towards Irish people.
The Feeling Thermometer-Black	63.33	0-100	Medium warmth towards Black people.
Discrimination Scale	35.2	24-42	Not racially biased against Black people.
Diversity Scale	12.7	9-17	Not racially biased against Black people.
Modified Modern Racism Scale (MMRS)	-4.4	-9 to +1	Not racially biased against Black people.
The Racial Experience Questionnaire Sub-Scales			
Black Acquaintance	92.4%= Yes 7.6% = No		Most participants knew Black people.
Black Friend	62.7%= Yes 36.3% = No		Two thirds of participants had a Black friend.
Black Colleague	23.1%= Yes 75.9% = No		One quarter of participants had a Black colleague.
Black Relationship	36.3%= Yes 62.7% = No		One third of participants had a close Black relationship.
Black Neighbour	75.9%= Yes 23.1% = No		Three quarters of participants had Black neighbours.
Black Encounters	42.9% = Positive 46.2%= Neutral 9.9% = Negative		Just under half of participants had positive and equally positive/negative Black experiences.
Likert Scales	3.83	+16 to -13	Black persons perceived as safe rather than dangerous.

On the majority of the explicit measures, participants indicated the absence of racial bias against Black people (e.g. on the DDS and the MMRS). Furthermore, on the Feeling Thermometer, they rated Irish people only slightly higher than Black people and on the Likert scales they more readily associated Black people with safe than dangerous. The other explicit measures appear to provide some explanation for these outcomes in terms of direct experiences participants reported with Black persons. Specifically, the majority of participants had positive acquaintances, friends and neighbours who were Black and a smaller percentage had Black work colleagues and close relationships with Black people.

Correlations Among Explicit Measures

A series of Fishers r to z correlations were undertaken among the explicit measures in order to determine the potential relationships therein. The correlation coefficients and significance values for all of the planned comparisons that proved significant are presented in Table 5.

Table 5
Observed Correlations Among the Explicit Measures Employed in Experiment 2

Measures Compared	**Correlation**	**P-Value**
Correlations Expected to be Positive		
Discrimination vs. Diversity	.508	.004
Discrimination vs. Feeling Thermometer Irish	-.448	.012
Discrimination vs. Feeling Thermometer Black	.522	.003
Correlations Expected to be Negative		
MMRS vs. Discrimination Scale	-.612	0.00
MMRS vs. Diversity	-.415	.022
Likert Scales vs. Feeling Thermometer Black	-.511	.003
Discrimination vs. Likert Scales	-.506	.004
Diversity vs. Encounters with Black People	-.404	.026
Feeling Thermometer Black vs. Black Acquaintance	- .417	.021
Likert Scales vs. Black Acquaintance	-.517	.003

Overall, there were numerous correlations amongst the explicit measures that suggested that they may have been measuring the same attitudes. For example, the correlations were as expected and significant between the DDS sub-scales and the MMRS; and the Discrimination Scale with the Likert scales. Interestingly, the Feeling Thermometer for Black persons correlated also with the Discrimination Scale and with the Likert scales, indicating that they were tapping into the same attitudes to Black people. The negative correlation between the Discrimination Scale and the Irish Feeling Thermometer suggested that favourability towards Irish did not assume disfavour towards Black. This view was also supported by the positive correlation between the Discrimination Scale and the Black Feeling Thermometer. The correlation between the Likert scales and responding on the Black Acquaintance sub-

scale of the REQ suggested that having a Black acquaintance increased the likelihood of responding to Black as safe. Furthermore, the correlation between the Diversity Scale and the Encounters with Black People sub-scale suggested that greater neutrality towards Black people was associated with greater diversity. Again, the negative correlation between the Black Feeling Thermometer and the Black Acquaintance sub-scale suggested that more acquaintances were correlated with greater warmth.

Explicit and Implicit Correlations

Although the majority of IRAP studies frequently do not find significant correlations between the IRAP and explicit measures, a series of Fishers r to z correlations were conducted here to correlate the overall D-$_{IRAP}$ scores with the explicit measures. Only the Black Encounters sub-scale ($r = .433$, $n = 29$, $p = .016$) and the Black Acquaintance sub-scale ($r = .356$, $n = 29$, $p = .053$) correlated significantly with the IRAP (all p's > .105). That is, high overall D-$_{IRAP}$ scores showing stronger racial bias correlated with more negative experiences with Black people.

Results Summary

The D-$_{IRAP}$ trial-type scores were in part consistent with experimental predictions in showing some element of a pro- White bias. Specifically, participants significantly confirmed that White was safe and had mixed views about Black as dangerous. However, the two Black trial-tpes were unexpectedly similar to the White trial-types. That is, paarticipants significantly confirmed that Black was safe and had mixed views of Black as dangerous. Hence, there was no anti-Black bias. On

practically all of the explicit measures, participants indicated the absence of racial bias against Black people. While there were numerous correlations amongst the explicit measures that suggessted that they may be measuring the same attitudes, overall $D\text{-}_{IRAP}$ score correlated only with the level of Black acquaintances.

DISCUSSION

The primary aim of Experiment 2 was to examine the IRAP as an implicit measure of race, and in particular to examine evaluations of Black vs. White as safe and dangerous. In the current study, this was achieved by examining the relationships between the phrases WHITE PERSON and BLACK PERSON and a range of evaluative terms, readily dichotomised as safe or dangerous (e.g. FRIENDLY and HARMLESS). In this context, responding on both White trial-types showed similar pro- White and pro- Black effects in terms of evaluating both as safe, but they had mixed views on both as dangerous. Hence there was no anti- Black effect.

The current findings in terms of the absence of an anti-Black bias are inconsistent with practically all of the IAT studies of race (McConnell et al., 2008, Dasgupta, McGee, Greenwald, & Banaji, 2000). This discordance of evidence in the latter may have occurred for a variety of reasons. (1) The current experiment was conducted with an Irish population, whereas the majority of IAT studies have employed American participants, who live in a much more multi-cultural environment. (2) Numerous IAT studies have employed picture, rather than word, stimuli and the latter may be particularly useful in distinctions that involve safe vs. dangerous (rather than more simplistic good vs. bad distinctions -- Greenwald, Oakes, & Hoffman, 2002). Experiment 3 attempted to address this latter possibility directly by employing a picture-based, rather than word-based, IRAP.

EXPERIMENT 3

The primary aim of Experiment 3 was to extend the preliminary work in Experiment 2 that had used word stimuli to examine implicit attitudes to race in terms of evaluations of safe and dangerous. In short, we wanted to determine if stronger IRAP outcomes, particularly in terms of anti-Black, would be recorded with picture stimuli. Hence, in this study, participants were presented with pictures of Black and White men. In order to enhance the Black vs. White distinction and to examine the possible role of attentional weapon bias, Experiment 3 presented pictures in which both the Black and White men were holding guns. However, it is important to emphasise that consistent and inconsistent responding on the IRAP were not differentiated by the presence of the guns. Put simply, in this IRAP consistent responding required relations in which White was always safe and Black was always dangerous and inconsistent responding required the reverse.

Participants

A sample of twenty-seven participants, aged between 18 and 30 years of age (M: 24 years and 2 months) were recruited for Experiment 3. All were Caucasian Irish undergraduates at NUIM and were recruited through class announcements and Departmental notice boards. They had no experience of the IRAP and received no remuneration for their involvement in the research. All participants presented with high levels of fluency in English and normal or corrected to normal vision. Only one participant was removed from the data analyses for failing to reach criterion during the IRAP practice blocks, thus leaving a sample of twenty-six.

Setting

The setting was identical to Experiment 2.

Materials and Apparatus

Experiment 3 involved the same two basic sets of materials as Experiment 2, except for the new IRAP stimuli and the related adjustments to the visual aids and Likert scales.

Likert scales. The Likert scales employed here each comprised of a 13-point scale that assessed participants' explicit attitudes towards the six pictures of Black and White men holding guns that would subsequently be presented in the IRAP. On each rating, participants were required to indicate the extent to which they rated the man in the picture as SAFE or DANGEROUS from -6/Extremely Safe to +6/Extremely Dangerous, see Appendix H.

The IRAP. The IRAP employed in Experiment 3 was identical in format to the previous study, except for the different sample and target stimuli (see Table 6). In the current case, the sample stimuli were the words SAFE and DANGEROUS. The target stimuli were three pictures of Black men holding guns and three pictures of White men holding (the same three) guns. The response options were TRUE and FALSE. The six pictures comprised of digital coloured photographs taken from the International Affective Picture System (IAPS: Lang, Bradley, & Cuthbert, 1999) and had been previously employed in IAT research by Greenwald et al., (2003). All pictures were 259 pixels wide, 305 pixels tall, in 256-colour greyscale format and presented against a white background.

Table 6
The Stimulus Arrangements Employed in Experiment 3

Sample 1 SAFE	**Sample 2** DANGEROUS
Response Option 1 TRUE	**Response Option 2** FALSE
Targets deemed consistent with Sample 1	**Targets deemed consistent with Sample 2**
White man with gun 1	Black man with gun 1
White man with gun 2	Black man with gun 2
White man with gun 3	Black man with gun 3
White man with gun 1	Black man with gun 1
White man with gun 2	Black man with gun 2
White man with gun 3	Black man with gun 3

Procedure

All participants completed the study in a single experimental session that lasted between 30 and 40mins. in total. Although made available to them, none of the participants opted for short breaks at any point.

The IRAP. All procedural aspects of the IRAP employed in Experiment 3 were identical to the previous studies. The trial-types presented here were referred to as: SAFE/White (top left of Figure 6); SAFE/Black (top right); DANGEROUS/White (bottom left); and DANGEROUS/Black (bottom right). As before, each block of trials was composed of six exposures to each of the four trial-types, hence two exposures to each of three photos for each trial-type.

Correct responses on consistent trials involved indicating that the White pictures were SAFE and the Black pictures were DANGEROUS (irrespective of the fact that both held guns). That is, participants were required to respond as follows: SAFE/White/TRUE; SAFE/Black/FALSE; DANGEROUS/White/FALSE and DANGEROUS/Black/TRUE. In contrast, correct responses on inconsistent trials involved indicating that the Black pictures were SAFE and the White pictures were DANGEROUS. That is, participants were required to respond as follows:

SAFE/White/FALSE; SAFE/Black/TRUE; DANGEROUS/White/TRUE; and DANGEROUS/Black/FALSE.

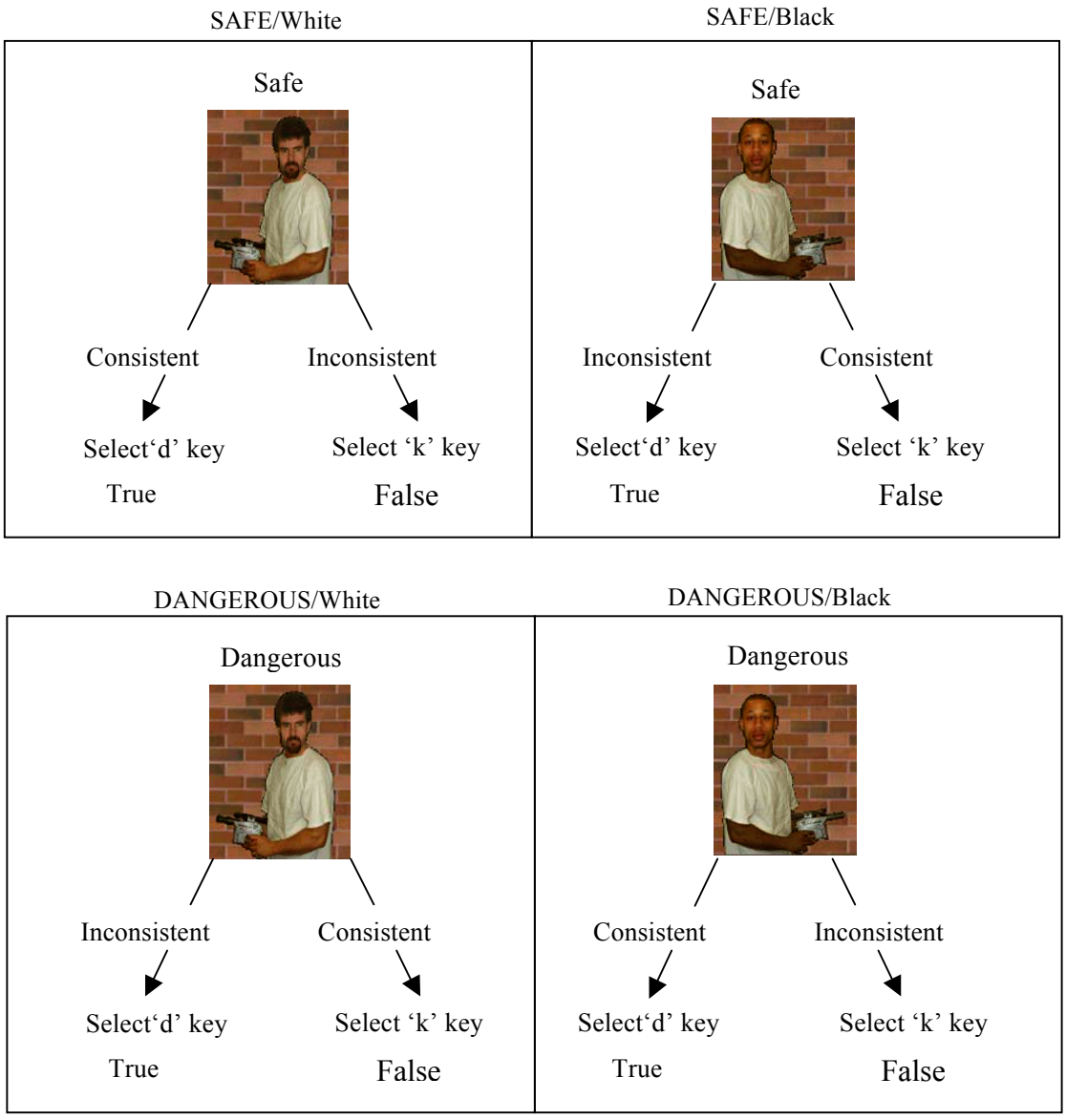


Figure 6. A schematic representation of the four basic trial-types in Experiment 3.

RESULTS

The mean $D\text{-}_{IRAP}$ scores for the four trial-types in Experiment 3 are presented in Figure 7. Again,responding in the predicted direction (that was, responding that White was safe and Black was dangerous) was indicated in the graphs by bars above a

D-IRAP score of zero (positive D-IRAP scores) and responding that was contrary to prediction (responding that Black was safe and White was dangerous) was indicated by bars below a D-IRAP score of zero (negative D-IRAP scores). The largest D-IRAP scores were recorded for the SAFE/White and DANGEROUS/Black trial-types, both of which were in the predicted direction. That is, participants more readily confirmed than denied that White was safe (thus showing a pro- White bias) and that Black was dangerous (thus showing an anti-Black bias). The D-IRAP for SAFE/Black was also sizable, but was in the non-predicted direction. That is, participants more readily confirmed than denied that Black was safe, hence showing a pro-Black bias. The D-IRAP scores for DANGEROUS/White was the smallest and was also in the non-predicted direction. That is, participants more readily confirmed than denied that White was dangerous.

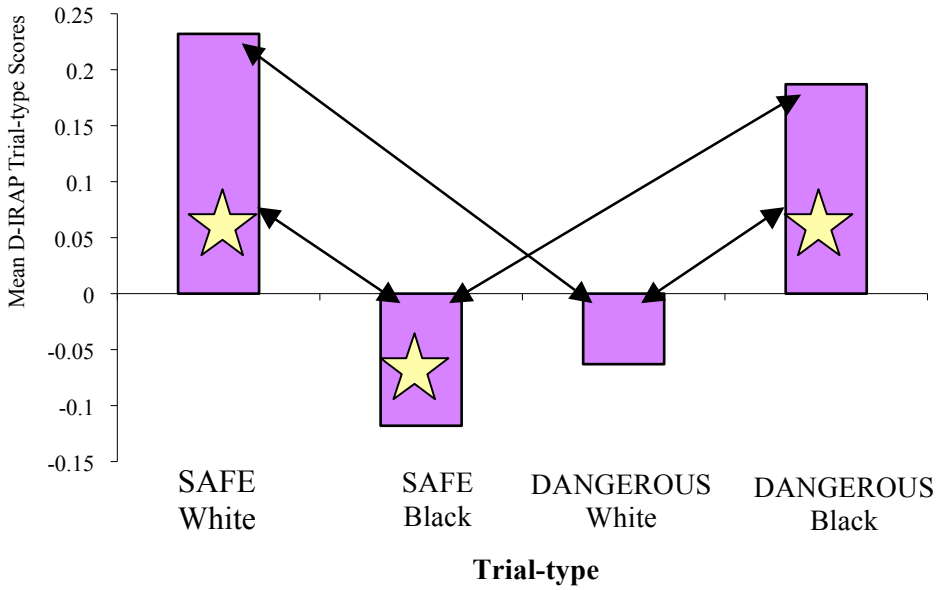


Figure 7. The mean D-IRAP trial-type scores recorded in Experiment 3.*Indicates that the trial-type was significant relative to zero. Arrow indicates trial-types that were significantly different from each other.

The D-IRAP trial-type scores were subjected to a 2x3x4 mixed repeated ANOVA. Trial-type was significant [$F(3, 66) = 6.486$, $p = .001$, $\eta_p^2 = 0.235$], but test

order and sequence were not (both p's > .277), all other p's > .164. Post-hoc tests (Fishers' PLSD) indicated that SAFE-White differed significantly from SAFE/Black (p = .001) and from DANGEROUS/White (p = .005). DANGEROUS/Black differed significantly from SAFE/Black (p = .004) and from DANGEROUS/White (p = .002), all other p's > .592.

Four one-group t-tests were conducted to determine whether the D-$_{IRAP}$ scores for each trial-type differed significantly from zero. SAFE/White [t (23) = 3.358, p < .003] and DANGEROUS/Black [t (23) = 2.483, p < .021] were significant, and SAFE/Black was approaching significance [t (23) = -1.779, p = .088]. Only DANGEROUS/White was not significant (p = .411).

IRAP Summary

Large and significant D-$_{IRAP}$ scores were recorded for SAFE/White and DANGEROUS/Black and both were in the predicted direction. That is, participants more readily confirmed than denied that White was safe and that Black was dangerous showing a pro-White and anti-Black bias respectively. The D-$_{IRAP}$ score for SAFE/Black was also significant but was not in the predicted direction, indicating that participants confirmed rather than denied that Black was safe (thus showing a pro-Black bias). Participants had mixed views on White as dangerous.

Explicit Measures

A summary of the outcomes (i.e. means, percentages etc.) recorded on the explicit measures and a brief interpretation of these outcomes is presented in Table 7. On the majority of explicit measures, participants again indicated the absence of racial bias against Black people. Again, most had numerous and positive experiences with Black persons.

Table 7
A Summary and Interpretations of Outcomes on the Explicit Measures in Experiment 3

Explicit Measures	Mean/Score	Participants Range	Interpretation of Attitudes
The Feeling Thermometer-Irish	70	65-80	Medium warmth towards Irish people.
The Feeling Thermometer-Black	60	50-70	Medium warmth towards Black people.
Discrimination Scale	39	37-43	Not racially biased against Black people.
Diversity Scale	12	10-13	Not racially biased against Black people.
Modified Modern Racism Scale (MMRS)	-10	-6 to -11	Not racially biased against Black people.
The Racial Experience Questionnaire Sub-Scales			
Black Acquaintance	100%= Yes 0% = No		All participants knew Black people.
Black Friend	66%= Yes 33% = No		Two thirds of participants had a Black friend.
Black Colleague	25%= Yes 75% = No		One quarter of participants had a Black colleague.
Black Relationship	33%= Yes 66% = No		One third of participants had a close Black relationship.
Black Neighbour	25%= Yes 75% = No		One quarters of participants had Black neighbours.
Black Encounters	50% = Positive 50%= Neutral 0% = Negative		Half of participants had positive and half had equally positive/negative Black experiences.
Likert Scales	-12	-6 to -14	Black and White persons with guns perceived as dangerous.

