Developing the Implicit Relational Assessment Procedure (IRAP) as a Measure of Implicit Racial Bias

NUI MAYNOOTH

Óllscoil na h Éireann Má Nuad

Patricia M Power B.A. (Hons)

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

July, 2010

Head of Department

Dr Fiona Lyddy

Supervisor

Professor Dermot Barnes-Holmes

Department of Psychology

Table of Contents

ACKNOWLEDGEMENTS

ACKNOWLEDGMENTS

There are several people to whom I am truly indebted for their support and encouragement over the last number of years:

My Supervisor,

Professor Dermot Barnes-Holmes

Thanks D for being an outstanding mentor. Your enthusiasm and commitment are contagious- resistance is futile!

My Family, Friends and the Merrigans

for your endless patience, encouragement, love and support,

especially my husband,

Jim

Psychology department staff, undergrads, postgrads, past and present,

&

My fellow IRAPers!

For the advice and laughs… and for providing welcome distractions.

Finally, thank you to everyone who took the time to participate in my research and to those whose research has served to inspire and motivate the current work.

ABSTRACT

ABSTRACT

The aim of the current programme of research was to determine if the IRAP, a recently developed methodology for the assessment of implicit cognition, is a useful measure of implicit racial bias in the Irish context. Over the course of a series of experiments, the research refined the IRAP and examined its relationship with various alternative attitudinal indices, including self-report measures and measures of behavioural intentions. In addition, the fifth experiment explored the predictive validity of the IRAP using known-groups, while Experiment 6 investigated the relationship between neural activity, measured with electroencephalograms, and IRAP responses. In the final experiment, the malleability of IRAP performances, as a result of acceptance and education-based interventions, was investigated. Overall, the research reported in the current thesis provides support for the reliability and validity of the IRAP and suggests that it is a relatively robust measure that could be used in subsequent research in the study of implicit racial bias.

CHAPTER 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

This chapter serves as a road map for the current thesis. It begins by outlining the pertinent literature and highlighting the relevance of the current research programme. Following this, each of the chapters of the thesis is mapped out. The principal goal of the thesis is to determine if the Implicit Relational Assessment Procedure (IRAP: Barnes-Holmes, Barnes-Holmes, Hayden, Milne, Power, & Stewart, 2006), a recently developed methodology for the assessment of implicit cognition, is a useful measure of implicit racial bias in the Irish context.

The Socio-cognitive Approach to Understanding Attitudes

From a socio-cognitive psychological perspective attitudes are hypothetical constructs which cannot be directly observed but rather, are inferred from observable responses (Eagly & Chaiken, 1993; Rosenberg & Hovland, 1960). Attitudes as hypothetical constructs, by their very nature, are difficult to define. However, Fazio and Petty (2008) suggest that an "attitude is a person's evaluation of an object -- favorability or unfavorability toward the object" (p.3). It is this evaluative component that many assume to be the most important aspect of an attitude (see Baron, Byrne, & Watson, 2004). The socio-cognitive approach contends that such evaluations are stored in memory (Fazio, Sanbonmatsu, Powell, & Kardes, 1986), exist across time and are automatically activated from memory upon mere observation of the attitude object (Fazio & Petty, 2008; Fazio, 2001). In fact, cognitive neuroscientific evidence indicates that evaluative judgments produce different patterns of neuronal activity when compared with non-evaluative

judgments (Ajzen, 2001).

The study of attitudes has long preoccupied social scientists and continues to do so, not least of all because of their assumed relationship with behaviour. From this stance, positive attitudes are deemed to be associated with approach behaviours and negative attitudes with avoidance behaviours (Eagly & Chaiken, 1993). This belief accounts for at least some of the preoccupation with attitude research. The following section will review socio-cognitive research investigations of the structure of attitudes, their relationship with behaviour and how this perspective has addressed the measurement of attitudes.

The structure of attitudes. Within mainstream attitude research there remains a theoretical debate concerning the structure and dimensionality of attitudes (Chaiken & Stangor, 1987). The one-component view suggests that attitudes are primarily affective evaluations (Thurstone, 1928). More recently, Fishbein and Ajzen (1975) suggest a two-component perspective in which cognitions also have a large role to play in evaluation. The three-component or tripartite model is the most popular and widely accepted model of attitudes (Krech & Crutchfield, 1948; Katz & Stotland, 1959; Rosenberg & Hovland, 1960; Triandis, 1971). This model sees attitudes as being comprised of cognitive, affective and behavioural components. However, although these three components are conceptually and empirically distinct (Breckler, 1984), research has demonstrated that they are not easily distinguishable from each other in simple assessments (see Eagly & Chaiken, 2007). Despite these inconsistent findings, the model remains the most popular mainstream approach to the structure of attitudes and has served to provide an important conceptual framework (Fazio & Petty, 2008).