On the majority of the explicit measures, participants again indicated the absence of racial bias against Black people. Again, most had numerous and positive experiences with Black persons. Participants were moderately motivated to respond in a non-prejudiced way and were slightly more motivated by internal than external constructs in this regard.

Correlations Amongst Explicit Measures

Statistical analyses (Fisher's r to z) were conducted to determine whether any correlations existed among the explicit measures. The results are presented in Table 8.

Table 8
Statistically Significant Correlations among Explicit Measures in Experiment 3

Planned Comparisons between Explicit Measures	**Correlation**	**P-value**
Correlations Expected to be Positive		
Discrimination vs. Diversity	.557	.024
Diversity vs. REQ 1	.641	.003
Correlations expected to be Negative		
Discrimination vs. MMRS	-.623	.002

A number of expected correlations were observed amongst the explicit measures that indicated that they were likely tapping into the same explicit attitudes. Not surprisingly, there was a strong correlation between the DDS sub-scales. The Diversity scale was also positively correlated with Black acquaintances, indicating that more acquaintances was correlated with greater diversity. The Discrimination scale was positively correlated with Black acquaintances, indicating that more acquaintances was correlated with greater diversity. There was a negative correlation between the MMRS and Discrimination scale, but this is expected because a low MMRS score and a high Discrimination score both equate to non-racist views.

Explicit-Implicit Correlations

A series of correlations (Fisher's r to z) were conducted between the overall D-$_{IRAP}$ score and the explicit measures. None of these proved significant (all p's > .125).

Results Summary

Large and significant D-$_{IRAP}$ scores were recorded for SAFE/White and DANGEROUS/Black and both were in the predicted direction. That is, participants more readily confirmed than denied that White was safe (pro-White bias) and that

Black (anti-Black bias) was dangerous. The $D\text{-}_{IRAP}$ score for SAFE/Black was also significant but was not in the predicted direction, indicating that participants confirmed rather than denied that Black was safe (Pro- Black bias). Participants had mixed views on White as dangerous. On the majority of explicit measures, participants indicated that absence of racial bias against Black people. While there were a number of significant correlations among the explicit measure none of these correlated with the IRAP.

DISCUSSION

One of the primary aims of Experiment 3 was to assess racial bias by juxtaposing Black and White pictures in an IRAP and to determine the extent to which guns presented in all pictures might interfere with the participants' implicit reactions to the Black and White men. The findings indicated that participants showed significant pro-White and pro-Black bias for both the White and Black men in the pictures as safe even when they held guns. However, participants differed in their attitudes to Black and White as dangerous in a manner that may have reflected the influence of the guns. That is, they had mixed views about whether the White men with guns were dangerous, but significantly confirmed that the Black men with guns were dangerous.

A key difference, therefore, between Experiments 2 and 3 lay in the responses to Black and White as dangerous, because in both studies participants significantly confirmed that White and Black were safe. However, in Experiment 2, they had mixed views about both White and Black as dangerous, whereas in Experiment 3 when the pictures contained guns, they had mixed views about White as dangerous but significantly confirmed that Black was dangerous.

One alternative interpretation of the findings, of course, could simply be that the differences across the studies resulted from the fact that one IRAP presented words, while the other presented pictures. While this seemed unlikely given that the guns quite predictably appeared to affect the dangerous trial-types, rather than the safe trial-types, the IRAP in Experiment 4 presented pictures of Black and White men, all holding mundane objects.

EXPERIMENT 4

In order to determine if the weapons presented within the IRAP pictures in the previous study had facilitated different Black and White discriminations in terms of dangerousness, Experiment 4 presented Black and White pictures involving mundane objects.

Participants

A sample of twenty-five participants was recruited for Experiment 4. All were Caucasian Irish undergraduate students at NUIM and were aged between 18 and 28 years old (M: 20 years and 9 months). All participants were recruited through class announcements and notice boards in the Psychology Department. No participants had previous experience of the IRAP and they received no remuneration for their part in the research. All were required to have high fluency in English, as well as normal or corrected to normal vision. In this study, all participants reached criterion in practice trials for the IRAP, hence all are included in the data analyses.

Setting

The setting was identical to Experiment 3.

Materials and Apparatus

Experiment 4 involved the same two basic sets of materials employed in Experiment 3, with the exception of the new IRAP stimuli and the related modifications to the IRAP visual aids and the Likert scales.

The IRAP. The IRAP employed in Experiment 4 was identical in format to the previous study. Again, the sample stimuli were the words SAFE and DANGEROUS. The target stimuli here were three pictures of Black men holding mundane objects (a torch, a camera and a beer bottle) and three pictures of White men holding (the same three) mundane objects. Again, the response options were TRUE and FALSE (see Table 9). The six pictures were again digital coloured photographs taken from the IAPS and previously employed by Greenwald et al. (2003) and the White and Black men pictured here were identical to the previous study (only the items they held were different).

Table 9
The Stimulus Arrangements Employed in Experiment 4

Sample 1 SAFE	**Sample 2** DANGEROUS
Response Option 1 TRUE	**Response Option 2** FALSE
Targets deemed consistent with Sample 1	**Targets deemed consistent with Sample 2**
White Man with Torch	Black Man with Torch
White Man with Camera	Black Man with Camera
White Man with Bottle	Black Man with Bottle
White Man with Torch	Black Man with Torch
White Man with Camera	Black Man with Camera
White Man with Bottle	Black Man with Bottle

Procedure

All participants completed the study in a single experimental session that lasted between 30 and 40mins. in total. Although made available to them, none of the

participants opted for short breaks at any point. All participants were presented with the same experimental sequence that involved exposure to the explicit measure followed after a short break by the IRAP.

The IRAP. All procedural aspects of the IRAP employed in Experiment 4 were identical to the previous study, except for the different target stimuli. Because the change in the target stimuli only affected the items held by the men in the pictures, the trial-types remained the same -- SAFE/White (top left of Figure 8); SAFE/Black (top right); DANGEROUS/White (bottom left); and DANGEROUS/Black (bottom right).

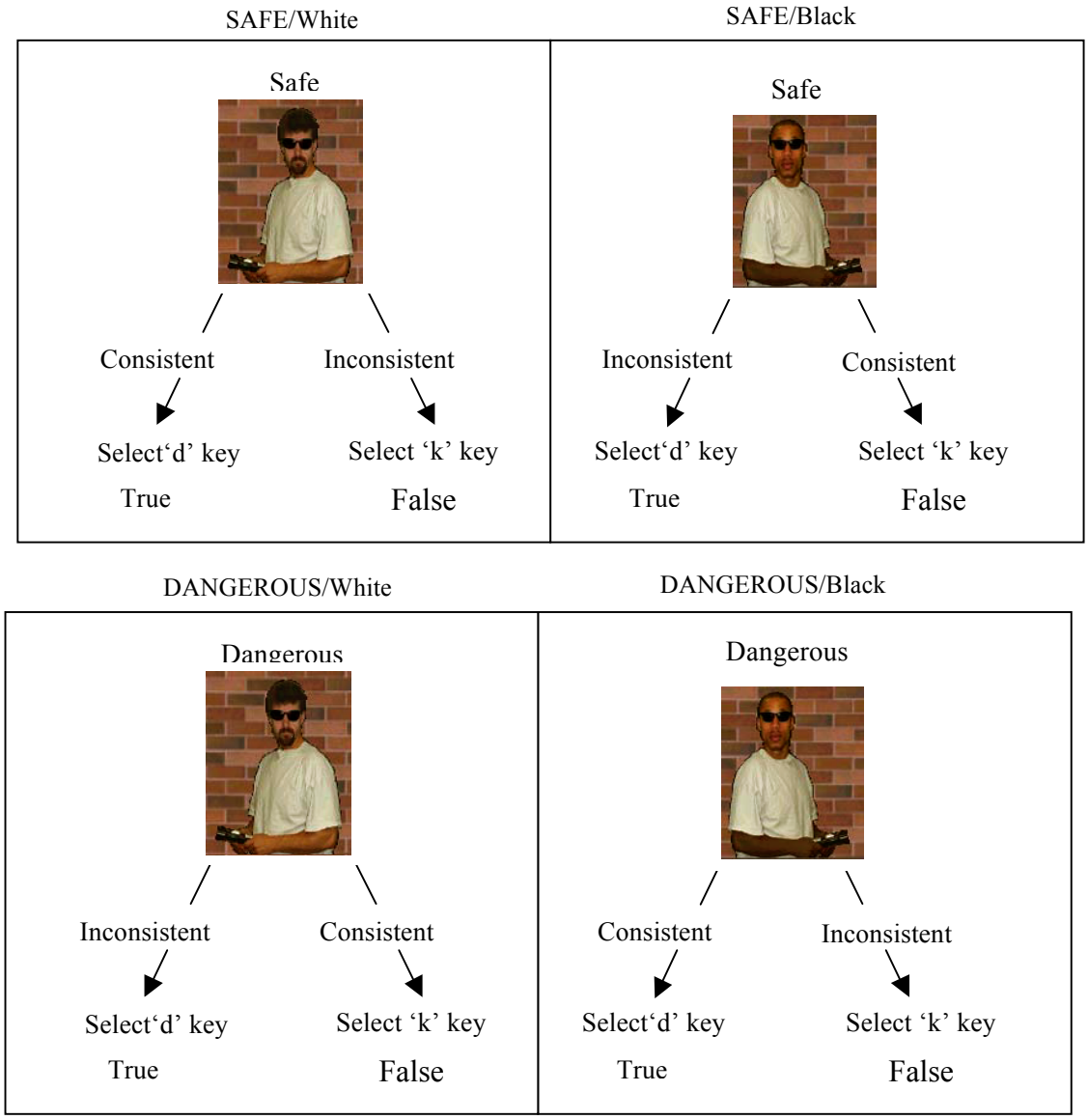


Figure 8. A schematic representation of the four basic trial-types in Experiment 4.

Once again, correct responses on consistent trials involved indicating that White was SAFE and Black was DANGEROUS. That is, participants were required to respond as follows: SAFE/White/TRUE; SAFE/Black/FALSE; DANGEROUS/White/FALSE and DANGEROUS/Black/TRUE. Once again, on these trials, White was always safe and Black was always dangerous. Again, correct responding on inconsistent trials involved indicating that Black was SAFE and White

was DANGEROUS. That is, participants were required to respond as follows: SAFE/White/False; SAFE/Black/TRUE; DANGEROUS/White/TRUE and DANGEROUS/Black/FALSE. Once again, on these trials, White was always dangerous and Black was always safe.

RESULTS

The mean D-$_{IRAP}$ scores for the four trial-types in Experiment 4 are presented in Figure 9. Responding in the predicted direction (that was, responding that White was safe and Black was dangerous) was indicated in the graphs by bars above a D-$_{IRAP}$ score of zero (positive D-$_{IRAP}$ scores) and responding that was contrary to prediction (responding that Black was safe and White was dangerous) was indicated by bars below a D-$_{IRAP}$ score of zero (negative D-$_{IRAP}$ scores). The largest D-$_{IRAP}$ score was recorded in the SAFE/White trial-type, which was in the predicted direction. That is, participants more readily confirmed than denied that White was safe (pro= White bias). A smaller D-$_{IRAP}$ score was recorded in the SAFE/Black trial-type, but responding here was not in the predicted direction. Contrary to the experimental predictions then, participants more readily confirmed than denied that Black was safe (pro- Black bias). The D-$_{IRAP}$ scores for the DANGEROUS/Black and DANGEROUS/White trial-types were considerably smaller and indeed both were in the predicted direction. Participants had mixed view on Black as dangerous. On DANGEROUS/White the effect was so small that it is reasonable to assume that participants neither confirmed nor denied that White was dangerous.

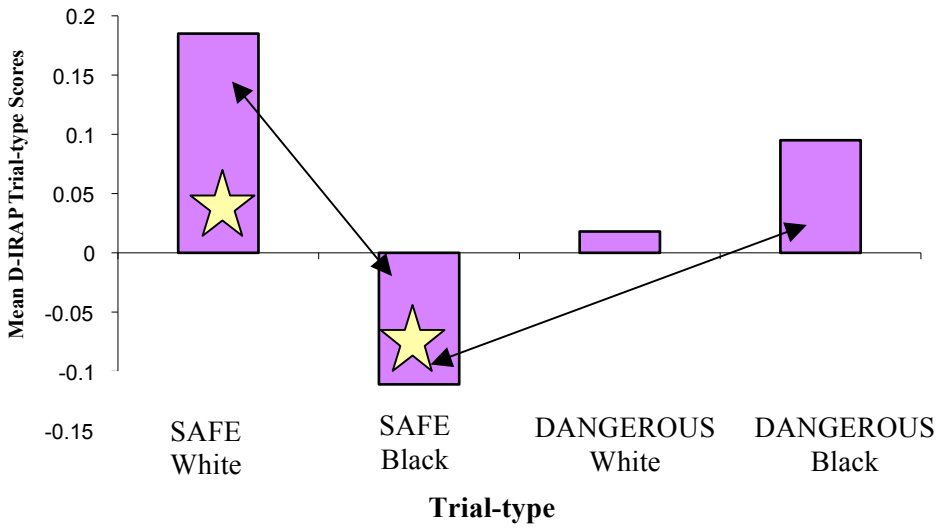


Figure 9. Mean D-$_{IRAP}$ scores recorded for the four trial-types in Experiment 4.

The D-$_{IRAP}$ trial-type scores were subjected to a 2x3x4 mixed repeated measures ANOVA. Trial-type was significant [$F(3, 66) = 1.481$, $p = .01$, $\eta_p^2 = .156$], but test order, sequence and the interaction effects were not, all p's > .174. Post-hoc tests (Fishers' PLSD) indicated that the only significant trial-type differences were between SAFE/White and SAFE/Black ($p = .005$) and between SAFE/Black and DANGEROUS/Black ($p = .048$), all other p's > .105 (see Figure 9).

Four one-sample t-tests (one per trial-type) were conducted to determine whether each of the D-$_{IRAP}$ scores differed significantly from zero. The results indicated that only SAFE/White [$t(23) = 2.642$, $p = .015$] was significant. SAFE/Black approached a significant difference from zero ($p = .081$), but DANGEROUS/White and DANGEROUS/Black were not significant (both p's > .16).

IRAP Summary

The $D_{\text{-IRAP}}$ scores in the current experiment indicated that participants significantly confirmed that White persons were safe (pro-White bias) and to a lesser, marginally significant extent they also confirmed that Black was safe (pro-Black bias). On the Dangerous/Black trial-type participants more readily confirmed than denied that Black was dangerous, but this was not significant. Participants had mixed views about White as dangerous.

Explicit Measures

The mean outcomes on the explicit measures along with a brief interpretation of each are presented in Table 10. On the majority of the explicit measures, participants indicated the absence of racial bias against Black people. Again, the majority reported many positive relationships with Black people.

Table 10
A Summary and Interpretations of Outcomes on the Explicit Measures in Experiment 4

Explicit Measures	Mean/Score	Participants Range	Interpretation of Attitudes
The Feeling Thermometer-Irish	84.38	70-100	High warmth towards Irish people.
The Feeling Thermometer-Black	64.79	30-90	Medium warmth towards Black people.
Discrimination Scale	35.63	26-47	Not racially biased Against Black people.
Diversity Scale	12.9	10-16	Not racially biased Against Black people.
Modified Modern Racism Scale (MMRS)	-5.6	-10- +1	Not racially biased Against Black people.
The Racial Experience Questionnaire Sub-Scales			
Black Acquaintance	95.8%= Yes 4.2% = No		Most participants knew Black people.
Black Friend	75% = Yes 25% = No		Three quarters of participants had a Black friend.
Black Colleague	20.8% = Yes 79.2% = No		One fifth of participants had a Black colleague.
Black Relationship	62.5% = Yes 37.5% = No		Almost two thirds of participants had a close Black relationship.
Black Neighbour	62.5% = Yes 37.5% = No		Almost two thirds of participants had Black neighbours.
Black Encounters	50% = Positive 37.5%= Neutral 12.5% = Negative		The majority of participants had positive or equally positive/negative Experiences with Black people
Likert Scales	3.83	+16- to -13	Black persons perceived as safe rather than dangerous.

Correlations Among Explicit Measures

Statistical analyses (Fisher's r to z) were conducted to determine whether any correlations existed among the explicit measures. The results are presented in Table 11.

Table 11
The Statistically Significant Correlations Among Explicit Measures in Experiment 11.

Planned Comparisons between Explicit Measures	**Correlation**	**P-value**
Correlations Expected to be Positive		
Discrimination vs. Diversity	.457	.0238
Discrimination vs. Feeling Thermometer Black	.578	.0025
Discrimination vs. REQ- Experience	.564	.0034
Diversity vs. Feeling Thermometer Black	.456	.0241
Diversity vs. Feeling Thermometer Irish	-.464	.0212
Feeling Thermometer Black vs. REQ-Experience	.488	.0144
Correlations expected to be Negative		
Discrimination vs. MMRS	-.591	.0018
REQ- Experience vs. MMRS	-.493	.0133

A number of expected correlations were observed amongst the explicit measures that indicated that they were likely tapping into the same explicit attitudes. Again, there was a strong correlation within the DDS and Discrimination also correlated with the MMRS. Interestingly, the Black Feeling Thermometer correlated positively with the Diversity Scale, but there was a negative correlation between the Irish Feeling Thermometer and the Diversity Scale, indicating that participants were both pro-Irish and pro-diversity. There were also correlations between the Experience sub-scale of the REQ and the MMRS and the Black Feeling Thermometer, suggesting that the low racism on the latter measures correlated with direct positive experiences with Black people.

Explicit-Implicit Correlations

A series of correlations (Fisher's r to z) were conducted here between the overall $D\text{-}_{IRAP}$ score and the explicit measures. No significant correlations were found (all p's > .336).

Results Summary

On the IRAP, participants significantly confirmed that the White men with mundane objects, hence showing a pro-White bias. However, they also significantly confirmed that the Black men in the pictures with mundane objects were safe, thus showing a pro-Black bias.. Although they more readily confirmed than denied that Black was dangerous, this was not significant and participants also had mixed views about White as dangerous. On the majority of explicit measures, participants indicated the absence of racial bias against Black people.

DISCUSSION

One of the primary aims of Experiment 4 was to assess racial bias by juxtaposing Black and White pictures in an IRAP, as we had done in the previous study, and to determine the extent to which mundane objects (rather than guns) presented in all pictures might influence participants' implicit reactions to the Black and White men. The findings indicated that participants showed significant pro-White and pro-Black bias for both the White and Black men in the pictures as safe when they held mundane objects. Participants showed mixed views about White and Black as dangerous.

A key difference, therefore, between Experiments 3 and 4 lay in the responses to Black and White as dangerous, because in both studies participants significantly confirmed that White and Black were safe, hence the presence of guns or mundane objects appeared to have no influence on these discriminations. However, in Experiment 3 when the pictures contained guns, they had mixed views about White as dangerous but significantly confirmed that Black was dangerous, whereas in Experiment 4 when the men held mundane objects, they had mixed views about both

White and Black as dangerous. Indeed, the outcome overall for Experiment 4 was almost identical to the word-based IRAP in Experiment 2. This latter similarity also suggested that the differences between Experiments 2 and 3 could not be attributed to contrasts per se between word- and picture-based IRAPs.

The three race studies conducted thus far were the first to employ the IRAP as a measure of implicit attitudes to race, so there was no existing IRAP research to which it could be compared. However, there exists numerous IAT studies of race, which have generally reported a combination of pro-White and anti-Black bias. However, the latter was not recorded with the IRAP's here. Several IAT studies reporting these effects have presented pictures of Black and White men holding both guns and mundane objects, which we had not yet done with the IRAP (Greenwald et al., 2003). We addressed this issue in Experiment 5.

EXPERIMENT 5

In line with previous IAT studies of race, Experiment 5 presented pictures of Black and White men holding both guns and mundane objects. We were also concerned with the potential overlap of findings from such a manipulation we the data we had recorded for the IRAP and race thus far.

Participants

A sample of twenty-seven participants was recruited for Experiment 5. All were Caucasian Irish undergraduate students at NUIM and were aged between 18 and 30 years old (M: 25 years and 4 months). All participants were recruited through class announcements and notice boards in the Psychology Department. No participants had

previous experience of the IRAP and they received no remuneration for their part in the research. All were required to have high fluency in English, as well as normal or corrected to normal vision. In this study, three participants failed to reach criterion in practice trials for the IRAP, hence the data from twenty-four participants are included in the analyses.

Setting

All aspects of the setting were identical to previous experiments.

Materials and Apparatus

Experiment 5 involved two basic sets of materials, identical to those used previously, but with relevant modifications to the IRAP stimuli and visual aids and the Likert scales.

The IRAP. The sample stimuli in Experiment 5 were the words SAFE and DANGEROUS. The target stimuli were twelve pictures -- six of White men and six of Black men. Three pictures involved a White man holding a gun (with a different gun in each picture) that were taken from Experiment 3 and three involved a White man holding a mundane object (a beer bottle, a camera or a torch) taken from Experiment 4. Similarly, three pictures involved a Black man holding a gun and three involved a Back man holding a mundane object (see Table 12). Again, the response options comprised the words TRUE and FALSE.

Table 12
The Stimulus Arrangements Employed in Experiment 5

Sample 1 SAFE	**Sample 2** DANGEROUS
Response Option 1 TRUE	**Response Option 2** FALSE
Targets deemed consistent with Sample 1	**Targets deemed consistent with Sample 2**
White man with beer bottle	Black man with beer bottle
White man with camera	Black man with camera
White man with torch	Black man with torch
White man with gun 1	Black man with gun 1
White man with gun 2	Black man with gun 2
White man with gun 3	Black man with gun 3

Procedure

All participants completed the study in a single experimental session that lasted 30-40mins and none opted for short breaks available at any point. All were presented with the same experimental sequence that involved exposure to the explicit measures followed after a short break by the IRAP.

The IRAP. All procedural aspects of the IRAP employed in Experiment 5 were identical to the previous studies, in the context of the new stimuli. However, because the picture stimuli employed here contained both guns and mundane objects, it was necessary to sub-divide the standard four trial-types into eight. This would permit an analysis of whether participants were discriminating Black and White on the basis of either type of object presented in the pictures. The eight trial-types are presented in Figure 10 and may be summarised as follows: SAFE/White/Mundane; SAFE/White/Gun; SAFE/Black/Mundane; SAFE/Black/Gun; DANGEROUS/White/Mundane; DANGEROUS/White/Gun; DANGEROUS/Black/Mundane; and DANGEROUS/Black/Gun.

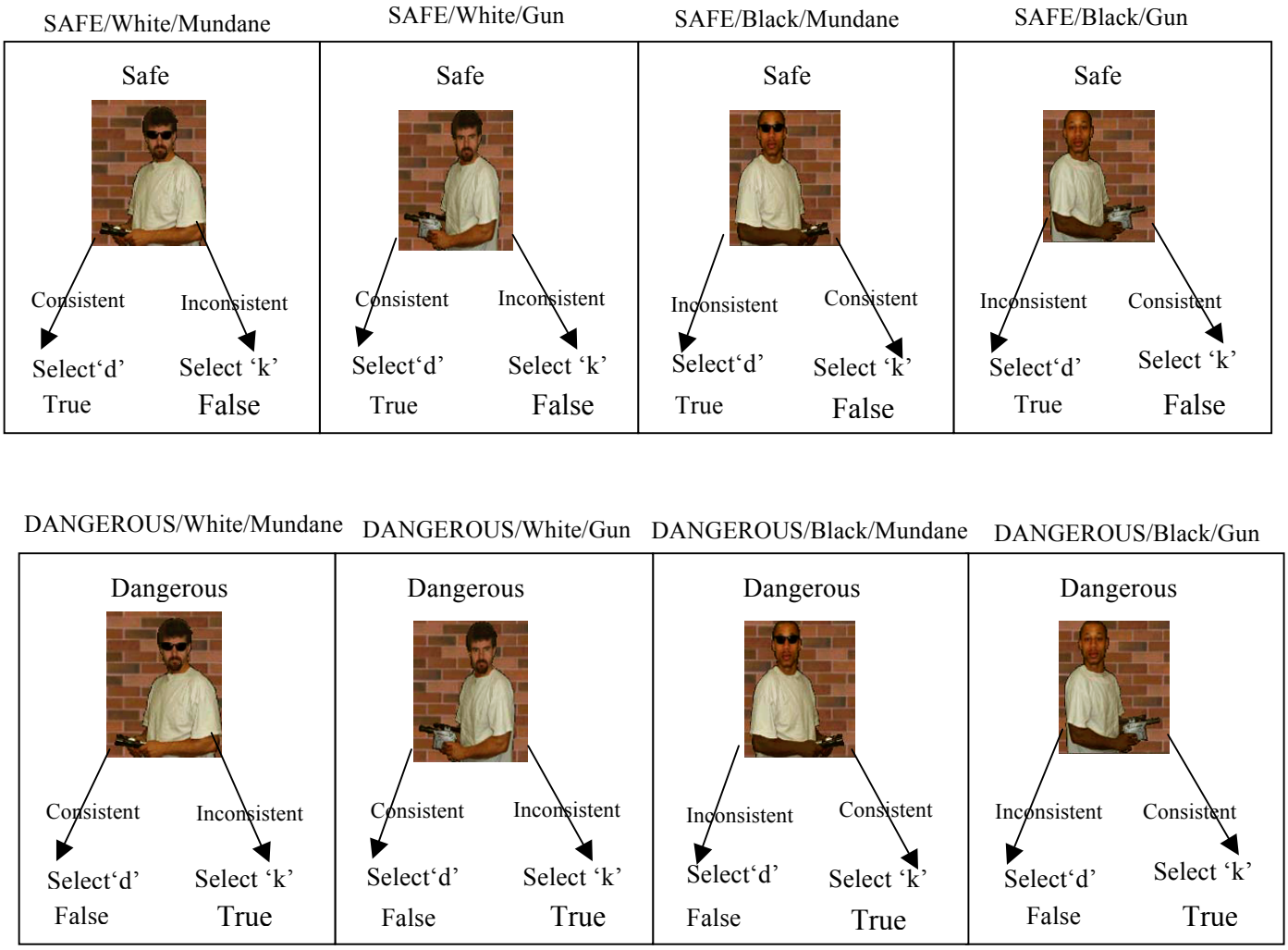


Figure 10. A schematic representation of the eight trial-types in Experiment 5.

In the current study, correct responding on consistent trials *always involved responding to White as safe and Black as dangerous*, irrespective of what item appeared in the pictures (e.g. SAFE/White/Gun/TRUE; SAFE/Black/Mundane/FALSE; DANGEROUS/White/Gun/FALSE; and DANGEROUS/Black/Mundane/TRUE). In contrast, correct responses on inconsistent trials always involved responding to Black as safe and White as dangerous (e.g. SAFE/White/Gun/FALSE; SAFE/Black/Mundane/TRUE; DANGEROUS/White/Gun/TRUE; and DANGEROUS/Black/Mundane/FALSE).

RESULTS

In Experiment 5, the data was analysed in terms of eight trial-types. The mean D-$_{IRAP}$ scores for the eight trial-types were calculated through the same process as the four trial-types before: (i) only response-latency data from test-blocks were used; (ii) latencies above 10,000 ms. were eliminated from the dataset; (iii) all data for participants were removed if they produced more than 10% of test block trials with latencies less than 300ms.; (iv) eight standard deviations for the eight trial-types were computed: one for each trial-type using the response latencies from the three Test Blocks combined; (v) two mean latencies for the eight trial-types in each test block were calculated; (vi) difference scores were calculated for each of the eight trial-types, by subtracting the mean latency of the consistent block from the mean latency of the corresponding inconsistent block; (vii) each difference score was divided by its corresponding standard deviation from step 4, yielding eight D-$_{IRAP}$ scores; one score for each trial-type for each pair of test blocks.

The mean D-$_{IRAP}$ scores for the eight trial-types in Experiment 5 are presented in Figure 11. Responding consistent with predictions (that was, responding that White was safe and Black was dangerous) was indicated in Figure 11 by bars above zero (positive D-$_{IRAP}$ scores) and responding inconsistent with predictions (responding that Black was safe and White was dangerous) was indicated by bars below zero (negative D-$_{IRAP}$ scores). The largest D-$_{IRAP}$ score was recorded for SAFE/White/Mundane and this was in the predicted direction, thus indicating that participants more readily confirmed than denied that White with a mundane object was safe (pro-White bias). A smaller D-$_{IRAP}$ score was recorded for SAFE/White/Gun, but this was not in the predicted direction, indicating that participants more readily denied than confirmed that White with a gun was safe. Responding on SAFE/Black/Mundane was not

noticeably different from zero (hence the lack of data on the graph), thus indicating that participants had mixed views about Black with a mundane object as safe. A small D-$_{IRAP}$ score was recorded for SAFE/Black/Gun, but again this was not in the predicted direction, indicating that participants more readily confirmed than denied that Black with a gun was safe. Responding on DANGEROUS/White/Mundane was barely noticeably different from zero although in the predicted direction, thus indicating that participants had mixed views about White with a mundane object as dangerous. A sizeable D-$_{IRAP}$ score was recorded for DANGEROUS/White/Gun, but again this was not in the predicted direction, indicating that participants more readily confirmed than denied that White with a gun was dangerous. A sizeable D-$_{IRAP}$ score was also recorded for DANGEROUS/Black/Mundane, but again this was not in the predicted direction, indicating that participants more readily denied than confirmed that Black with a mundane object was dangerous. A small D-$_{IRAP}$ score was recorded for DANGEROUS/Black/Gun and this was in the predicted direction, indicating that participants more readily confirmed than denied that Black with a gun was dangerous.

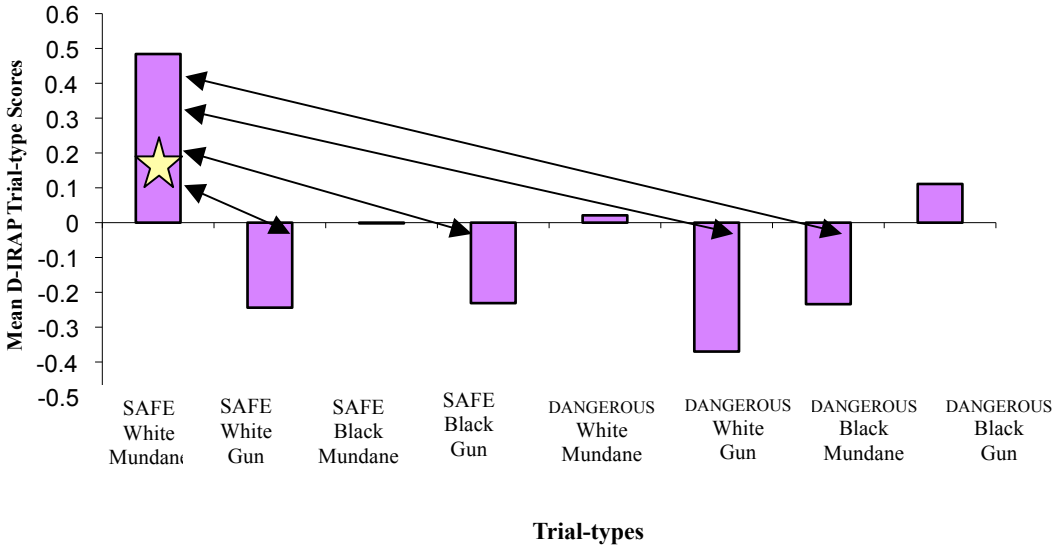


Figure 11. Mean D-$_{IRAP}$ scores for the eight trial-types in Experiment 5. * Indicates that the trial-type was significant relative to zero.

In order to determine if these differences were significant, the $D\text{-}_{IRAP}$ scores were subjected to a 2x8 mixed ANOVA with sequence (consistent-first vs. inconsistent-first) as the between participant variable and trial-type as within-participant variables. This analysis did not include test order, because doing so would have resulted in a sample size that was too small. The results revealed no significant main effects for trial-type or sequence and the interaction effect was also non-significant (all p's > .134).

Eight one sample t-tests were conducted to determine if the $D\text{-}_{IRAP}$ scores for trial-type differed significantly from zero. Only SAFE/White/Mundane was significant ($t(23) = 3.195$, $p = .024$, $\eta_p^2 = 0.19$), all other p's > .137.

IRAP Summary

Only one IRAP effect was significant in Experiment 5, where participants significantly confirmed that White men with mundane objects were safe (pro-White bias). Overall there were more pro-Black effects than expected.

Explicit Measures

The mean outcomes on the explicit measures along with a brief interpretation of each are presented in Table 13. On the majority of the explicit measures, participants indicated the absence of racial bias against Black people (e.g. on the DDS and the MMRS). On the Likert scales, they more readily associated men holding mundane objects as safe and men holding weapons as dangerous, irrespective of race.

Table 13
A Summary and Interpretations of Outcomes on the Explicit Measures in Experiment 5

Explicit Measures	**Mean/Score**	**Participants Range**	**Interpretation of Attitudes**
The Feeling Thermometer-Irish	75	60-100	Medium warmth towards Irish people.
The Feeling Thermometer-Black	75	50-100	Medium warmth towards Black people.
Discrimination Scale	37.8	30-45	Not racially biased against Black people.
Diversity Scale	15.4	11 - 20	Not racially biased against Black people.
Modified Modern Racism Scale (MMRS)	-7	-6 to -9	Not racially biased against Black people.
The Racial Experience Questionnaire Sub-Scales			
Black Acquaintance	100%= Yes 0% = No		All participants knew Black people.
Black Friend	50% = Yes 50% = No		Half of participants had a Black friend.
Black Colleague	66.6% = Yes 33.6% = No		Two thirds of participants had a Black colleague.
Black Relationship	66.6% = Yes 33.3% = No		Two thirds of participants had a close Black relationship.
Black Neighbour	66.6% = Yes 33.3% = No		Almost two thirds of participants had Black neighbours.
Black Encounters	66.6% = Positive 33.3%= Neutral 0% = Negative		Two thirds of participants had positive or equally positive/negative Experiences with Black people
Likert Scales -Mundane	14.5	-30 to 36	Pictures with Mundane objects perceived as safe rather than dangerous.
-Weapon	24.6	0 to -33	Pictures with Weapons perceived as Dangerous rather than safe.

Correlations among Explicit Measures

Statistical analyses (Fisher's r to z) were conducted to determine whether any correlations existed among the explicit measures. Only the DDS sub-scales correlated significantly ($r = .457$, $n = 23$, $p < .012$)

Explicit and Implicit Correlations

Planned comparisons (Fisher's r to z) were performed in order to determine if any significant correlations existed between the implicit (i.e. the final D-$_{IRAP}$ score) and explicit measures. The correlations found are reported in Table 14.

Table 14
Correlations Between the Implicit and Explicit Measures in Experiment 5

Planned Comparisons between Implicit and Explicit Measures	**Correlation**	**P-value**
Correlations Expected to be Positive		
IRAP vs. MMRS	.512	.051
Correlations expected to be Negative		
IRAP vs. REQ 2	-.956	.002
IRAP vs. REQ 5	-.965	.045

The overall D-$_{IRAP}$ score was positively correlated with racist views on the MMRS. The overall D-$_{IRAP}$ score was negatively correlated with the number of Black friends and neighbours, suggesting that the more Black acquaintances participants had in this regard, the less racist they were on the IRAP.

Results Summary

The D-$_{IRAP}$ trial-type scores were only partly consistent with experimental predictions. Specifically, only responding on the SAFE/White/Mundane trial-type indicated the expected pro-White bias and none of the other effects were significant. On closer inspection, however, participants were more pro-Black than expected, for example by indicating that Black with a gun was safe. On practically all of the explicit measures, participants indicated the absence of racial bias against Black people. While, there were numerous correlations amongst the explicit measures that

suggested that they may be measuring the same attitudes, the overall $D\text{-}_{IRAP}$ score did not correlate with the majority of explicit measures.

DISCUSSION

Several IAT studies have reported strong pro-White and anti-Black bias using pictures of Black and White men holding both guns and mundane objects, which we had not yet done with the IRAP (e.g. Greenwald et al., 2003). This was the focus of Experiment 5. The data indicated that the only significant IRAP effect was for White men with a mundane object as safe and all other effects were non-significant. However, closer inspection of the data actually indicated that participants showed a greater than pro-Black bias. Hence, the data here were not consistent with existing IAT studies, but did suggest that the negative influence the guns had on attitudes to the Black men in the pictures in Experiment 3 was softened when Black men held both mundane objects and guns. In short, the data from Experiment 5 more closely resembled those from Experiment 4 than those from Experiment 3.

In all of the IRAP's conducted thus far, correct responding was always based on race such that consistent trials required that participants categorise White as safe and Black as Dangerous, while inconsistent trials require the reverse. In the current series of studies the word-based IRAP in Experiment 2 failed to show an anti- Black bias and this was also the case when the pictures in Experiment 4 contained mundane objects. An anti- Black bias, however, began to emerge in Experiment 3 when the pictures contained guns. Nonetheless, this effect was not recorded in Experiment 5 that presented both guns and mundane objects. Taken together, it was difficult to determine to what extent participants had been responding on the basis of race, even though the experimental contingencies required this. In order to address this issue Experiment 6 replicated Experiment 5 but had the IRAP now presented contingencies

in which correct responding was based on the items held in the pictures are not on the race of the people who held them. Un short, race was not irrelevant and we wanted to determine the potential overlap of findings between this and Experiment 5.

EXPERIMENT 6

In Experiment 6, we created an IRAP in which participants were required to discriminate the objects in the pictures as the basis for correct responding and ignore race. That is, on consistent trials guns were always dangerous and mundane objects were always safe (irrespective of whether the men holding them were Black or White). Correct responding on inconsistent trials was reverse (guns always safe and mundane objects always dangerous), but again race was irrelevant to correct responding.

Participants

A sample of twenty-nine naive participants was recruited for Experiment 6. All were Caucasian Irish undergraduate students at NUIM and were aged between 18 and 23 years old (M: 20 years and 6 months). All participants were recruited through class announcements and notice boards in the Psychology Department. No participants had previous experience of the IRAP and they received no remuneration for their part in the research. All were required to have high fluency in English, as well as normal or corrected to normal vision. In this study, five participants failed to reach criterion in practice trials for the IRAP, hence the data from twenty-four participants are included in the analyses.

Setting

All aspects of the setting were identical to previous studies.

Materials and Apparatus

All aspects of the materials and apparatus from Experiment 6 were identical to Experiment 5.

Procedure

All participants completed the study in a single experimental session that lasted between 30 and 40mins. in total. Although made available to them, none of the participants opted for short breaks at any point. All participants were presented with the same experimental sequence that involved exposure to the explicit measure followed by the IRAP.

All procedural aspects of the IRAP here were identical to Experiment 5, except that correct responding was determined by the items in the pictures, not by the race of the person holding them. That is, on consistent trials guns were always categorised as dangerous and mundane objects as safe (e.g. SAFE/White/Mundane/TRUE; SAFE/White/Gun/FALSE; DANGEROUS/Black/Gun/TRUE; and DANGEROUS/Black/Mundane/FALSE). The reverse was true for inconsistent trials (e.g. SAFE/White/Mundane/FALSE; SAFE/White/Gun/TRUE; DANGEROUS/Black/Gun/FALSE; and DANGEROUS/Black/Mundane/TRUE).