The attitude-behaviour relationship. The relationship between attitudes and behaviour has been widely researched. As discussed previously, the three-component view tends to assume that attitudes have some predictive utility and it is at least in some part because of this belief that the study of attitudes is considered theoretically important. However, early research indicated that the relationship between attitudes and behaviour is tenuous (LaPiere 1934; Wicker 1969). Later, research began to focus on the context in which attitudes and behaviour were related and on variables that moderated this relationship. Findings indicated that the magnitude of the attitude-behaviour relation varies as a consequence of how the relation is measured (i.e. the assessment tool), where or in what context the relation is measured and on various salient aspects of the individual being assessed (Fazio, 2001; Krosnick, 1988). The moderating power of some of these variables has been widely reported in the research literature.

Studies have shown, for example, that participants who are highly motivated to control prejudice can report attitudes that they consider more socially desirable, and consequently these reported attitudes might not correlate with actual behaviours towards the attitudinal object (Ajzen & Fishbein, 2005; Campbell, 1950). In addition, low correlations may be the product of assessing only one component of the tripartite model, which would not fully represent the complexity of the attitude construct thus resulting in a weaker attitude behaviour relationship (Ajzen & Fishbein, 2005). Finally, according to Greenwald (1989) low correlations may also result from a failure to measure attitudes and behaviours at comparable levels of specificity.

More recent attitude research has been concerned with revealing the direction of

attitude-behaviour relations. Social psychology research findings indicate that this relationship is bi-directional with attitudes influencing behaviour at times and vice versa (e.g., Holland, Verplanken, & Van Kippenberg, 2002). Within the cognitive literature numerous processes have been proposed as explanations for how behaviour can influence attitudes. Bem (1972), for example, suggested that when we are unsure of our evaluations of a particular object, we may infer our attitudes from our behaviours towards that object. Such inferences may be made when the attitude is weak, ambiguous or un-interpretable. Other cognitive theorists suggest that behaviour can influence attitudes when we attempt to reduce the experience of cognitive dissonance. That is, after performing a behaviour that is inconsistent with an attitude we feel discomfort, but we cannot change the behaviour we have performed and thus we tend to change our attitude so it becomes consistent with our behaviour (Festinger, 1957).

Several socio-cognitive theories including the theory of reasoned action (TRA; Fishbein & Ajzen, 1975), and its successor the theory of planned behaviour (TPB; Ajzen, 1991), have attempted to model the relationship between attitudes and behaviours. These models each emphasise the moderating role of behavioural intention in determining whether or not a particular behaviour is carried out in response to a specific attitude. Socio-cognitive research by Ajzen (1991) and Beale and Manstead (1991) indicate that the attitude-behaviour relationship is stronger than was previously indicated in the early literature. Typically, the stronger a person's intention to engage in a particular behaviour the more likely they are to perform the behaviour.

The measurement of implicit attitudes. Explicit self-report measures have long

been the principal means of assessing attitudes from Likert's (1932) summated rating technique to Osgood, Suci and Tannenbaum's (1957) semantic differential scale. However, more recent research has highlighted a number of confounds intrinsic to such self-report measures (e.g., de Jong, 2002; Gemar, Segal, Sagratti, & Kennedy, 2001; Raja & Stokes, 1998; Teachman, Gregg, & Woody, 2001). According to Dawes (1972) the limitations of self-report methods result from the fact that when completing such measures participants are given time and opportunity to control their responses and these are thus subject to various cognitive, motivational and situational factors. The limitations of self report measures can be conceptualised as falling into two main categories; (i) those caused by the frailties of introspection and (ii), those that result from various demand characteristics (Orne, 1962). For example, individuals may be unaware that they hold a particular attitude and, accordingly, may fail to report it (Nisbett & Wilson, 1977; Dambrun & Guimond, 2004). Conversely, individuals may hold attitudes that they believe to be socially undesirable and may therefore attempt to conceal these from researchers (Paulhus, 1984; Rust & Golombok, 1999). These limitations are exacerbated by the fact that the way questions are framed may influence an individual's response (Rasinski, 1989).