RESULTS

The mean D-$_{IRAP}$ scores for the eight trial-types in Experiment 6 are presented in Figure 12. Responding consistent with predictions (that was, responding that mundane objects were safe and weapons were dangerous) was indicated in Figure 12 by bars above zero (positive D-$_{IRAP}$ scores) and responding inconsistent with predictions (responding that weapons were safe and mundane objects were dangerous) was indicated by bars below zero (negative D-$_{IRAP}$ scores). Responding on each of the eight trial-types will now be described in the order they appear in Figure 12. SAFE/White/Mundane generated only a small D-$_{IRAP}$ score, although it was in the predicted direction, thus indicating that participants more readily confirmed than denied that White with a mundane object was safe. SAFE/White/Gun was only negligibly different from zero and going in the wrong direction, indicating that participants had mixed views about White men with a gun as safe. There was a larger D-$_{IRAP}$ score for SAFE/Black/Mundane that was in the predicted direction, indicating that participants more readily confirmed than denied that a Back man with a gun was safe. The largest D-$_{IRAP}$ score was recorded for SAFE/Black/Gun and again this was in the predicted direction, indicating that participants more readily denied than confirmed that Black with a gun was safe. A small D-$_{IRAP}$ score was recorded on DANGEROUS/White/Mundane and this was not in the predicted direction, indicating that participants more readily confirmed than denied that a White man with a mundane object was dangerous. A large D-$_{IRAP}$ score was recorded for DANGEROUS/White/Gun and this was in the predicted direction, indicating that participants more readily confirmed than denied that a White man with a gun was dangerous. DANGEROUS/Black/Mundane generated a small D-$_{IRAP}$ score that was not in the predicted direction, indicating that participants more readily confirmed than

denied that a Black man with a mundane object was dangerous. The D-$_{IRAP}$ score for DANGEROUS/Black/Gun was negligibly different from zero, indicating that participants had mixed views about whether a Black with a gun was dangerous.

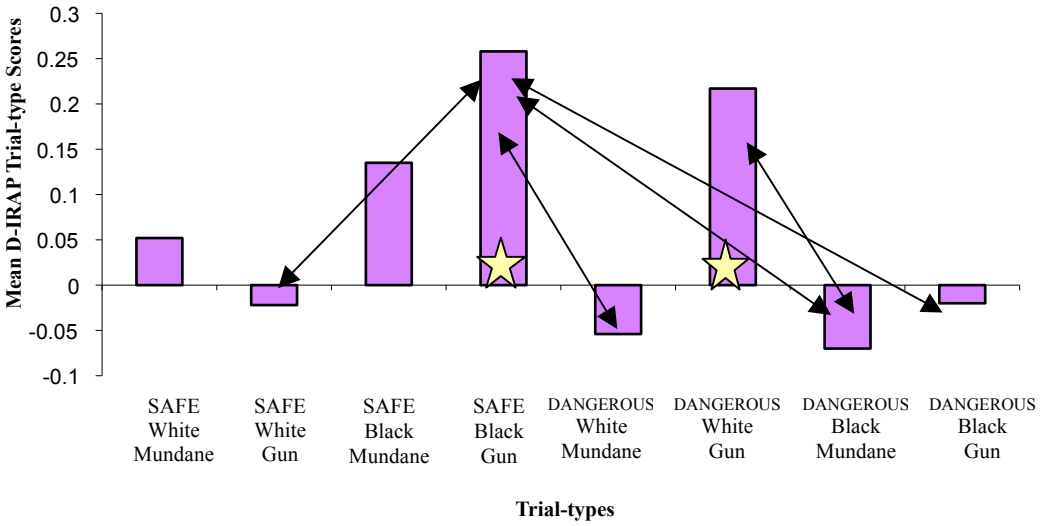


Figure 12. Mean D-$_{IRAP}$ scores for the eight trial-types in Experiment 6. * Indicates that the trial-type was significant relative to zero. Arrow indicates trial-types that were significantly different from each other.

In order to determine if these differences were significant, the D-$_{IRAP}$ scores were subjected to a 2x8 mixed ANOVA with sequence as the between participant variable and trial-type as the within-participant variable. The results revealed no significant main effect for trial-type, sequence or the interaction effect (all p's > .134).

Eight one sample t-tests were conducted to determine if the D-$_{IRAP}$ scores for trial-type differed significantly from zero. Only SAFE/Black/Gun was significant ($t(23) = 2.241$, $p = .035$) and DANGEROUS/White/Gun approached significance ($t(23) = 2.023$, $p = .054$), all other p's > .251.

IRAP Summary

The only significant IRAP effects recorded in the current study were consistent with predictions and indicated that a Black man with a gun is not safe and that a White man with a gun is dangerous, all other effects were non-significant.

Explicit Measures

A summary of the outcomes obtained across the explicit measures are presented in Table 17. On the majority of the explicit measures, participants indicated the absence of racial bias against Black people and on the Likert scales they more readily associated Black people with safe than dangerous.

Table 17
A Summary and Interpretations of Outcomes on the Explicit Measures in Experiment 6

Explicit Measures	**Mean/Score**	**Participants Range**	**Interpretation of Attitudes**
The Feeling Thermometer-Irish	75.6	40-100	High warmth towards Irish people.
The Feeling Thermometer-Black	65.2	40-100	Medium warmth towards Black people.
Discrimination Scale	37	21-48	Not racially biased against Black people.
Diversity Scale	12	5-16	Not racially biased against Black people.
Modified Modern Racism Scale (MMRS)	-5	-10- +1	Not racially biased against Black people.
The Racial Experience Questionnaire Sub-Scales			
Black Acquaintance	75%= Yes 25% = No		Most participants knew Black people.
Black Friend	66.6% = Yes 33.3% = No		Three quarters of participants had a Black friend.
Black Colleague	16.7% = Yes 83.3% = No		One fifth of participants had a Black colleague.
Black Relationship	66.6% = Yes 33.3% = No		Almost two thirds of participants had a close Black relationship.
Black Neighbour	62.5% = Yes 37.5% = No		Almost two thirds of participants had Black neighbours.
Black Encounters	50% = Positive 41.67%= Neutral 8.33% = Negative		The majority of participants had positive or equally positive/negative Experiences with Black people
Likert Scales -Mundane	14.5	-30 to 36	Pictures with Mundane objects perceived as safe rather than dangerous.
-Weapon	24.6	0 to –33	Pictures with Weapons perceived as Dangerous rather than safe.

Correlations among Explicit Measures

Statistical analyses (Fisher's r to z) were conducted to determine whether any correlations existed among the explicit measures. The results are presented in Table 16.

Table 16
Planned Comparisons Among Explicit Measures in Experiment 6

Planned Comparisons between Explicit Measures	**Correlation**	**P-value**
Correlations Expected to be Positive		
Discrimination vs. Diversity	.515	.009
Diversity vs. Black Encounters	.446	.028
Correlations expected to be Negative		
MMRS vs. Discrimination	**-.549**	**.005**
MMRS vs. Diversity	**-.779**	**< .001**
MMRS vs. Black Encounters	**-.496**	**.013**

Again, there was a positive correlation between the DDS sub-scales. The DDS sub-scales also correlated with the MMRS, indicating lack of racism on all three measures. The MMRS was also correlated with number of Black acquaintances suggesting that more acquaintances correlated with lower racism scores.

Explicit and Implicit Correlations

Planned comparisons (Fisher's r to z) were performed in order to determine if any significant correlations existed between the overall $D\text{-}_{IRAP}$ score and the explicit measures. No significant correlations were recorded (all p's > .238).

Results Summary

The data from the current study showed in one respect that participant confirmed that guns were dangerous and not safe, but this effect was not related to race. Overall discriminations on the basis of race or items held were not strong.

DISCUSSION

Experiments 5 and 6 did not show clear patterns of responding like those found in Experiments 2 to 4. One possible explanation noisy patterns of responding was the increased complexity of the target stimuli employed in those experiments. The target stimuli employed in these Experiments 5 and 6 had two critical features whereas in previous experiments the target stimuli has only had one critical feature. For example, in Experiments 3 the target stimuli showed pictures of Black and White men holding weapons, therefore the critical features was race of the man in the picture. In Experiments 5 and 6 the target stimuli showed pictures of Black and White men holding weapons *and* mundane objects, therefore the critical features was race of the man in the picture *and* also the object the man held. The increased complexity of the stimuli would mean it was more difficult for participants to respond quickly and accurately and therefore may account for the noise in the data.

A more probable reason for the noisy data in these experiments is the fact that participants received fewer exposures to each of the target stimuli and thus we did not have enough data to get a clear picture of participant's attitudes to each of the stimuli. This point is illustrated when you consider that in all previous experiments participants received six exposures to the target stimuli in each trial type in each of the three test blocks, a total of eighteen exposures over the course of the IRAP. However, in Experiments 5 and 6 participants only received three exposures to the target stimuli in each trial type in each of the three test blocks, a total of nine exposures over the course of the IRAP. Since these studies were completed a new IRAP software package has become available which will allow participant receive more exposures to the target stimuli. This may dramatically increase the scope for allowing researcher's employ complex stimuli in their experiments**.**

CHAPTER DISCUSSION

Across the five studies of race conducted here using the IRAP, it is fair to say that the effects were mixed. Even in Experiment 2 when participants were required to discriminate Black and White words, there was no anti-Black bias. This pattern was matched in three of the remaining experiments, with the exception of Experiment 3 where an anti- Black bias emerged but this appeared to be largely the result of attentional weapon bias. Overall, the studies generally showed pro-White effects, but anti-Black remained elusive. In the latter two studies the data from the eight trial-types was particularly mixed and showed no clear patterns of responding on the basis or items held in the pictures.

Chapter 4

Using the IRAP to Explore

Clinically Relevant Phenomena

Chapter 4
Using the IRAP to Explore Clinically Relevant Phenomena
Experiments 7 and 8

The study of the self, and beliefs that are potentially relevant to clinical phenomena have traditionally been examined using explicit measures. However, these explicit measures appear to be no less sensitive to the types of bias commonly observed with explicit measures in general (e.g. self presentation bias ect, see de Jong, 2001; Sackheim, &Gur, 1978). As a result, researchers of implicit cognition have argued for the potential role of measures of implicit attitudes in clinical research and practice (de Jong, Pasman, Kindt, & van den Hout, 2001).

Several studies have reported discordance between implicit and explicit measures in assessments of clinical phenomena. For example, de Jong (2002) recorded implicit-explicit discrepancies on self-esteem evaluations in a study of socially anxious women. In short, the women displayed relatively low self-esteem on the explicit measure, but high self-esteem on the implicit measure. In response to this type of discrepancy, Haeffel et al. (2007) have suggested that implicit and explicit cognitions may play separate roles in depressive reactions to stressful life events. Specifically, implicit processes may be critical in determining immediate affective reactions to stress, while explicit processes may be more relevant to long-term depressive reactions. In any case, even this latter evidence points to a potentially important role for the study of implicit cognitions that relate to clinical phenomena.

There is now a considerable number of studies that have used implicit measures to explore clinically relevant phenomena. These include: anxiety (de Jong,

van den Hout, Rietbroek, &Huijdig, 2003); phobias (Teachman, Gregg, & Woody, 2001); eating behaviour (Vartanian, Herman, &Polivy, 2004); paedophilia (Gray, Brown, MacCulloch, & Smith, 2005); alcohol abuse (Wiers, van Woerden, &Smulders, 2002); drug use (Wiers, de Jong, Haverman, &Jelicic, 2004); depression (Haeffel, Abramson, Brazy, Shah, Teachman et al., 2007); and sexual risk behaviour (Stacy, Newcomb, & Ames, 2000).

Several studies have also differentiated participant groups on the basis of implicit attitudes to clinical phenomena. For example, Gray et al. (2005) used the IAT to successfully distinguish between paedophilic offenders and a control group. Specifically, the paedophile sample demonstrated an association between children and sex, while non-paedophile controls showed an association between adults and sex. Similar results were recorded by Nunes, Firestone and Baldwin (2007) who reported that paedophiles viewedchildren as more sexually attractive than non-sexual offenders. Furthermore, the extent to which children were viewed as sexually attractive was correlated with greater risk of sexual recidivism.

Perhaps one of the most promising applications of implicit measures in clinical phenomena is their apparent sensitivity to clinical change. For example, Teachman and Woody (2003) investigated treatment-related changes in the implicit cognitions of spider phobic participants compared to control participants over a course of exposure therapy. The results indicated that associations for spider-fear and spider-disgust were reduced for the phobic participants after treatment, but remained stable across a similar time frame for the controls.

Self-Esteem

Self-esteem is one of the most widely used concepts in psychology. Common definitions include: a self-feeling that is determined by a comparison between the self and the ideal self (James, 1980); an individual's sense of own value or worth (Pelham, & Swann, 1989); the extent to which a person values, approves of, appreciates, prizes, or likes themselves (Blascovich, &Tomaka, 1991); and a favourable or unfavourable attitude toward the self (Rosenberg, 1965).

A body of literature attests to the link between low self-esteem and psychological ill-health, including depression, social anxiety, loneliness and alienation. For example, Kernis, Weisenhunt, Waschull, Greenier, Berry et al. (1998) reported that individuals with fragile self-esteem was less equipped to deal with the hassles of daily life, which in turn resulting in increases in depressive symptoms. On the flip side, Bernard, Hutchison, Lavin and Pennington (1996) reported that high self-esteem correlated positively with self-efficacy, ego strength, hardiness, optimism and maladjustment, and that all of these constructs were significantly related to positive health.

The majority of studies of self-esteem to date have relied almost entirely on explicit methodologies that assume that respondents are giving honest impressions of themselves and that they have enough accurate insight into themselves to be able to do so (Blascovich, et al., 1991). However, explicit measures of self may be particularly sensitive to the types of bias commonly observed with explicit measures in general, including self-presentational bias and self-deception (Sackheim, &Gur, 1978).

As a result, researchers of implicit cognition have argued for the potential role of measures of implicit attitudes in the broader study of the self (Greenwald et al.,

2000). A number of IAT researchers have attempted to use the procedure to examine implicit attitudes to self and to assess the potential overlap between these and explicit outcomes. For example, Farnham, Greenwald and Banaji (1999) presented target words relating to the self (ME) and not relating to the self (NOT ME) in the context of positive and negative attributes. Because they assumed that their sample of 'normal' participants also presented as high in self-esteem on explicit measures, they predicted that these individuals would more readily associate positive attributes with self and negative with not-self. And this is exactly what they observed. In spite of this apparent overlap between the explicit and implicit measures, however, the researchers failed to find any significant correlations when they compared the IAT outcomes with the Rosenberg Self-esteem Scale (RSES: Rosenberg, 1965) and the Self-Affect Scale (SAS). Indeed, they accounted for these findings by suggesting that the implicit and explicit measures in this case accessed different aspects of the self.

The Current Research

Experiment 7 was the first study to employ the IRAP as a measure of implicit attitudes to the self. Participants were required to indicate their agreement or disagreement with a range of statements that reflected high self-esteem (i.e. were positive about the self) or reflected low self-esteem (i.e. were negative about the self). Because the study was also concerned with whether or not the IRAP would correlate with an explicit measure of self, the positive and negative statements presented in the IRAP were taken directly from the RSES and participants also completed this as an explicit measure. We predicted that participants high on explicit self-esteem would also show high implicit self-esteem.

EXPERIMENT 7

Participants

A sample of thirty participants was recruited for Experiment 7. All were undergraduate students at NUIM and were aged between 19 and 23 years old (M: 21 years and 7 months). All participants were recruited through class announcements and notice boards in the Psychology Department. Participants had no experience of the IRAP and received no remuneration for their involvement in the research. All were required to have high fluency in English, as well as normal or corrected to normal vision. In the current study, six participants failed to reach criterion during the practice blocks, hence the data from twenty-four participants were contained in the analyses.

Setting

The setting was identical to the previous studies.

Apparatus and Materials

Experiment 8 involved two basic sets of materials. The explicit measure comprised of the RSES and the implicit measure comprised of the IRAP.

The RSES. The RSES is a measure of global self-esteem consisting of ten statements (copy provided in Appendix I). Five of the statements indicate positive self-regard or high self-esteem (e.g. "I take a positive attitude toward myself"), while the other five indicate negative self-regard or low self-esteem (e.g. "I certainly feel useless at times"). Participants identified with each statement using a four-point Likert scale that ranged from 1/Strongly Agree to 4/Strongly Disagree. The maximum RSES score is 40, the minimum is 0. A score of <15 indicates low self-esteem and a

score >15 indicates high self-esteem. The RSES generally has high reliability with test-retest correlations typically in the range of .82 to .88, and Cronbach's alpha for various samples in the range of .77 to .88 (Blascovich et al., 1991).

The IRAP. The IRAP employed in Experiment 7 was identical in format to previous word-based IRAPs, with the exception of the new stimuli (see Table 17). The sample stimuli comprised of the phrases I AGREE and I DISAGREE. The target stimuli comprised of six statements taken directly from the RSES -- three positive statements that signified high self-esteem (e.g. ON THE WHOLE, I AM SATISFIED WITH MYSELF) and three negative statements that signified low self-esteem (e.g. AT TIMES I THINK I AM NO GOOD AT ALL). The response options were TRUE and FALSE. The previous screen shots employed as visual aids were adapted for current purposes.

Table 17
The Stimulus Arrangements Employed in Experiment 7

Sample 1 I AGREE	**Sample 2** I DISAGREE
Response Option 1 TRUE	**Response Option 2** FALSE
Positive Target Phrases	**Negative Target Phrases**
On the whole, I am satisfied with myself.	At times, I think I am no good at all.
I feel I have a number of good qualities.	I certainly feel useless at times.
I take a positive attitude toward myself.	I wish I could have more respect for myself.
On the whole, I am satisfied with myself.	At times, I think I am no good at all.
I feel I have a number of good qualities.	I certainly feel useless at times.
I take a positive attitude toward myself.	I wish I could have more respect for myself.

Procedure

All participants completed the study in a single experimental session that lasted between 30 and 40mins. in total. Although made available to them, none opted

for short breaks at any point. All participants were presented with the same experimental sequence that involved exposure to the RSES followed after a short break by the IRAP.

All procedural aspects of the IRAP employed in Experiment 7 were identical to the previous studies. The four trial-types employed here were referred to as: I AGREE/Positive (top left of Figure 13); I AGREE/Negative (top right); I DISAGREE/Positive (bottom left); and I DISAGREE/Negative (bottom right). The six exposures to each trial-type comprised of two exposures to each of the three target statements.

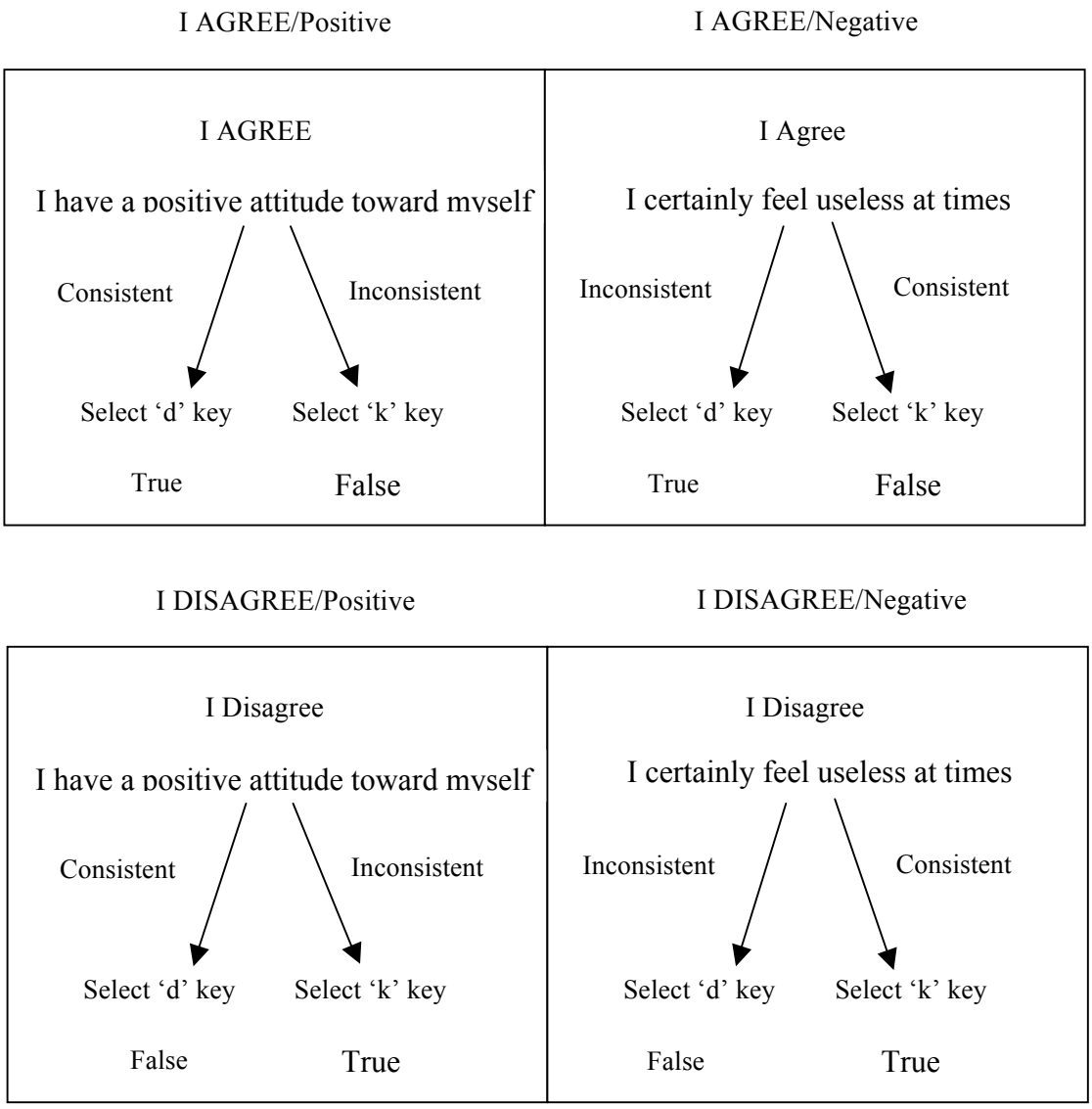


Figure 13. A schematic representation of the four basic trial-types in Experiment 7.