Recently, in an effort to circumvent these problems, researchers have devoted increasing attention to studying the nature of implicit attitudes. Greenwald and Banaji (1995) define implicit attitudes as "introspectively unidentified or inaccurately identified traces of past experience that mediate favorable or unfavorable feeling, thought, or action toward social objects" (p. 8). Despite ongoing debate surrounding the adequacy of this definition (see De Houwer, 2006), the core argument holds that because implicit attitudes

are often unconscious, their influence on subsequent behaviours may go unnoticed. Insofar as implicit attitudes may be unconscious, traditional self-report explicit measures will quite probably fail to capture them. Consequently, researchers have endeavoured to develop reaction-time based methodologies in which implicit attitudes are inferred on the basis of response speed and accuracy (see De Houwer, 2006).

In the socio-cognitive literature implicit attitude measures are referred to as indirect attitude measures. The term 'indirect' comes from the fact that these measures do not directly ask the individual to report their attitude (Petty, Fazio, & Brinol, 2008). To date, the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) is the most widely used indirect, reaction-time based measure in the assessment of implicit attitudes in socio-cognitive and clinical research. The IAT is based upon the assumption that it should be easier and therefore will take less time to respond, when two closely associated concepts in memory are assigned to the same response key, than when these two concepts are assigned to different keys.

Evidence in support of the psychometric properties of the IAT have been comprehensively investigated for more than a decade and are generally accepted (e.g., Fazio & Olsen, 2003; Nosek, Banaji, & Greenwald, 2002). However, across a recent meta-analysis, Greenwald, Poehlmann, Uhlmann, and Banaji, (2009), found that the relative predictive validities of the IAT and self-report measures varied depending on the content domains, but overall, that each measure provided a gain in predictive validity compared with using the other alone. For the IAT, this gain was found to be greater when socially sensitive topics were being investigated. Greenwald et al. (2009) thus recommend that

both the IAT and self-report measures be employed together, as predictors of behaviour.

The IAT has been applied to fields and content domains both within and beyond social psychology, from clinical, developmental and health psychology research (e.g., Teachman & Brownell, 2001; Baron & Banaji, 2006; and Teachman, Gapinski, Brownell, Rawlins, & Jeyaram, 2003) to neuroscientific research (e.g., Cunningham, Johnson, Raye, Gatenby, Gore, & Banaji, 2004), and market research (e.g., Maison, Greenwald, & Bruin, 2004). In particular, it has been used extensively to examine prejudice (e.g., Dasgupta & Greenwald, 2001; Rudman, Feinberg & Fairchild, 2002) and stereotypes (e.g., Rudman, Greenwald, & McGhee, 2001).

Numerous studies using the IAT, have indicated that white participants tend to show a relatively strong pro-white/anti-black bias (see Dasgupta, McGhee, Greenwald & Banaji, 2000; Greenwald, Banaji, Rudman, Farnham, Nosek, & Mellott, 2002; Monteith, Voils, & Ashburn-Nardo, 2001; Livingston, 2002; Ottoway, Hayden, & Oakes, 2001). In the study conducted by Dasgupta et al., for example, findings showed that participants responded faster on tasks that categorized pleasant words with 'white' (faces or names) and unpleasant words with 'black' (faces and names) than vice versa. Furthermore, this pro-white/ anti-black bias occurred for participants who explicitly stated that they held no racist attitudes,

The IAT often reveals levels of bias not registered at self-report (e.g., Chambliss, Finley, & Blair, 2004; O'Brien, Hunter, Halberstadt, & Anderson, 2007). As discussed above, implicit assessment methodologies, such as the IAT, were developed in part to overcome the problems associated with self-report measures. However, despite its utility,

several limitations of the IAT have been identified. (see Lowery, Hardin, & Sinclair, 2001; Richeson & Ambady, 2003). The following limitations are of particular interest. Firstly, the IAT cannot be used to measure the valence of individual concepts as it was designed to be a measure of relative associative strength (De Houwer, 2002; Nosek, Greenwald and Banaji, 2005). That is, the IAT is relativistic in that it can indicate that x is preferred to y, but it cannot reveal to what extent x and y are liked or disliked, per se. In the context of a race IAT, for example, each trial involves presenting both of the relevant categories, (i.e., Black People and White People), and consequently the IAT effect is based on responses that occur in the context of both categories, rather than each independently. Thus, a pro-white/anti-black IAT effect could indicate, for example, that a participant has a positive attitude to "White People" and a neutral attitude to "Black People", or it could indicate a neutral attitude to "White People" and a negative attitude to "Black People".

Secondly, the IAT offers a relatively indirect measure of implicit attitudes. De Houwer (2002), for example, has argued that the IAT measures associations rather than relations among stimuli or events, and as such can provide only an indirect measure of beliefs:

> 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* [italics added] between the concept "self" and the concept "bad," namely *a directional association*

> [italics added] which specifies that "bad" is a property or characteristic of "self." IAT effects do not reflect the *nature or directionality of an association between concepts* [italics added], 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 (pp. 117–118).

Thirdly, although there is considerable evidence that the IAT may reveal levels of bias not recorded using self-report measures, there is recent research that suggests that the IAT is not completely impervious to at least some of the confounds that affect explicit measures. For example, evidence that the IAT is sensitive to the context in which it is completed was provided by a study investigating homonegativity (Boysen, Vogel, & Madon, 2006) in which the measure was administered in both public and private assessment situations. In the public condition, participants believed that the experimenter would know their level of bias. In contrast, participants in the private condition did not hold this belief. The results showed that the public context significantly decreased the

level of bias toward homosexuality relative to the private context. Critically, this finding suggests that, similar to explicit measures, racial bias on the IAT may be reduced when participants are more motivated to conceal their prejudice (i.e., when they think their prejudice will be public; see Gawronski, LeBel, & Peters, 2007, for a review that questions the common assumption that implicit measures are immune or less sensitive to social desirability concerns).

Although the IAT is by far the most widely used measure of implicit attitudes, substantial empirical attention has been directed towards developing alternative implicit measures that aim to address some of its limitations. So-called implicit measures such as the Go/No-go Association Task (GNAT; Nosek & Banaji, 2001), Evaluative Priming (Fazio, Sanbonmatsu, Powell, & Kardes, 1986), the Emotional Stroop (Pratto & John, 1991) and the Extrinsic Affective Simon Task (EAST; De Houwer, 2003b) have been offered, but they all share a common feature. Specifically, they were all designed to measure associations in memory. For example, evaluative priming involves presenting a positively or negatively valenced prime before presenting the relevant attitude object. If the prime and target share the same valence (e.g., they are both positive) then the prime and target will be associated in memory, and thus the former will activate the latter leading to a relatively short response time. If the prime and target do not share the same valence, associative strength will be low and response time will be longer. In effect, the vast majority of implicit measures are based on the assumption that they are revealing underling associative strengths represented in memory. However, as we shall see in the next section, an alternative approach has recently been offered.

An Alternative Behavioural Approach to Understanding Attitudes

The previous section detailed the mainstream socio-cognitive approach to the measurement of attitudes. The widely used IAT was discussed and several limitations of the IAT where highlighted. The current section will introduce a recent alternative implicit attitude measure -- the Implicit Relational Assessment Procedure (IRAP), which aims to address some of the IAT's limitations. The IRAP emerged from behaviour analysis, a distinctly different psychological tradition to the socio-cognitive tradition which gave rise to the IAT. The following section will provide an account of the behaviour analytic approach to attitudes. The final part of the chapter will provide a detailed account of the IRAP and of research that has employed the IRAP for the measurement of implicit attitudes across a range of domains.

Behaviour analysis. Behaviourism is a philosophy of science (Baum, 1994). The scientific study of the behaviour of organisms is known as behaviour analysis (Leslie & O'Reilly, 1999). Behaviour analysis, thus involves the study of behaviour and the variables that influence it (Grant & Evans, 1994) and behaviour analysts seek to predict and influence behaviour (defined as any and all activities that an organism can engage in, both overtly and covertly). Behaviour analysis aims to identify functional relationships between manipulable independent variables found in the environment and behaviour (dependent variable). When relevant manipulable variables are identified, the experimenter may then have the opportunity to influence and change the functionally related behaviour as they wish. The classification of functional relations produces a scientific account of behaviour, independent of unobservable mental events or

hypothetical constructs (Baum, 1994). It enables the application of experimental analyses to both overt and covert behaviours and shuns so-called "explanatory fictions" of the mind and mental states (Nye, 1975). Within the behaviour-analytic tradition, mentalisitc concepts like attitudes are not offered as explanations for behaviour. However, the functional relationships involved in behaviours (which are typically accepted as indicators of attitudes) do necessitate systematic empirical analysis. Derived stimulus relations are particularly relevant in this regard.