Correct responses on consistent trials involved indicating agreement with positive statements and disagreement with negative statements (i.e. responding here was biased towards high self-esteem). That is, participants were required to respond as follows, for example: I AGREE/I HAVE A POSITIVE ATTITUDE TOWARD MYSELF/TRUE; I AGREE/I CERTAINLY FEEL USELESS AT TIMES /FALSE; I DISAGREE/I HAVE A POSITIVE ATTITUDE TOWARD MYSELF/FALSE; and I DISAGREE/I CERTAINLY FEEL USELESS AT TIMES /TRUE. In contrast, correct

responses on inconsistent trials involved indicating agreement with negative statements and disagreement with positive statements (i.e. responding here was biased towards low self-esteem). That is, participants were required to respond as follows, for example: I AGREE/I HAVE A POSITIVE ATTITUDE TOWARD MYSELF/FALSE; I AGREE/I CERTAINLY FEEL USELESS AT TIMES/TRUE; I DISAGREE/I HAVE A POSITIVE ATTITUDE TOWARD MYSELF/TRUE; and I DISAGREE/I CERTAINLY FEEL USELESS AT TIMES/FALSE.

RESULTS

All analyses and interpretations of the IRAP data were consistent with previous analyses involving four trial-types. As in previous experiments, responding consistent with predictions (that was, responding I agree with positive statements, and the reverse, I disagree with negative statements) was indicated in Figure 13 by bars above zero (positive D-$_{IRAP}$ scores) and responding inconsistent with predictions (responding that I agree negative statements, and the reverse, I disagree with positive statements) was indicated by bars below zero (negative D-$_{IRAP}$ scores).The mean D-$_{IRAP}$ trial-type scores are presented in Figure 13. I AGREE/Positive and I DISAGREE/Positive produced similarly large D-$_{IRAP}$ scores that were both in the predicted direction. That is, participants readily confirmed that they agreed with the positive statements and disconfirmed that they disagreed with them. Responding on I AGREE/Negative produced a lower D-$_{IRAP}$ score, but this was also in the predicted direction. Hence, participants disconfirmed that they agreed with the negative statements. I DISAGREE/Negative yielded a D-$_{IRAP}$ score that was only marginally different from zero, and which was not in the predicted direction. This indicated that

participants had no strong reaction to whether they disagreed with the negative statements.

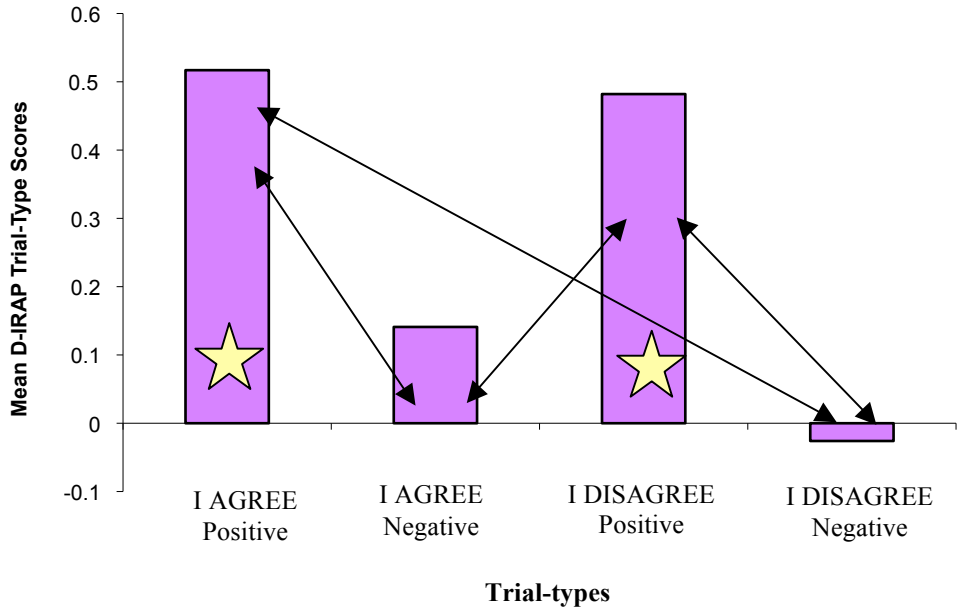


Figure 13. Mean D-$_{IRAP}$ trial-type scores in Experiment 7. * Indicates that the trial-type was significant relative to zero. Arrow indicates trial-types that were significantly different from each other.

A 2x3x4 mixed repeated measures ANOVA revealed a significant main effect for trial-type [$F(3, 66)=4.323$, $p < .001$, $\eta_p^2 = .418$], but not for test or sequence (p's > = .101), and the interaction effects were non-significant (all p's > .102). Fishers's post hoc tests revealed that the significant differences in the trial-types lay between: I AGREE/Positive and I AGREE/Negative ($p = .002$); I AGREE/Positive and I DISAGREE/Negative ($p < .001$); I AGREE/Negative and I DISAGREE/Positive ($p = .005$); and I DISAGREE/Positive and I DISAGREE/Negative ($p < .001$).

Four one sample t-tests were conducted on the D-$_{IRAP}$ trial-type scores in order to determine whether each differed significantly from zero. Only I AGREE/Positive and I DISAGREE/Positive were significant (both p's < .001), the remaining two trial-types were not (both p's > .339).

Explicit-Implicit Correlations

Responding on the RSES ranged from 19-26 (M: 21.95, SD: 1.82), indicating as expected that all participants presented with explicit high self-esteem. A Fishers r to z correlation was calculated to determine if the RSES scores correlated with the overall $D\text{-}_{IRAP}$ score. Although the correlation was positive, it was not significant [r =.148; n = 23; p = .504].

DISCUSSION

The primary aim of Experiment 7 was to assess the utility of the IRAP as an implicit measure of self-esteem and to assess potential correlations between explicit and implicit measures of the same. The research was also the first attempt to modify the IRAP by presenting target stimuli that were whole statements, rather than words or pictures. In order to facilitate greater concordance between the IRAP as an implicit measure of self-esteem and the RSES as an explicit measure, the IRAP presented statements as target stimuli that had been taken directly from the RSES. The findings indicated that the IRAP did appear to tap into self-esteem when responding on two of the target trial-types was in the predicted direction. That is, participants significantly confirmed that they agreed with the positive (high self-esteem) statements and significantly disconfirmed that they disagreed with them. They had no strong reaction to whether they disagreed with the negative statements. Although these findings were largely as expected from participants who all scored as high in self-esteem on the explicit RSES, there were no significant correlations between the two measures.

In Experiment 8, we continued with our investigations of using the IRAP to explore clinical phenomena and extending the target stimuli to full statements. In line with the previous study, we once again attempted to enhance the potential the overlap

between the explicit and implicit measures by taking whole statements from the former and presenting them in the latter. In Experiment 8, we examined implicit attitudes to acceptance and avoidance in a sample of participants who presented with a high orientation towards acceptance on an explicit measure.

EXPERIMENT 8

There is a general consensus among psychologists that clinically relevant phenomena are influenced by individual beliefs and that many forms of psychopathology result from dysfunctional beliefs or misinterpretations of events, emotions, etc. (e.g. Beck, 1976). This general perspective is consistent with RFT, which would argue that the bi-directionality of stimulus relations ensures that there is almost no type of action that does not participate in relational frames with other actions, memories, thoughts, emotions, sensations, etc. Indeed, according to Acceptance and Commitment Therapy (ACT; a program of clinical treatment that is functionally related to RFT), attempts to avoid these relations likely serve only to increase, rather than decrease, the hold our internal content has over our external actions (Wilson, Hayes, Gregg, & Zettle, 2001). Of course, many studies of emotional and cognitive suppression support this view (Wegner, 1994) and suggest, in line with ACT, that the most effective strategy for dealing with problematic internal cognitions is to accept, rather than avoid, them and move on with valued living. That way, the problematic content may not go away, but one has the opportunity to achieve values even if it doesn't (Hayes, & Strosahl, 2004).

The Current Research

In Experiment 8, we used the IRAP to explore implicit attitudes towards acceptance and avoidance. For this purpose, we selected undergraduate students who showed an explicit propensity towards high acceptance/low avoidance on the Acceptance and Action Questionnaire (AAQ-2). In order to enhance the potential overlap between the two types of measure, as we had done in Experiment 7, we took opposing statements directly from the AAQ and inserted them into the IRAP as target stimuli.

Participants

A sample of thirty-four participants was recruited for Experiment 8. All were undergraduate students at NUIM and were aged between 18 and 58 years old (M: 24 years and 9 months). All participants were recruited through class announcements and notice boards in the Psychology Department. Participants had no experience of the IRAP and received no remuneration for their involvement. All were required to have high fluency in English, as well as normal or corrected to normal vision. Participants received no remuneration for their involvement. In the current study, ten participants were excluded from the analyses because they had failed to reach criterion during the practice blocks. This left a sample of twenty-four participants.

Setting

The setting employed here was identical to Experiment 7.

Apparatus and Materials

Experiment 8 involved two basic sets of materials. The explicit measure comprised of the AAQ-2 and the implicit measure comprised of the IRAP.

The AAQ-2 is a written self-report measure of emotional avoidance and acceptance (Hayes, Bissett, Roget, Padilla, Kohlenburg et al., 2004). It comprises of a 10-item list of statements that assess avoidance (e.g. 'Emotions cause problems in my life') or acceptance (e.g. 'It's okay if I remember something unpleasant'). Participants indicate the truth of each statement using a 7-point Likert scale ranging from 1/Never True to 7/Always True. A copy of the AAQ-2 can be found in Appendix J. A high score (maximum 70) indicates high acceptance/low avoidance, whereas a low score (minimum 7) indicates high avoidance/low acceptance. The mean scores for a sample of university students is 50.72 (*SD* = 9.19). The alpha value of the AAQ-II is strong at .83, while the scales also possess good test-retest reliability: .80 (at 3 months) and .78 (at 12 months -- Bond, Hates, Baer, Carpenter, Orcutt, Waltz, et al., manuscript under submission). The overall AAQ score can also be sub-scored into two sub-scores, one for acceptance (maximum=21) and one for avoidance (maximum=49), by collating the scores for each type of statement.

The IRAP employed here was identical in format to Experiment 7, with the exception of the new stimuli. The sample stimuli comprised the phrases I AGREE and I DISAGREE. The target stimuli comprised of six AAQ statements, three categorised as high acceptance/low avoidance (e.g. IT IS OK IF I REMEMBER SOMETHING UNPLEASANT) and three categorised as low acceptance/high avoidance (e.g. WORRIES GET IN THE WAY OF MY SUCCESS). The response options comprised of the words TRUE and FALSE (see Table 18). The screen shots previously employed as visual aids were adapted for current purposes. The six exposures to each trial-type comprised of two exposures to each of the three target statements.

Table 18
The Stimulus Arrangements Employed in Experiment 8

Sample 1 I AGREE	**Sample 2** I DISAGREE
Response Option 1 True	**Response Option 2** False
Target words consistent with I Agree	**Target words consistent with I Disagree**
It's ok if I remember something unpleasant	I'm afraid of my feelings
My thoughts and feelings do not get in the way of how I want to live my life	Emotions cause problems in my life
I am in control of my life	Worries get in the way of my success
It's ok if I remember something unpleasant	I'm afraid of my feelings
My thoughts and feelings do not get in the way of how I want to live my life	Emotions cause problems in my life
I am in control of my life	Worries get in the way of my success

Procedure

All participants completed the study in a single experimental session that lasted between 30 and 40mins. in total. Although made available to them, none opted for short breaks at any point. All participants were presented with the same experimental sequence that involved exposure to the AAQ-2 followed after a short break by the IRAP.

All procedural aspects of the IRAP were identical to the previous studies, with the exception of the new stimuli. In this case, the four basic trial-types were referred to as: I AGREE/Acceptance (top left of Figure 15); I AGREE/Avoidance (top right); I DISAGREE/Acceptance (bottom left); and I DISAGREE/Avoidance (bottom right).

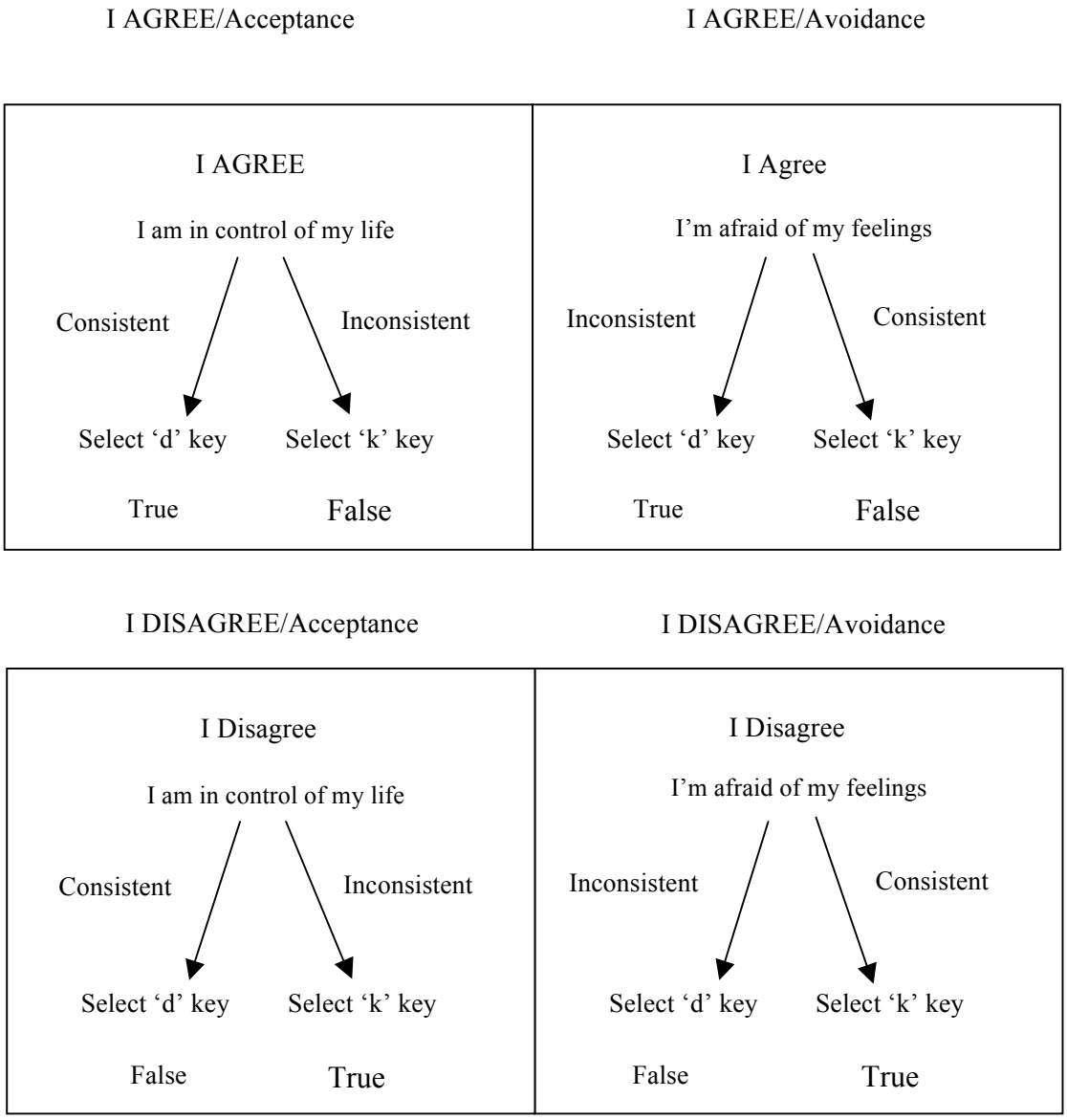


Figure 15. A schematic representation of the four basic trial-types in Experiment 8.

Correct responses on consistent trials involved indicating agreement with acceptance-based statements (i.e. high acceptance/low avoidance) and disagreement with avoidance statements (i.e. low acceptance/high avoidance). That is, participants were required to respond as follows, for example: I AGREE/I AM IN CONTROL OF MY LIFE/TRUE; I AGREE/I'M AFRAID OF MY FEELINGS/FALSE; I DISAGREE/I AM IN CONTROL OF MY LIFE/FALSE; and I DISAGREE/I'M AFRAID OF MY FEELINGS/TRUE. In contrast, correct responses on inconsistent trials involved indicating agreement with avoidance statements and disagreement with

acceptance statements. That is, participants were required to respond as follows, for example: I AGREE/I AM IN CONTROL OF MY LIFE/FALSE; I AGREE/I'M AFRAID OF MY FEELINGS/TRUE; I DISAGREE/I AM IN CONTROL OF MY LIFE/TRUE; and I DISAGREE/I'M AFRAID OF MY FEELINGS/FALSE.

RESULTS

The mean D-$_{IRAP}$ trial-type scores obtained across participants are presented in Figure 16. Again, responding consistent with predictions (that was, responding I agree with acceptance statements, and the reverse, I disagree with avoidance statements) was indicated in Figure 16 by bars above zero (positive D-$_{IRAP}$ scores) and responding inconsistent with predictions (responding that I agree avoidance statements, and the reverse, I disagree with acceptance statements) was indicated by bars below zero (negative D-$_{IRAP}$ scores). The largest D-$_{IRAP}$ score was recorded for I AGREE/Acceptance, with a smaller D score recorded on I DISAGREE/Acceptance, but both were in the predicted direction. Hence, participants confirmed that they agreed with the acceptance statements and disconfirmed that they disagreed with them. Similarly small D scores were recorded for I AGREE/Avoidance and I DISAGREE/Avoidance, but again both of these effects were in the predicted direction. That is, participants disconfirmed that they agreed with the avoidance statements and confirmed that they disagreed with them.