Relational Frame Theory. Relational Frame Theory (RFT; Hayes, Barnes-Holmes & Roche, 2001) is a modern behavior analytic account of human language and cognition. RFT has its roots in investigations of derived relational responding (Hayes et al., 2001). Sidman (1971) was the first to demonstrate the emergence novel behavior that had not been directly trained or reinforced. When he trained participants in a series of related conditional matching tasks using arbitrary stimuli he found that several untaught performances emerged according to a pattern which he called "stimulus equivalence". For example, if a participant was taught to choose A in the presence of B, and to choose B in the presence of C, then several untrained performances would typically emerge including choosing B with A and C with B, thus reversing the taught relations (referred to as symmetry) and choosing A with C (transitivity) and C with A (combined symmetry and transitivity). Sidman named the overall pattern stimulus equivalence because the participant appeared to be responding to the stimuli as equivalent.

The concept of stimulus equivalence received considerable attention because it offered a way to rapidly advance response repertoires. Furthermore, empirical research

revealed a strong link between stimulus equivalence and human language across a variety of contexts (Barnes-Holmes, Barnes-Holmes, Smeets, Cullinan, & Leader, 2004; Cowley, Green, & Braunling-Mc Morrow, 1992; Devany, Hayes, & Nelson, 1986; Kendall, 1983; Wulfert & Hayes, 1989). Barnes (1994) outlined five research areas that appear to lend empirical support for this view: (i) derived equivalence is readily demonstrated by verbally-able humans, whereas, it has not been unequivocally demonstrated by non-verbally-able humans or nonhumans (Barnes, McCullagh, Keenan, 1990; Devany et al., 1986; Dugdale & Lowe, 2000; Hayes, 1989; Sidman & Tailby, 1982); (ii) learning to label stimuli may make equivalence responding possible in young children (Dugdale & Lowe, 2000); (iii) human language impairments can be treated through the use of equivalence procedures (e.g., Cowley et al., 1992); (iv) a behaviour analytic understanding of both symbolic meaning and the generative nature of grammar has been generated through stimulus equivalence (Barnes & Holmes, 1991; Barnes-Holmes, Barnes-Holmes, & Cullinan, 2000; Hayes & Hayes, 1989; Wulfert & Hayes, 1988); (v) equivalence phenomena have been applied to human verbal behaviors such as logical reasoning and social categorization (e.g., Barnes & Hampson, 1993; Roche & Barnes, 1996; Watt, Keenan, Barnes, & Cairns, 1991) . In addition, neuropsychological research has recently uncovered similar brain activation patterns during the semantic processing underlying language and the formation of equivalence relations (e.g., Dickins, Singh, Roberts, Burns, Downes, Jimmieson & & Bentall, 2001). In general, these findings indicate that the control exerted over behaviour by stimuli participating in equivalence classes appears to be comparable with the control that verbal stimuli exert over human behaviour (Hayes &

Hayes, 1989).

From an RFT perspective, derived stimulus relations make up the core of what has been absent from a satisfactory behavioural account of human language. RFT appeals to the concept of arbitrarily applicable relational responding in its account for derived equivalence relations (Barnes-Holmes et al., 2004). This idea grew out of the basic finding that organisms, from insects to primates, can learn to respond to the *non-arbitrary* relations among stimuli (e.g., bigger than, smaller than; see Reese, 1968). This is now referred to as non-arbitrary relational responding. Furthermore, RFT holds that in the context of an appropriate history of multiple exemplar training, verbally-able humans can also respond to *arbitrary* relations among and between stimuli.

According to RFT, arbitrary relations are defined not by the formal properties of the stimuli involved but rather by additional features of the *context* outside of the stimuli being related. For example, imagine that a mother shows her verbally-able child a picture of a dog (stimulus A) and she says "This *is* a dog" (stimulus B). She might also tell the child that a dog (stimulus B) makes the sound "woof" (stimulus C). RFT proposes that contextual cues such as the spoken word "*is*" can bring a repertoire of arbitrarily applicable relational responding to bear on the stimuli such that the child will subsequently regard these stimuli as "going together" and will be able to derive novel relations between the stimuli that were not explicitly trained. For example, if the mother later gives the child pictures of different animals and asks "Which one says 'woof'?" then the child may well readily point to dog, despite this being an untrained response.