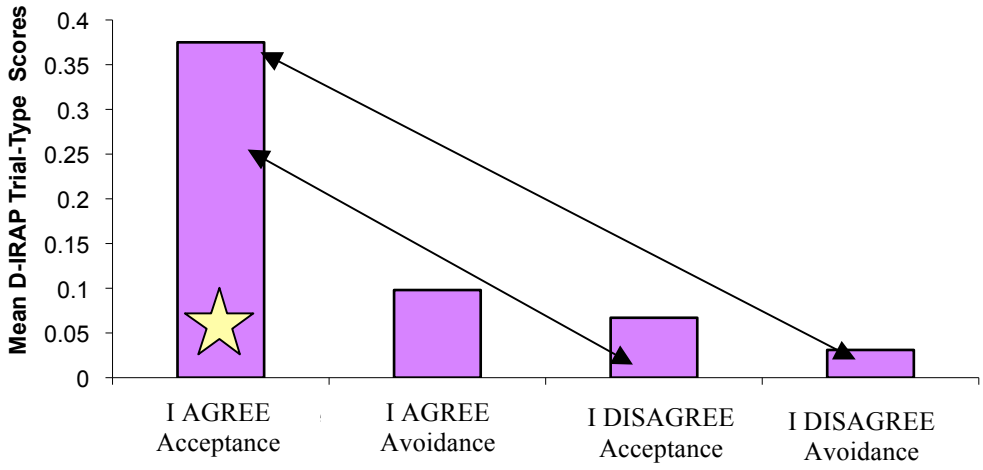


Figure 15. Mean D-$_{IRAP}$ trial-type scores recorded in Experiment 8. * Indicates that the trial-type was significant relative to zero. Arrow indicates trial-types that were significantly different from each other.

The mean D-$_{IRAP}$ trial-type scores were subjected to a 2x3x4 mixed repeated measures ANOVA. Trial-type was significant [$F(3,66)=4.254$; $p = .008$; $\eta_p^2 = .162$], test order, sequence and the interaction effects were not (all p's > .398) were not. Fisher's post-hoc tests indicated the significant trial-type differences lay between I AGREE/Acceptance and I DISAGREE/Acceptance ($p = .025$), and I AGREE/Acceptance vs. I DISAGREE/Avoidance approached significance ($p = .056$) all other p's > .10.

Four one-sample t-tests were used to determine if any of the D-$_{IRAP}$ trial-type scores differed significantly from zero. Only I AGREE/Acceptance was significant [$t(23)=5.236$, $p < .001$], all other p's > .34.

Explicit-Implicit Correlations

Participants' AAQ scores ranged from 28-70 (M: 49.29, SD: 10.76), indicating as expected a propensity towards high acceptance/low avoidance. The sub-scores were as follows: acceptance: M: 14.39, SD: 3.27 and avoidance: M: 35.22, SD: 8.9. A series of Fishers r to z correlations was conducted to determine if the AAQ

data correlated with D_{-IRAP} trial-type scores. The correlation overall proved to be negative and non-significant [r=-.117, p = .589].

DISCUSSION

The primary aim of Experiment 8 was to assess the utility of the IRAP as a measure of implicit attitudes to acceptance and avoidance and to determine the potential correlation between this and the AAQ as an explicit measure of the same. In order to facilitate the overlap between the two types of measure, high vs. low acceptance statements were taken directly from the AAQ and presented as target stimuli within the IRAP. The IRAP outcomes indicated that participants significantly confirmed that they agreed with the acceptance statements and disconfirmed that they disagreed with them (although the latter was not significant). They also disconfirmed that they agreed with the avoidance statements and confirmed that they disagreed with them (but again neither of these was significant). This implicit propensity towards acceptance and away from avoidance was not surprising with this sample of participants who all scored as high acceptance/low avoidance on the explicit AAQ. In spite of this apparent overlap, however, the statistical correlation between the two measures proved non-significant.

Chapter 5

General Discussion

Chapter 5

General Discussion

The aim of the current chapter is to provide a summary of the empirical work presented in Chapters 2 to 4, as well as articulating the main theoretical issues arising from each. A generic issue that spans all aspects of the research programme will then be discussed.

Chapter 2: Summary of Findings from Experiment 1

The first experimental aim of the current research programme was methodological and attempted to determine the stimulus presentation that would produce the strongest IRAP effect. Practically all available IRAP studies to date have employed a presentation format in which both the sample stimuli and response options randomly switch positions across trials within each block. However, there is no empirical evidence to suggest the superiority of this format over any others (e.g. keeping either of these features static).

Experiment 1 of the current research set out to address this simple experimental question. Its main aim was to systematically manipulate the presentation of stimuli on-screen to determine what influence this might have on implicit responding. Naturally, a simple IRAP, rather than one on race for example, was chosen for this purpose. Experiment 1 comprised of four conditions: Moving-Moving (where sample stimuli and response options were both randomised); Moving-Static (where the sample stimuli were randomised, but response options remained in fixed locations); Static-Moving (sample stimuli were fixed and response options were randomised); and Static-Static (sample stimuli and response options were both fixed).

The results of the study were consistent with previous IRAP evidence, with strong and predicted IRAP effects. More importantly, the comparisons across the four conditions indicated that the randomisation of the sample stimuli combined with fixing the response options (Moving-Static) generated the strongest D-$_{IRAP}$ score. However, randomising the samples appeared to be the more critical variable because the D-$_{IRAP}$ score recorded for randomising both samples and response options (Moving-Moving) was only marginally smaller and was also significant. Fixing the samples and randomising the response options (Static-Moving) also yielded a significant effect, although it was smaller than the two previous effects (but not significantly). In contrast, fixing *both* samples and response options (Static-Static) generated a much smaller IRAP effect that was *not* significant.

These outcomes suggested that the consistent use of the Moving-Moving format in IRAP research to date had been wise and had at least to some extent facilitated the strong IRAP effects commonly reported. In line with these findings and the existing literature's use of this format, all subsequent studies in the current thesis thereafter employed the Moving-Moving presentation.

Theoretical Issue Arising From Chapter 2

Stimulus presentation plays an important role in any experimental methodology, but may be particularly important to implicit procedures. This sensitivity results, at least in part, from the delicate balance between the need to make the task challenging (rather than turning it into an explicit measure because the time constraint is loose), while ensuring that the task is do-able. Indeed, several studies have indicated that disruptions in the sequence of implicit learning can arise from alterations in the locations of the target stimuli (Schmidtke, & Heuer, 1997; Stadler, 1995; Tavassoli, &

Lee, 2004). In Experiment 1 we manipulated the screen presentation of sample stimuli and response options, and in doing this manipulated task difficulty. Since the completion of this thesis

new research has emerged that manipulates the maximum response latency allowed in the IRAP, another means of manipulating task difficulty. Barnes-Holmes, Murphy, Barnes-Holmes, & Stewart (in press) have presented evidence that suggests that reducing the maximum response latency allowed on each trial to 2000ms rather than 3000ms may be better at targeting implicit responding. Barnes-Holmes et al conducted an identical experiment to Experiment 3 of the current work under two conditions, one when the maximum response latency was 3000ms and one when the maximum response latency was 2000ms. The results showed that in the 3000ms context that participants were responding that they were pro-White and also pro-Black, in the 2000ms condition significantly greater pro- White and anti- Black stereotyping was revealed. In Experiment 3 of this thesis we found a pro- White, a pro-Black and also an anti- Black bias. While these results show a more mixed attitude toward Black persons at 3000ms than was found by Barnes-Holmes et al, we may also have found a stronger anti-Black bias predicted if we had used 2000ms as a maximum response latency. This research by Barnes-Holmes et al (in press) highlights the necessity to continue experimenting with the experimental features of the IRAP.

The findings from Experiment 1 were perhaps more positive than anticipated in suggesting that only one method of organising the sample stimuli and response options (i.e. Static-Static) failed to generate a significant IRAP effect, although even this was in the predicted direction. These outcomes were perhaps more important for the IRAP than for the IAT, for example, because the screen presentations in the

former are simple and do not change across trial blocks, as is the case for the IAT. Hence, although the data here did not suggest that great disruption of effects would arise from the use of any of the four presentation formats, it did suggest that significant effects are facilitated more by some formats than others.

The fixed or random nature of the stimuli presented in the IRAP also raises the related issue about the degree of flexibility that participants need to have, particularly in switching across blocks between consistent and inconsistent responding. The data here suggest that the greater the randomisation (e.g. Moving-Moving), the better the IRAP effect and the less the randomisation, the weaker the effect (Static-Static). Hence, one might assume that greater flexibility within the screen presentations *across trials* permits greater flexibility in responding *across blocks,* hence improved accuracy and/or speed.

There is recent empirical evidence that supports the importance of flexibility in this regard. O'Toole and Barnes-Holmes (in press) used the IRAP to target similar/different and before/after relations in conjunction with the explicit Kaufman Brief Intelligence Test. Consistent with their predictions, participants who produced higher scores on the intelligence test responded more quickly on the IRAP. Specifically, the inconsistent IRAP trials produced a larger number of significant correlations with the explicit measure than consistent trials. In other words, participants who performed better on the intelligence test were not only faster at IRAP responding, but also demonstrated a greater degree of relational flexibility. It is reasonable to assume, therefore, that this flexibility across blocks may be facilitated by requiring considerable flexibility in responding to rapidly changing features of screen presentations across trials. The current findings support this view.

Chapter 3: Summary of Findings from Experiments 2-6

The second aim of the current thesis was to determine the utility of the IRAP as a measure of implicit attitudes to race and this was the core focus of Experiments 2 to 6 reported in Chapter 3. In the IRAP in *Experiment 2*, participants were simply required to relate the phrases WHITE PERSON and BLACK PERSON to a range of positive and negative terms that could be readily categorised as safe and dangerous. In this study, we predicted that Irish participants would readily relate White with safe and Black with dangerous, hence showing a pro-White and anti-Black bias.

The data indicated that responding on the White trial-types showed the expected pro-White bias where White was safe and not dangerous, although the latter effect was not significant. Unexpectedly, responding on the two Black trial-types was contrary to predictions and failed to show an anti-Black bias. Specifically, participants significantly confirmed (rather than denied) that Black was safe and had mixed views about Black as dangerous. Evidence from explicit measures was consistent with this lack of anti-Black bias. But in spite of this apparent overlap, the overall $D\text{-}_{IRAP}$ score failed to correlate with many of the explicit measures, although the latter correlated with each other.

Several IAT studies of race have reported stronger pro-White and anti-Black bias with the presentation of pictures, rather than words. Hence, *Experiment 3* presented a picture-based race IRAP to determine if this would result in any improvement on the IRAP outcomes, particularly with regard to an anti-Black bias. However, in order to assess the possible role of attentional weapon bias that had also been observed in IAT evidence on race, and to create the possibility that this might also facilitate anti-Black bias, both the Black and White men in the pictures were holding guns. Nonetheless, the experimental contingencies remained as before, with

correct responding based entirely on race, rather than on the guns (in any case, all pictures contained guns). We simply wanted to determine whether pictures of Black and White men with guns would enhance the weak IRAP effects recorded with words.

The results indicated significant D-$_{IRAP}$ scores for responding on SAFE/White and SAFE/Black trial-types, but the former and not the latter was in the predicted direction. That is, participants confirmed that both White and Black men with guns were safe, showing a pro-White and pro-Black bias even when they held guns. However, participants' implicit views diverged to some extent on whether the Black and White men holding guns were dangerous. Specifically, participants had mixed views on whether White men with guns were dangerous, but they significantly confirmed that Black men with guns were dangerous. The effect for danger, therefore, showed an anti-Black bias. In short, the anti-Black bias recorded here (but not in Experiment 2) suggested that the presence of the guns had, to some extent, negatively influenced participants' implicit attitudes to the Black men in the pictures. Again, however, the majority of explicit measures reflected the absence of explicit racial bias and in spite of a number of significant correlations among explicit measures, none correlated with the IRAP.

In order to confirm that the anti-Black effect in Experiment 3 had resulted from the guns in the pictures, *Experiment 4* presented pictures of the same Black and White men holding mundane objects. If the guns had not enhanced the anti-Black bias, then we predicted that the outcomes here would be consistent with those from Experiment 2. Again, correct responding was always based on race, not on the items held.

The results indicated that participants significantly confirmed that both the White and Black men holding mundane objects in the pictures were safe. Similarly,

they did not differentiate between the two races in terms of dangerousness. That is on both the DANGEROUS/White and Dangerous/Black trial-types they denied that the White and Black men were dangerous. In short, the anti-Black bias from Experiment 3 was lost. This outcome did suggest that the guns presented in the previous study had negatively biased participants against the Black men in the pictures as dangerous. Again, on the majority of explicit measures, participants indicated the absence of racial bias and although the explicit measures did correlate with each other, none correlated with the IRAP.

In order to further examine the potential effect of the guns on anti-Black bias and in line with previous IAT studies of race, *Experiment 5* presented pictures of Black and White men holding guns *and* mundane objects. However, it is important to remember that correct responding was again based on race, not on the items held in the pictures. We wanted to determine whether the guns and mundane objects would differentially influence discriminations of race and to what extent to combination of the two types of items would yield effects that overlapped with those recorded in the previous experiments (in which only one type of item was held). In order to examine the relative impact of each type of held item, the IRAP analyses were further sub-divided into eight (rather than four) trial-types.

The majority of effects were weak. Specifically, only responding on SAFE/White/Mundane was significant and indicated that participants confirmed that White men with mundane objects were safe. Interestingly, however, they had mixed views on whether Black men with mundane objects were safe, thus showing some evidence of a pro-White and anti-Black bias in this regard. Unexpectedly however, when asked about White and Black men with mundane objects as dangerous, participants had mixed views about White as dangerous, but denied that Black was

dangerous. Similarly, when asked about Black and White men holding guns as safe, participants unexpectedly confirmed that White was not safe, but Black was safe. Finally, when asked about White and Black men holding guns as dangerous, participants confirmed that both White and Black were dangerous. Hence, although the majority of effects were non-significant, there was greater than expected evidence of pro-Black bias, irrespective of whether the items held were guns or mundane objects.

In *Experiment 6*, participants were presented with the same IRAP as the previous study, but they were now required to discriminate the objects in the pictures as the basis for correct responding and ignore race. That is, on consistent trials guns were always dangerous and mundane objects were always safe and correct responding on inconsistent trials was reverse (guns always safe and mundane objects always dangerous). Our aim here was to determine whether this would yield a different outcome from the previously weak but pro-Black effects. Again, the majority of effects overall were weak. The only significant effects were recorded on SAFE/Black/Gun and DANGEROUS/White/Gun. Hence, participants significantly denied that the Black men with guns were safe (but had mixed views on White with guns as safe) and significantly confirmed that White men with guns were dangerous (but had mixed views on Black with guns was dangerous). In addition, they confirmed that White and Black men with mundane objects were safe. They also unexpectedly confirmed that White and Black men with mundane objects were dangerous. This weak array of effects was consistent with Experiment 5, but showed no obvious bias in either direction.

Theoretical Issues Arising from Chapter 3

In general the IRAP race outcomes here were not entirely consistent with IAT evidence, particularly with regard to anti-Black bias. On several occasions, we found sound evidence of pro-White bias, but struggled to record anti-Black bias, with the only exception for the latter being the presentation of guns only in Experiment 3. One possible reason for the discordance of evidence may concern the fact that the majority of IAT studies are conducted with American participants, whereas participants here were all Caucasian Irish. Indeed, the US is a historically multi-cultural society, but has an equally long history of Black slavery and Black discrimination. As a result, perhaps this climate has served to strengthen anti-Black bias and pro-White bias simultaneously (see Dovidio, Kawakami, & Beach, 2001). Indeed, Dovidio and Gaertner (2003) argued that social categorisation by race is automatic for US citizens.

By contrast, Ireland is historically a more homogeneous culture that only very recently was subject to considerable inward migration, particularly from persons native to the African continent. Hence, we have as yet not experienced levels of migration of Black persons or related variables that contribute to the types of anti-Black bias demonstrated by American participants of IAT studies. Interestingly, anti-Black only emerged in our research in the context of threat and even Irish participants have considerable exposure to American television where crime is almost synonymous with Black, but not White, persons. As a result, a visual context involving weapons would perhaps be adequate to capturing anti-Black bias in Irish participants, but a less threatening context would not. Indeed, the weak and mixed effects on the remaining four IRAP experiments here support this view.

It might be argued that the anti-Black effect recorded only with the presentation of guns in Experiment 3 is evidence of attentional bias that enhanced

anti-Black effects. This would be consistent with evidence reported by Duncan (1976) who demonstrated that even mildly aggressive behaviour was perceived as more threatening when performed by a Black person than by a White person (see also Sagar, & Schofield, 1980). Attentional *weapon* bias is a particularly well-established effect in the context of eye-witness testimony and several implicit studies have demonstrated that the presence of weapons serves to enhance anti-Black bias (Payne, Lambert, & Jacoby, 2001). The IRAP data from Experiment 3 indicate a similar effect and suggests that the weapons influenced the discriminations of race, even though they were actually irrelevant to correct responding.

One of the key contrasts between Experiments 2 and 3 attempted to determine which would generate better IRAP effects for racial prejudice. Although the data favoured the latter in this regard, it was difficult to determine whether this resulted from the use of pictures per se, or the presentation of weapons. To some extent, the data from Experiment 4 in particular resolved this issue and indicated that racial pictures per se do not necessarily generate strong IRAP effects.

However, several IAT studies have indicated that stronger race effects are generally observed with pictures, rather than words. For example, Dasgupta et al. (2000) reported that faces of Black vs. White persons produced stronger racial bias on the IAT compared to Black vs. White names. According to Greenwald (1999), IAT effects are not as strong for words as pictures because pictures are aided by visual discriminations, while words are not. The current research was the first IRAP work to present picture stimuli, rather than words, so more research will be needed to determine which generates the better effects. To some extent, however, this will also likely depend on the types of implicit attitudes that are targeted.

The use of pictures, however, creates the added possibility that the outcomes do not reflect genuine racial preferences, but simply participants' greater familiarity with White faces over Black. This suggestion is supported by findings in which frequent exposure to previously neutral stimuli enhances self-reported preference for those stimuli (Zajonc, 1968). However, several studies suggest that this possibility may not explain IAT effects for race. First, Ottaway, Hayden and Oakes (2001) controlled for familiarity while assessing attitudes to Black, White and Hispanic names and found that White participants still expressed automatic preferences for White names over the others. Second, American participants have been found to exhibit greater automatic preferences for unfamiliar American faces than familiar Russian faces (Rudman, Greenwald, Mellott, & Schwartz, 1999). Third, White Americans who have had extensive exposure to Black and other racial groups, still show significantly racist IAT effects. Taken together, these facts suggest that obvious familiarity with the in-group cannot discount in-group vs. out-group preferences.

Another issues relating to the to use of picture stimuli when investigating race is highlighted by Eberherdt et al. This study shows that participants show more prejudice to Black persons with very black skin compared to Black persons of relatively pale skin. In Experiments 3 to 6 the picture stimuli we employed depicted Black men with relatively pale sink. Given the research by Eberherdt et al it is likely that we would have found a stronger anti-Black effect if we had employed picture stimuli depicting very Black men.

Perhaps a better question for the IRAP research program here concerns the impact of the safe/dangerous dimension we chose to target in all five studies of race. Indeed, we selected this over, for example, the more basic good vs. bad distinction, because we thought that the more extreme dimension would generate stronger implicit

outcomes. Our data suggest that this was not necessarily the case, although it is difficult to tell what role the dimension played here, because we did not systematically compare White and Black on any other dimension. However, perhaps our selection was counter-productive in that the specificity of safe vs. dangerous requires more experience of different racial groups in order to create a strong IRAP effect, but more simplistic categorisations such as White-good/Black-bad may create these effects more readily. This view would lend support to our previous argument that our race IRAP did not show as strong anti-Black effects as the IAT because we involved Irish rather than American participants. Again, more IRAP research will be needed to separate out the different magnitude of effects that may be obtained when racial groups are compared along different dimensions and with different participant groups.

Although the weak IRAP effects in Experiments 5 and 6 are perhaps most readily accounted for in terms of unacceptably small cell sizes, the sub-division of the more common four trial-types into eight was a useful addition to these studies involving complex visual arrays. Without analysis at this level, it would have been very difficult to tease apart the potential influence of the two types of items held by the men in the pictures. Indeed, analysis at this level would have been basically impossible without using the D-IRAP algorithm, which appears to offer a significant analytic advance over previous analyses of the response latency data.

Another issue for discussion that arose in Chapter 3 was the question of explicit measures. Five explicit measures were employed in Chapter 3 (the Modified Modern Racism Scale (MMRS), the Discrimination and Diversity Scale (DDS), the Racial Experience Questionnaire (REQ), the Feeling Thermometer and a series of Likert scales). We considered it necessary to use all of these measures as each

measure gave us a different perspective of participant's attitudes to race. The Modern Racism Scale (MRS) and the Discrimination and Diversity Scale (DDS) are two of the explicit measures of racial attitudes that have been most commonly used in conjunction with implicit measures of race and therefore allow us to compare our research(e.g., Dasgupta et al., 2000; Greenwald et al., 1998; Wittenbrink et al., 1997). In Chapter 2, we employed a modified version of the MRS, which is shorter than the original version and is more relevant to an Irish population. We also employed the DDS, a measure of discrimination on the basis of race and also openness to embracing cultural diversity in our country. By using these measures we had the benefit of being consistent with established research.

The feeling thermometer and Likert measures give a reflection of participant's global or general attitudes. These measures were included as there is evidence to suggest that implicit measures correlated more frequently with explicit measures of global attitudes such as semantic differential ratings or feeling thermometers than to more complex policy related measures, like the MRS (Dovidio, Kawakami, & Beach 2003). This is illustrated by Dasgupta, McGee, Greenwald, & Banaji (2000) who found that two IAT measures of racial prejudice correlated more strongly with semantic differential (rs = .24 and .38) and feeling thermometer (rs = .23 and .21) measures than with the MRS (rs = -.13 and .02) and a diversity scale (rs = -.02 and.14). This effect has also been found by Greenwald and Farnham (2000), who found a significant correlation between implicit self-esteem and feeling thermometers (r = .06) but no significant correlations with any other explicit measure (average r = .17). Likewise, in IRAP research Vahey, Barnes-Holmes, Barnes-Holmes, & Stewart (in press) showed a correlation between implicit self-esteem and feeling thermometers.

The REQ was a closed questionnaire was designed specifically for our experimental purposes. The purpose of this measure was to explore the relationship between racial bias and actual exposure contact with members of another race. Each of these explicit measures was important in its own way, although in some respects it could be argued that the DDS was one of the most important measures to use in the current research. The DDS is the only measure used in this research that is identical to that used in other implicit studies, as we used a modified version of the MRS and the Likert measures, feeling thermometers and REQ were specific to this research.

Chapter 4: Summary of Findings from Experiments 7 and 8

The third aim of the current thesis was to examine the utility of the IRAP as a measure of clinically relevant implicit cognitions. In *Experiment 7,* participants were required to indicate their agreement or disagreement with a range of statements that reflected high self-esteem (i.e. were positive about the self) or reflected low self-esteem (i.e. were negative about the self). Because the study was also concerned with whether or not the IRAP would correlate with an explicit measure of self, the positive and negative statements presented in the IRAP were taken directly from the Rosenberg Self-Esteem Scale (RSES) and participants also completed this as an explicit measure.

The findings indicated that participants significantly confirmed that they agreed with the positive (high self-esteem) statements and significantly disconfirmed that they disagreed with them. They also disconfirmed that they agreed with the negative statements, although this effect was non-significant. They had no strong reaction to whether they disagreed with the negative statements. Although these findings were largely as expected from participants who all scored as high in self-

esteem on the explicit RSES, there were no significant correlations between the two measures.

In *Experiment 8*, we used the IRAP to explore implicit attitudes towards acceptance and avoidance. For this purpose, we selected undergraduate students who showed an explicit propensity towards high acceptance/low avoidance on the Acceptance and Action Questionnaire (AAQ-2). In order to enhance the potential overlap between the two types of measure, as we had done in Experiment 7, we took opposing statements directly from the AAQ and inserted them into the IRAP as target stimuli. The results indicated that participants significantly confirmed that they agreed with the acceptance statements, but all other effects were non-significant. Closer inspection of the trial-type data, however, were as expected and showed that participants disconfirmed that they agreed with the avoidance statements and confirmed that they disagreed with them. In spite of this overlap of implicit evidence with the outcomes on the AAQ, there was no significance correlation between the two measures.

Theoretical Issue Arising From Chapter 4

Participants in Experiments 7 and 8 appeared to find it more difficult to respond to negative statements than positive and this finding is consistent with the literature on reasoning and problem-solving, which provides evidence for a "top-down" cognitive processing bias toward confirming rather than disconfirming a postulate (see Eysenck & Keane, 2000, for a review). Hence, increased cognitive load on these trials may have reduced automaticity, thus weakening IRAP effects.

Theoretical Issue Arising from the Thesis

One of the secondary aims of the current research program was to examine the relationship between implicit and explicit measures. This is a common aim in all published and unpublished IRAP studies and practically all published IAT studies.

There is considerable debate within the field of implicit cognition concerning whether or not researchers 'should' expect to find explicit-implicit correlations in a given context and in simple terms the general trend is that sometimes there are correlations and sometimes there are not. Indeed, there is not even clear evidence of a pattern one way or the other in the measurement of specific attitudes. According to Greenwald et al. (2003), the greater the concordance between explicit and implicit measures, the greater the validity of what is being measured. Alternatively, Dovidio and Fazio (1992) have argued that the magnitude of the correlation depends upon participants' motivation and opportunity to deliberate, such that high motivation and opportunity produce low explicit-implicit correlations. This latter view has considerable support from the facts that some of the weakest correlations in the area are reported in the measurement of the most socially-sensitive attitudes and it is well-established that participants have a strong desire to try to control self-perceived prejudices (Ziegert & Hanges, 2005). Hence, it is not surprising that discrepancies between explicit and implicit measures are observed in the context of studying race, as was the case here for the majority of the data.

Discordance between explicit and implicit measures has led to further questions about what is actually being targeted across the different methodologies (Cameron, Alvarez, & Bargh, 2000). Some researchers, for example, have argued that the different measures simply target different evaluations of the same attitude and Fazio and Olson (2003) suggested that this difference results from the fact that

implicit attitudes are probably not available to conscious processing or introspection. Indeed, Greenwald et al. (1998) thus described the two types of evaluations as 'dissociated', while Wilson (2000) referred to them as 'independent'. Blaire (2001) has even gone so far as to suggest that implicit and explicit attitudes are independent cognitive representations that result from different psychological processes. There is considerable support for the idea that implicit and explicit cognitions operate like a dual processing system. Advocates of this argument include Devine (1989), Haeffel, Abramson, Brazy, Shah, Teachman and Nosek, (2007) and Neely (1977).

Structural similarity has also been shown to be a variable that effects correlations between measures. For example, Ajzen and Fishbein (1977) reported that structural similarity increased correlations between different explicit measures. However, the findings from the last two experiments here suggest that perhaps this relationship does not hold across explicit and implicit measures, when the two failed to correlate even though the statements from the explicit measure comprised the target stimuli in the implicit measure.

A behaviour analytic and RFT account of the relationship between implicit and explicit attitudes is referred to as the Relational Elaboration and Coherence (REC) Model. The REC model assumes that an instance of implicit responding, such as an IRAP effect, when produced under appropriate time-pressure, is driven largely by immediate and relatively brief relational responses, whereas explicit measures reflect extended and coherent relational networks.

Concluding Comments

The current thesis was ambitious in its aim to conduct three different types of IRAP research. The data from Experiment 1 were as expected and showed strong

IRAP effects for all conditions, as well as offering insight into the screen presentation format that facilitates significant effects. The data from the more ambitious later sets of experiments were not as straightforward. That is, while we often recorded effects that we had predicted, we equally often failed to record effects that we expected and on some occasions we even recorded effects that we had not expected. This variability was particularly apparent in the race studies. On balance, however, this was a complex array of studies that presented complex visual arrays on screen. Our findings on the clinical phenomena were more positive and generally more in line with predictions. Taken together, therefore, the IRAP did appear to meet a range of challenges as a measure of implicit attitudes to race, self-esteem and acceptance and demonstrated that the future for the methodology as a tool for measuring implicit attitudes is promising.

References

References

Ahern, A.L., & Hetherington, M.M., (2006). The thin ideal and body image: An experimental study of Implicit Attitudes. *Psychology of Addictive Behaviors*, *20*, 338–342.

Ajzen, I., & Fishbein, M. (1977). Attitude–behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84,* 888–918.

Amodio, D.M., & Devine, P.G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, *91,* 652-661.

Barnes, D. (1994). Stimulus equivalence and relational frame theory. *The Psychological Record*, *44*, 91-124.

Barnes, D. (1996). Naming as a technical term: Sacrificing behavior analysis at the altar of popularity. *Invited commentary for the Journal of the Experimental Analysis of Behavior*, *65*, 264-267.

Barnes, D., Browne, M., Smeets, P.M., & Roche, B. (1995). A transfer of functions and a conditional transfer of functions through equivalence relations in three to six year old children. *The Psychological Record*, *45,* 405-430.

Barnes, D., & Holmes, Y. (1991). Radical behaviorism, stimulus equivalence, and human cognition. *The Psychological Record, 41*, 19-31.

Barnes, D., & Keenan, M. (1993). Concurrent activities and instructed human fixed-interval performance. *Journal of the Experimental Analysis of Behavior,* 59, 501-520.

Barnes, D., & Roche, B. (1996). Relational frame theory and stimulus equivalence are fundamentally different: A reply to Saunders' commentary. *The Psychological Record, 46,* 489-507.

Barnes-Holmes, Y., Barnes-Holmes, D., Smeets, P.M., Strand, P., & Friman, P. (2004). Testing and training relational responding in accordance with the relational frame of opposite in young children. *International Journal of Psychology and Psychological Therapy*, *4,* 531-558.

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., Cochrane, A., Barnes-Holmes, Y., & Stewart, I. (2004). "Offer it up" and psychological acceptance: Empirical evidence for your grandmother's wisdom. *The Irish Psychologist, 31*, 72-78.

Barnes-Holmes, D., Dymond, S., Roche, B., & Grey, I. (1999). Language and cognition. *The Psychologist*, *12*, 500-504.

Barnes-Holmes, D., Murtagh, L., Barnes-Holmes, Y., & Stewart, I. (in press). Using the Implicit Association Test and the Implicit Relational Assessment Procedure toMeasure Attitudes Towards Meat and Vegetables in Vegetarians and Meat-Eaters. *Psychological Record.*

Beck, A.T. (1976). *Cognitive therapy and emotional disorders*. New York: International University Press.

Baron, A.S., & Banaji, M.R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 to 10 and adulthood. *Psychological Science, 17*, 53-58

Bernard, L.C., Hutchison, S., Lavin, A., & Pennington, P. (1996). Ego-strength, hardiness, self-esteem, self-efficacy, optimism, and maladjustment: Health-related personality constructs and the "Big Five" model of personality. *Assessment*, *3,* 115-131.

Blascovich, J., & Tomaka, J. (1991). Measures of self-esteem. In J. P. Robinson, P. R. Shaver, and L. S. Wrightsman, (Eds*.) Measures of personality and social psychological attitudes*, (*Volume* I, pp. 115-160). San Diego, CA: Academic Press.

Bond, F.W., Hayes, S.C., Baer, R.A., Carpenter, K.M., Orcutt, H.K., Waltz, T., et al. (2008). *Preliminary psychometric properties of the Acceptance and Action Questionnaire-II: A revised measure of psychological flexibility and acceptance.* Manuscript submitted for publication.

Correll, J., Park, B., Judd, C.M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology, 83,* 1314-1329.

Cullen, C., & Barnes-Holmes, D. (in press). Implicit pride and prejudice: A heterosexual phenomenon? In T.G. Morrison, and M.A. Morrison (Eds.), *Modern Prejudice*. New York: Nova Science.

Cullinan, V., Barnes-Holmes, D., & Smeets, P.M. (2000). A precursor to the relational evaluation procedure: Analyzing stimulus equivalence II. *The Psychological Record, 50*, 467-492.

Cullinan, V., Barnes-Holmes, D., & Smeets, P.M. (2001). A precursor to the relational evaluation procedure: The search for the contextual cues that control equivalence responding. *Journal of the Experimental Analysis of Behavior, 76,* 339-349.

D'Amato, M.R., & Colombo, M. (1985). Auditory matching-to-sample in monkeys (*Cebus apella*). *Animal Learning and Behavior, 13,* 375–382.

D'Amato, M.R., & Worsham, R.W. (1974). Retrieval cues and short term memory in Capuchin monkeys. *Journal of Comparative and Physiological Psychology, 86*, 274-282.

Dasgupta, N., McGhee, D.E., Greenwald, A.G., & Banaji, M.R. (2000). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology, 36,* 316-328.

Dawson, D.L., Barnes-Holmes, D., Gresswell, D.M., Hart, A.J.P., & Gore, N.J. (in press). Assessing the implicit beliefs of sexual offenders using the Implicit Relational Assessment Procedure: A first study. *Sexual Abuse: A Journal of Research and Treatment.*

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

de Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology, 50,* 77-85.

de Jong, P.J. (2002). Implicit self-esteem and social anxiety: Differential self-positivity effects in high and low anxious individuals. *Behaviour Research and Therapy, 40,* 501-508.

de Jong, P.J., Pasman, W., Kindt, M., & van den Hout, M.A. (2001). A reaction time paradigm to assess complaint-specific (implicit) dysfunctional beliefs. *Behaviour Research and Therapy, 39*, 101–113.

de Jong, P.J., van den Hout, M., Rietbroek, H., & Huijdig, J. (2003). Dissociations between implicit and explicit attitudes toward phobic stimuli. *Cognition and Emotion*, *17*, 521-545.

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

Donahoe, J.W., & Palmer, D.C. (1994). *Learning and complex behaviour*. Needham Heights, M.A.: Allyn and Bacon.

Duncan, B.L. (1976). Differiental social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of blacks. *Journal of Personality and Social Psychology, 34,* 590-598.

Dunham, Y., Baron, A.S., & Banaji, M.R. (2006). From American city to Japanese village: A cross-cultural investigation of implicit race attitudes. *Child Development, 77*, 1268-1281.

Dymond, S., & Barnes, D. (1994). A transfer of self-discrimination response functions through equivalence relations. *Journal of the Experimental Analysis of Behavior, 62,* 251-267.

Dymond, S., & Barnes, D. (1995). A transformation of self-discrimination response functions in accordance with the arbitrarily applicable relations of sameness, more-than, and less-than. *Journal of the Experimental Analysis of Behavior, 64*, 163-184.

Dymond, S., & Barnes, D. (1996). A transformation of self-discrimination response functions in accordance with the arbitrarily applicable relations of sameness and opposition. *The Psychological Record, 46*, 271-300.

Dymond, S., & Barnes, D. (1998). The effects of prior equivalence testing and detailed verbal instructions on derived self-discrimination transfer: A follow-up study. *The Psychological Record, 48*, 147-170.

Eberhardt, J.L., Goff, P.A., Purdie, V.J., & Davies, P.G. (2004). Seeing black: Race, crime and visual processing. *Journal of Personality and Social Psychology, 87*, 876-893.

Edwards, C.A., Jagielo, J.A., Zentall, T.R., & Hogan, D.E. (1982). Acquired equivalence and distinctiveness in matching to sample by pigeons: Mediation by reinforcer-specific expectancies. *Journal of Experimental Psychology: Animal Behavior Processes, 8,* 244-259.

Egloff, B., & Schmukle, S.C. (2002). Predictive validity on an Implicit Association Test for measuring anxiety. *Journal of Personality and Social Psychology, 83,*1141-1455.

Eysenck, M.W., & Keane, M.T. (2000). *Cognitive psychology: A student's handbook (4th ed)*. East Sussex, UK: Psychology Press Ltd.

Farnham, S.D., Greenwald, A.G., & Banaji, M.R. (1999). Implicit self-esteem. In D. Abrams, and M.A. Hogg (Eds.), *Social identity and social cognition* (pp. 230-248). Oxford, U.K.: Blackwell Publishers.

Gilbert., D.T., & Hixon. J.T. (1991). The trouble with thinking: activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60,* 509–17.

Gray, N.S., Brown, A.S., MacCulloch, M.J., & Smith, J. (2005). An implicit test of the associations between children and sex in pedophiles. *Journal of Abnormal Psychology, 114*, 304-308.

Greenwald, A.G. (1990). What cognitive representations underlie social attitudes? *Bulletin of the Psychonomic Society, 28*, 254-260.

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., & Breckler, S. J. (1985). To whom is the self presented? In B.R. Schlenker (Ed.), *The self and social life* (pp. 126-145). New York: McGraw-Hill.

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

Greenwald, A.G., McGhee, D.E., & Schwarz, 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., & Klauer, K.C. (2005). Validity of the salience asymmetry interpretation of the IAT: Comment on Rothermund and Wentura (2004). *Journal of Experimental Psychology: General, 134*, 420-425.

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.

Greenwald, A.G., Oakes, M.A., & Hoffman, H.G. (2003). Targets of discrimination: Effects of race on responses to weapons holders. *Journal of Experimental Social Psychology, 39,* 399-405.

Grover, V.P., Keel, P.K., & Mitchell, J.P. (2003). Gender differences in implicit weight identity. *International Journal of Eating Disorders, 34*, 125-135.

Hastings, R.P. (1997). Measuring staff perceptions of challenging behaviour: The Challenging Behaviour Attributions Scale (CHABA). *Journal of Intellectual Disabilities Research, 41*, 495–501.

Haeffel, G.J., Abramson, L.Y., Brazy, P.C., Shah, J.Y., Teachman, B.A., & Noesk, B.A. (2007). Explicit and implicit cognition: A preliminary test of a dual-process theory of cognitive vulnerability to depression. *Behavior Research and Therapy, 45,* 1155-1167.

Hayes, S.C., & Barnes, D. (1997). Analysing derived stimulus relations requires more than the concept of stimulus class. *Journal of the Experimental Analysis of Behavior, 68*, 235-270.

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

Hayes, S.C., Bissett, R., Roget, N., Padilla, M., Kohlenburg, B.S., Fischer, G., Masuda, A., Pistorello, J., Rye, A.K., Berry, K., & Nicols, R., (2004). The impact of acceptance and commitment training and multi-cultural training on the stigmatising attitudes and professional burnout of substance abuse counsellors. *Behaviour Therapy, 35,* 821-835.

Hayes, S.C., Bissett, R.T., Strosahl, K.D., Wilson, K.D., Pistorello, J., Dykstra, T. A., et al. (2000). *Psychometric properties of the Acceptance and Action Questionnaire (AAQ)*. Non-published manuscript. (Obtained by request from first author).

Hayes, S.C., & Hayes, L.J. (1989). The verbal action of the listener as a basis for rule-governance. In S.C. Hayes (Ed.) *Rule-governance behavior: Cognition, contingencies, and instructional control*. New York: Plenum Press.

Hayes, S.C., Strosahl, K.D., & Wilson, K.D. (1999). *Acceptance and Commitment Therapy. An experimental approach to behavior change*. New York: Guildford.

Hayes, S.C., Strosahl, K.D.(Eds.) (2004). *A practical guide to Acceptance and Commitment Therapy.* New York: Springer-Verlag.

Hayes, S.C., Wilson, K.G., Gifford, E.V., Follette, V.M., & Strosahl, K.D. (1996). Experiential avoidance and behavioral disorders: A functional dimensional approach to diagnosis and treatment. *Journal of Consulting and Clinical Psychology, 64*, 1152-1168.

James, W. (1890). *Principles of psychology (Vol. 1)*. New York: Henry Hold.

Karpinski, A. (2004). Measuring self-esteem using the Implicit Association Test: The role of the other. *Society for Personality and Social Psychology, 30,* 22-34.

Kernis, M.H., Weisenhunt, C.R., Waschull, S.B., Greenier, K.D., Berry, A.J., Herlocker, C.E., & Anderson, C.A. (1998). Multiple facets of self-esteem and their relations to depressive symptoms. *Personality and Social Psychology Bulletin, 24,* 657-688.