RFT proposes that this type of performance is based on a history of being rewarded

for responding relationally to pictures and words and to other pairs of objects in the presence of contextual cues (such as "*is*") that serve to manage or manipulate the response. In addition, following extensive training across multiple exemplars, relating becomes so abstracted that it can be arbitrarily applied to any stimuli. Arbitrarily applicable relational responding is also known as relational framing. This comes from the metaphor of an empty frame that could potentially be filled with any content.

For RFT, stimulus equivalence symbolizes an example of relational framing that is brought to bear by a certain feature of the context in which the task occurs. For example, in a matching to sample task in which a participant is trained to pick a stimulus consistently in the presence of another stimulus the context itself can function as a contextual cue signaling that the two stimuli are the same. As a result, further relational responses will be derived. This particular type of relational framing is referred to as framing in accordance with the relation of sameness or coordination. There are many diverse forms of relational framing including opposition, distinction, comparison, hierarchy, perspective, and so on, and the properties of the derived relational responses involved differ greatly (Barnes, 1994). A frame of opposition for example has the property that an opposite of an opposite is the same, an opposite of an opposite of an opposite is an opposite, and so forth (Hayes et al., 2001). In general therefore, RFT is far broader in scope than stimulus equivalence.

There are three central properties inherent in all forms of relational framing -- mutual entailment, combinatorial entailment and transformation of stimulus function. Mutual entailment refers to the bi-directionality of relational responding (Hayes et al.,

2001). That is, if X is related to Y in a given context, then a relationship between X and Y is entailed. The relationship between the stimuli can be symmetrical (i.e., as in the case of equivalence or coordination), but not necessarily so. For example, if X were smaller than Y, the relationship is not symmetrical but is mutually entailed. Therefore, two relations would exist, "X is smaller than Y" and "Y is bigger than X" (Hayes et al., 2001).

Combinatorial entailment refers to derived stimulus relations that involve two or more sets of relations. Combinatorial entailment makes it possible to define the relevant forms of relational frames (Hayes et al., 2001). For example, if X is related to Y in a particular context and Y is related to Z, then a relation is entailed between X and Z and equally between Z and X. This may include, but is not restricted to, the transitive relations found in stimulus equivalence. In mutually entailed relations, the specified relationship between X and Y always entails a relationship between Y and X at the same level of precision. Conversely, in combinatorial entailed relations, the derived relationship may be less precise than the original relationship. For example, if X is different to Y and Y is different to Z, then the relationship between X and Z and Z and X is clearly unknown. Moreover, the unknown nature of the latter relationships, in and of themselves, constitutes stimulus relations (i.e., identifying a relation as unspecified is a relational response).

The third defining property of a relational frame is termed the transformation of function. That is, any function associated with one of the stimuli involved in a relational frame may lead to the transformation of functions for any or all of the other stimuli participating in that frame (Barnes, 1994; Hayes et al., 2001; Hayes & Wilson, 1993). The functions are always transformed in terms of the specific relational frame involved. For

example, if two stimuli are involved in a frame of comparison, such that X is "more than" Y, and Y is known to have an aversive function, then X will acquire a stronger aversive function than Y.

RFT proposes that the three central properties of relational framing make up the basis of what an adequate behavioural account of stimulus equivalence and human language has been lacking (Hayes & Wilson, 1993). In particular, the belief that these processes are central to understanding language has provided a means of studying language and other complex behaviours in purely functional terms. From an RFT perspective, verbal behaviour constitutes the action of framing events relationally (Hayes et al., 2001, p.43). Moreover, this process involves two parties-- the speaker and the listener (Hayes & Hayes, 1989). When the speaker engages in this process they are speaking with meaning, and when a listener does so, they are listening with understanding (Hayes & Wilson, 1993). It is important to note that it is the framing of these events that indicates that the behaviour is verbal for both speaker and listener. Accordingly, verbal meaning is a highly specified behavioural process not a mental event (Hayes & Barnes-Holmes, 2004). Likewise, a verbal stimulus has its functions, in part, because it participates in relational frames.