Koole, S.L. Dijksterhuis, A., & van Knippenberg, A. (2001). What's in a name: implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology, 80,* 669-85.

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (1999). *International Affective Picture System (IAPS): Instruction manual and affective ratings.* Technical report A-4. Florida: University of Florida, The Center for Research in Psychophysiology.

Leader, G., Barnes-Holmes, D., & Smeets, P.M. (2000). Establishing equivalence relations using a respondent-type procedure III. *The Psychological Record, 50,* 63-78.

Leader, G., & Barnes-Holmes, D. (2001). Establishing fraction-decimal equivalence using a respondent-type training procedure. *The Psychological Record, 51*, 151-166.

Lindeman, M., & Verkasalo, M. (1995). Personality, situation, and positive-negative asymmetry in socially desirable responding. *European Journal of Personality, 9,* 125-34.

Marial, J.B. (2004). Spanish adaptation of the Acceptance and Action Questionnaire (AAQ). *International Journal of Psychology and Psychological Therapy, 4*, 505-515.

McConahay, J.B. (1983). Modern racism and modern discrimination: The effects of race, racial attitudes, and context on simulated hiring decisions. *Personality and Social Psychology Bulletin, 9*, 551-558.

McConaghy, J.B. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In J.F. Dovido, and S.L. Gaertner (Eds.), *Prejudice, discrimination and racism* (pp. 91-125). Orlando, FL: Academic Press.

McHugh, L., Barnes-Holmes, Y., O'Hora, D., & Barnes-Holmes, D. (2004). Perspective-taking: A relational frame analysis. *The Experimental Analysis of Human Behavior Bulletin, 22*, 4-10.

Neely, J.H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology, 106,* 226-254.

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

Nunes, K.L., Firestone, P., & Baldwin, M.W. (2007). Indirect assessment of cognitions of child sexual abusers with the Implicit Association Test. *Criminal Justice and Behavior, 34,* 454-475.

Nuttin, J.M. (1985). Narcissism beyond Gestalt and awareness: The name letter effect. *European Journal of Social Psychology, 15,* 353-61.

Ottaway, S.A., Hayden, D.C., & Oakes, M.A. (2001). Implicit attitudes and racism. Effect of word familiarity and frequency on the Implicit Association Test. *Social Cognition, 19,* 97-144.

Payne, B.K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology, 81*, 181-192.

Payne, B.K. (2006). Weapon bias: Split second decisions and unintended stereotyping. *Current Directions in Psychological Science, 15*, 287-291.

Payne, B.K., Burkley, M.A., & Stokes, M.B. (2008). Why do implicit and explicit attitudes test diverge? The role of structural fit. *Journal of Personality and Social Psychology, 94*, 16-31.

Payne, B.K., Lambert, A.J., & Jacoby, L.L. (2002). Best laid plans: Effects of goals on accessibility bias and cognitive control in race-based misperceptions of weapons. *Journal of Experimental Social Psychology, 38*, 384-396.

Pelham, B.W., & Swann, W. B. (1989). From self-conceptions to self-worth: On sources and structure of global self-esteem. Journal of Personality and Social Psychology, 57, 672-680.

Pine, D.S., Mogg, K., Bradley, B.P., Monegomery, L., Monk, C.S., McClure, E., Guyer, A.E., Ernst, M., Charney, D.S., & Kaufman, J. (2005). Attention bias to threat in maltreated children: Implications for vulnerability to stress-related psychopathology. *American Journal of Psychiatry, 162,* 291-296.

Plant, E.A., & Devine, P.G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology, 75*, 811-832.

Puliafico, A., & Kendall, P. (2006). Threat-related attentional bias in anxious youth: A review. *Clinical Child and Family Review, 3-4,* 162-180.

Reese, H.W. (1968). *The perception of stimulus relations: Discrimination learning and transposition*. New York: Academic Press.

Richeson, J.A., & Trawalter, S. (2008). The threat of appearing prejudiced and race-based attentional biases. *Association for Psychological Science, 19*, 98-102.

Rippere, V. (1977). "What's the thing to do when you're feeling depressed?" – a pilot study. *Behavior Research and Therapy, 15*, 185-191.

Roefs, A., & Jansen, A., (2002). Implicit and explicit attitudes toward high-fat foods in obesity. *Journal of Abnormal Psychology, 111*, 517-521.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rudman, L.A., & Ashmore, R.D. (2007). Discrimination and the Implicit Association Test. *Group Processes and Intergrroup Relations*, *10*, 359-372.

Sagar, H.A., & Schofield, J.W. (1980). Racial and behavioral cues in black and white children's perceptions of ambiguously aggressive acts. *Journal of Personality and Social Psychology, 19,* 590-598.

Scanlon, G. (2008). The utility of the Implicit Relational Assessment Procedure (IRAP): Measuring self-esteem profiles of Children with Special Educational Needs and Assessing the Attitudes of Teachers to Children with Emotional Behavioural Difficulties in Mainstream Education. *Unpublished doctoral thesis.*

Schlenker, B.R., & Britt, T.W. (1999). Beneficial impression management: Strategically controlling information to help friends. *Journal of Personality and Social Psychology*, *76,* 559-573.

Schlenker, B.R., & Britt, T.W. (2001). Strategically controlling information to help friends: Effects of empathy and friendship strength on beneficial impression management. *Journal of Experimental Social Psychology*, *37*, 357-372.

Sackeim, H.A., & Gur, R.C. (1978). Self-deception, self-confrontation, and consciousness. In G.E. Schwartz, and D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in Research* (Vol. 2, pp. 139-197). New York: Plenum.

Schmidtke, V., & Heuer, H. (1997). Task integration as a factor in secondary-task effects on sequence learning. *Psychological Research, 60,* 53-71.

Schuman, H., Steeth, C., & & Bobo, L. (1997). Racial attitudes in America: Trends and interpretations. Cambridge, MA: Harvard University Press.

Sidman, M. (1971). Reading audio-visual equivalences. *Journal of Speech and Hearing Research, 14*, 5-13.

Simon. J.R. (1990). The effects of an irrelevant directional cure on human information processing. In R.W. Proctor, and T.G. Reeve (Eds), *Stimulus-response compatibility: An integrated perspective* (pp. 31-86). New York: North Holland.

Smeets, P.M., Leader, G., & Barnes, D. (1997). Establishing stimulus classes with adults and children using a respondent training procedure: A follow-up study. *The Psychological Record, 47*, 285-308.

Smith-McAllen, A., Johnson, B.T., Dovidio, J.F., & Pearson, A.R. (2006). Black and white: The role of colour bias in implicit race bias. *Social Cognition*, *24,* 46-73.

Spalding, L.R., & Hardin, C.D. (1999). Unconscious unease and self-handicapping: Behavioral consequences of individual differences in implicit and explicit self-esteem. *Psychological Science, 10*, 535–539.

Stacy, A.W., Newcomb, M.D., & Ames, S.L. (2000). Implicit cognition and HIV risk behavior. *Journal of Behavioural Medicine, 23*, 475-499.

Stadler, M.A. (1995). Role of attention in implicit serial learning. *Journal of Experimental. Psychology: Learning, Memory and Cognition, 21,* 674-685.

Steele, D.L., & Hayes, S.C. (1991). Stimulus equivalence and arbitrary applicable relational responding. *Journal of the Experimental Analysis of Behavior, 56,* 555-561.

Stewart, I., Barnes-Holmes, D., Roche, B., & Smeets, P.M. (2001). Generating derived relational networks via the abstraction of common physical properties: A possible model of analogical reasoning. *The Psychological Record, 51*, 381-408.

Teachman, B.A., Gregg, A., & Woody, S.R. (2001). Implicit processing of fear-relevant stimuli among individuals with snake and spider fears. *Journal of Abnormal Psychology, 110*, 226-235.

Teachman, B.A., & Woody, S.R. (2003). Automatic processing in spider phobia: Implicit fear associations over the course of treatment. *Journal of Abnormal Psychology, 112,* 100-109.

Tavassoli, N.T., & Yih, H.L. (2004). The effect of attribute order on judgment in Chinese and English. *Journal of Experimental Psychology: Applied*, *10*, 258-266.

Tulving, E., Schacter, D., & Stark, H.A. (1982). Priming events in word fragment completion are independent on recognition memory. *Learning Memory and Cognition, 8,* 336-341.

Varitanian, L.R., Herman, C.P., & Polivy, J. (2004). Accuracy in the estimation of body weight: An alternate test of the motivated-distortion hypothesis. *The International Journal of Eating Disorders, 36,* 69-75.

Vitale, A., Barnes-Holmes, Y., Barnes-Holmes, D., & Campbell, C. (2008). Facilitating responding in accordance with the relational frame of comparison: Systematic empirical analyses. *The Psychological Record, 58*, 365-390.

Warrington, E.K., & Weinskrantz. L. (1968). New method of testing long-term retention with special inference to amnesic patients. *Nature, 217,* 972-974.

Wegner, D.M. (1994).*White bears and other unwanted thoughts.* New York: Guilford.

Wenzlaff, R.M., Wengar, D.M., & Klein, S.B. (1991). The role of thought suppression in the bonding of thought and mood. *Journal of Personality and Social Psychology, 60,* 500-508.

Wiers, R.W., de Jong, P.J., Haverman, R., & Jelicic, M. (2004). How to change implicit drug use-related cognitions in prevention: A transdisciplinary integration of findings from experimental psychopathology, social cognition, memory, and experimental learning psychology. *Substance Use and Misuse, 39,* 1625-1684

Wiers, R., van Woerden, N., & Smulders, F. (2002). Implicit and explicit alcohol-related cognitions in heavy and light drinkers. *Journal of Abnormal Psychology, 111,* 648-658.

Wilson, K.G., Hayes, S.C., Gregg, J., & Zettle, R.D. (2001). Psychopathology and psychotherapy. In S.C. Hayes, D. Barnes-Holmes, and B. Roche. (Eds). *Relational frame theory: A post-Skinnerian account of human language and cognition.* New York: Plenum.

Wittenbrink, B., Judd, C.M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology, 72,* 262-274.

Wulfert, E., & Hayes, S.C. (1988). The transfer of conditional sequencing through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior, 50*, 125-144.

Yamaguchi, S., Greenwald, A.G., Banaji, M.R., Murakami, F., Chen, D., Shiomura, K., Kobayashi, C., Cai, H., & Krendl, A. (2007). Apparent universality of positive implicit self-esteem. *Psychological Science, 18,* 498-500.

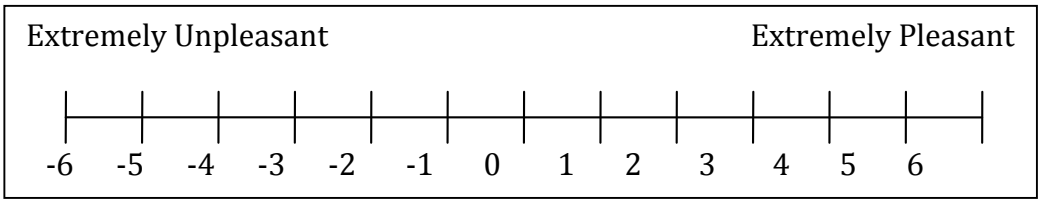
Yarmey, A.D., & Jones, H.T.P. (1983). Is eyewitness identification a matter of common sense? In S. Lloyd-Bostock, and B.R. Clifford (Eds), *Evaluating witness research: Recent psychological research and new perspectives*, (pp.13-40). New York: Wiley.

Table of Appendices

Appendix A

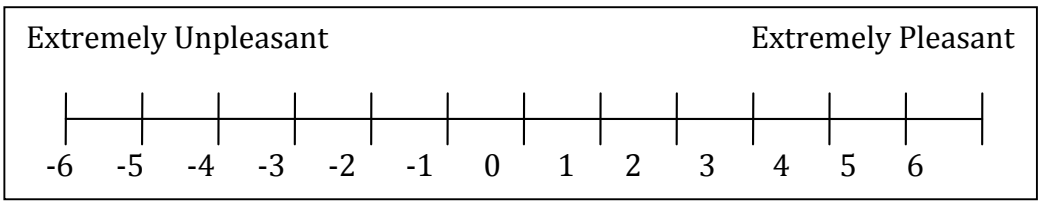
Examples of the Likert scales representing the stimuli employed in Experiment 1

Caress



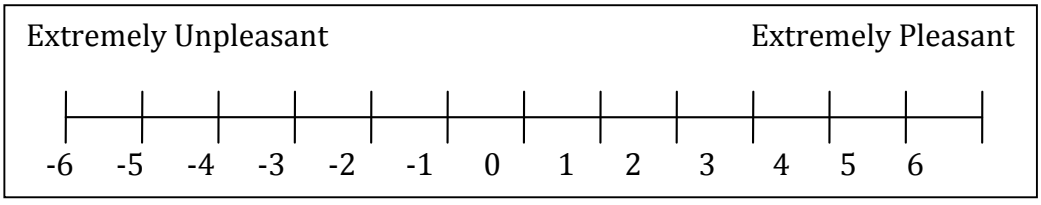
Please indicate the extent to which you would rate the word above as pleasant or unpleasant by circling the appropriate number.

Freedom



Please indicate the extent to which you would rate the word above as pleasant or unpleasant by circling the appropriate number.

Abuse



Please indicate the extent to which you would rate the word above as pleasant or unpleasant by circling the appropriate number.

Appendix B

The Modified Modern Racism Scale (MMRS)

Please indicate the extent to which you agree or disagree with each of these statements by circling a number on each scale, where -2 indicates that you strongly disagree, -1 indicates that you slightly disagree, 0 that you neither disagree nor agree, 1 indicates that you slightly agree and 2 indicates that you strongly agree.

It is easy to understand the anger of Black people

Strongly Disagree -2 -1 0 1 2 Strongly Agree

_____

Blacks are getting too demanding in their push for equal rights

Strongly Disagree -2 -1 0 1 2 Strongly Agree

_____

Over the past few years Blacks have gotten more economically than they deserve

Strongly Disagree -2 -1 0 1 2 Strongly Agree

_____

Appendix C

The Discrimination and Diversity Scale (DDS)

Discrimination Scale

Using the scale below as a guide please indicate how much you agree or disagree with each statement.

Members of ethnic minorities have a tendency to blame Whites too much for problems that are of their own doing

1----------2----------3----------4----------5

Strongly Disagree Neutral Strongly Agree

Members of ethnic minorities often exaggerate the extent to which they suffer from racial inequality

1----------2----------3----------4----------5

Strongly Disagree Neutral Strongly Agree

Black people often blame the system instead of looking at how they could improve their situation themselves

1----------2----------3----------4----------5

Strongly Disagree Neutral Strongly Agree

Appendix C continued

The Discrimination and Diversity Scale (DDS)

Diversity Scale

Using the scale below as a guide please indicate how much you agree or disagree with each statement.

There is a real danger that too much emphasis on cultural diversity will tear Ireland apart

1----------2----------3----------4----------5

Strongly Disagree Neutral Strongly Agree

_____

The desire of many ethnic minorities to maintain their cultural traditions impedes the achievement of racial equality

1----------2----------3----------4----------5

Strongly Disagree Neutral Strongly Agree

_____

Whites need to learn about Black culture if positive inter-ethnic relations are to be achieved

1----------2----------3----------4----------5

Strongly Disagree Neutral Strongly Agree

_____

Appendix D

The Racial Experience Questionnaire

Subject No:_____

Subject Age:_____

Male [ ] Female [ ]

1. Do you know any Black people in Ireland?

   None [ ] A handful [ ] Many [ ]

2. Do you have any friends who are Black?

   None [ ] A handful [ ] Many [ ]

3. Do you work with any Black people?

   None [ ] A handful [ ] Many [ ]

4. Would you say that you have had one or more close friendships with a Black person?

   Yes [ ] No [ ]

5. Do many Black people live in your neighbourhood?

   None [ ] A handful [ ] Many [ ]

6. Would you say that your experiences of meeting Black people have been positive or negative overall?

   Very Negative [ ]

   Negative [ ]

   Equally Negative and Positive [ ]

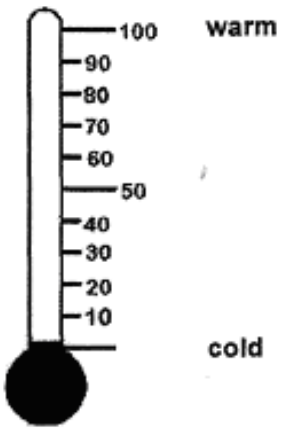
   Positive [ ]

   Very Positive [ ]

Appendix E

The Feeling Thermometers employed in Experiment 2-6

Using the scale shown on the thermometer, please rate how warm/favourable (100°) or cold/unfavourable (0°) you feel about Irish people and then immigrants living in Ireland.



Irish People: _____
(e.g. 0 = cold, 100 = warm)

Black People: _____
(e.g. 0 = cold, 100 = warm)

Appendix F

Examples of the Likert scales representing stimuli employed in Experiment 2

Please use the scale below to indicate which word in each pair is the most appropriate description of a *Black person*

-3indicates that the word on the left is the most appropriate description of a Black person, 3 indicates that the word on the right is the most appropriate description of a Black person and 0 indicates that neither word is appropriate

SAFE DANGERSOUS

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Please indicate the extent to which you would rate the word above as pleasant or unpleasant by circling the appropriate number.

_____

FRIENDLY HOSTILE

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Please indicate the extent to which you would rate the word above as pleasant or unpleasant by circling the appropriate number.

_____

Appendix G

An example of a Likert scale representing stimuli employed in

Experiments 2-6

1) For each picture you are shown, please state what object the person is holding

2) Also please indicate the extent to which you would rate the person in the picture as safe or dangerous by circling a number on the scale below.

Picture 1

The object is a :_____

I would rate the person in Picture 1 as:

Extremely Safe Extremely Dangerous

-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6

Please indicate the extent to which you would rate the person in each picture as safe or dangerous by circling the appropriate number.

Appendix H

The Rosenberg Self-Esteem Scale (RSES)

Instructions: Below is a list of statements dealing with your general feelings about yourself. If you strongly agree, circle **SA**. If you agree with the statement, circle **A**. If you disagree, circle **D**. If you strongly disagree, circle **SD**.

1	On the whole, I am satisfied with myself.	SA	A	D	SD
2*	At times, I think I am no good at all.	SA	A	D	SD
3	I feel that I have a number of good qualities.	SA	A	D	SD
4	I am able to do things as well as most other people.	SA	A	D	SD
5*	I feel I do not have much to be proud of.	SA	A	D	SD
6*	I certainly feel useless at times	SA	A	D	SD
7	I feel that I'm a person of worth, at least on an equal plane with others.	SA	A	D	SD
8*	I wish I could have more respect for myself.	SA	A	D	SD
9*	All in all, I am inclined to feel that I am a failure	SA	A	D	SD
10	I take a positive attitude toward myself.	SA	A	D	SD

*Indicates that item is negatively worded

Appendix I

The Acceptance and Avoidance Questionnaire-II (AAQ-II)

Below you will find a list of statements. Please rate how true each statement is for you by circling a number next to it. Use the scale below to make your choice.

1	2	3	4	5	6	7
never true	**very seldom true**	**seldom true**	**sometimes true**	**frequently true**	**almost always true**	**always true**

1.	Its OK if I remember something unpleasant.	1	2	3	4	5	6	7
2.	My painful experiences and memories make it difficult for me to live a life that I would value.	1	2	3	4	5	6	7
3.	I'm afraid of my feelings.	1	2	3	4	5	6	7
4.	I worry about not being able to control my worries and feelings.	1	2	3	4	5	6	7
5.	My painful memories prevent me from having a fulfilling life.	1	2	3	4	5	6	7
6.	I am in control of my life.	1	2	3	4	5	6	7
7.	Emotions cause problems in my life.	1	2	3	4	5	6	7
8.	It seems like most people are handling their lives better than I am.	1	2	3	4	5	6	7
9.	Worries get in the way of my success.	1	2	3	4	5	6	7
10.	My thoughts and feelings do not get in the way of how I want to live my life.	1	2	3	4	5	6	7