In summary, the development of an appropriate behavioural account of language has made a behavioural approach to the study of the verbal phenomenon of attitudes possible. In other words, RFT "provides an alternative, behaviour-analytic approach to verbal events that is theoretically consistent, is built on existing principles, is in contact with some of the latest empirical evidence, and is fully subject to behaviour analysis

directed toward prediction and control" (Hayes & Wilson, 1993, p. 228).

Relational Frame Theory and attitudes. RFT proposes that attitudinal behaviour is verbal responding to an attitude-object that involves transformation of *evaluative* stimulus functions with respect to that object. Grey and Barnes (1996) conducted the first empirical behaviour analytic study designed to model attitudes as verbal phenomena. This study involved two experiments which sought to examine the contribution of stimulus equivalence to the formation of attitudes towards stimuli that had not previously been directly paired with an attitude-forming event and in the absence of direct reinforcement with those stimuli. In Experiment 1, participants were trained using a match-to-sample procedure to form three three-member equivalence relations (A1-B1-C1; A2-B2-C2; A3-B3-C3) using nonsense syllables as stimuli. One member from two of these classes (B1 and B2) was placed on a label attached to one of two video cassettes. The videos contained scenes of either a religious or romantic nature. After viewing the videos, participants were presented with four new videos. The new videos were labelled with the remaining nonsense syllables from the equivalence training (i.e., A1, C1, A2, C2). In order to examine the influence of their participation in equivalence classes, the next task asked participants to categorize the four unseen videos as "good" or "bad". The task served as a model for the phenomenon in which an individual forms an attitude about an object for which they have no direct experience. Results revealed that participants had derived equivalence relations in that they categorized the unseen videos in accordance with their evaluations of the originally viewed videos.

In Experiment 2, Grey and Barnes (1996) demonstrated a stimulus equivalence

model of attitude formation and change. Firstly, in order to determine if performance on the categorization tasks could be manipulated, contextual control was incorporated into the procedure through equivalence relations. To achieve this, match to sample training was provided which served to make the phrases “moral content” and “dramatic presentation” members of two separate equivalence relations along with a number of arbitrary stimuli. Participants were then tested in these derived stimulus relations. Following this task, participants were presented with a sexually violent video that was labelled with one of the nonsense syllables in the remaining relation (i.e., B3) from the equivalence training provided in Experiment 1. The final part of the study involved presenting participants with the same categorization tasks as those completed in Experiment 1. Results showed that the categorization of the videos came under the contextual control of two arbitrary stimuli because of their participation in equivalence relations with the two phrases (i.e., “moral content” and “dramatic presentation”). More specifically, when a participant was required to categorize a sexually violent video given a contextual cue that participated in an equivalence relation with “moral content”, the video was categorized as “Bad.” However, when the contextual cue participated in an equivalence relation with “dramatic presentation”, the video was categorized as “Good”.

Grey and Barnes (1996) also established that watching the sexually violent content altered the evaluative functions of some of the videos. For example, participants who categorized the videos with sexual content as morally bad in the first categorization task, failed to retain this classification after watching the sexually violent material. In other words, participants changed their attitudes towards other stimuli in this response class.

Overall therefore, Grey and Barnes (1996) offered a basic empirical model of the formation of attitudes as a transformation of evaluative stimulus functions through stimulus equivalence. Moreover, they suggested that contextually controlled transfer, in particular, may explain the common finding that people report different attitudes on the same issues in different contexts (Eagly & Chaiken, 1993).

The Relational Elaboration and Coherence model: A functional account of implicit and explicit attitudes. RFT has recently offered the Relational Elaboration and Coherence (REC) model in an effort to formally account for the empirical and theoretical divergence between implicit and explicit attitudes (Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, in press). This approach has at its core, the notion that relational responses, like all behaviours, develop over time. Thus, when a stimulus is encountered, a relational response will occur relatively quickly, and this may be followed by additional relational responses. These additional relational responses may occur as a response to the stimulus itself or be directed toward the initial response to the stimulus. Given enough time, these additional relational responses will likely form a coherent relational network. Take for example a white participant who is presented with the image of a black man holding a gun. The first relational response to occur may well involve a negative evaluation based on a verbal history in which black men are repeatedly depicted (in the media) as violent and dangerous. On the other hand, further relational responding may involve quite a different evaluation, for example, "judging someone on the basis of their skin colour is wrong" and/ or "He may be a police man", and so on. Put simply, relational responding may be relatively quick and immediate or can involve broader relational networks.

It is these brief and immediate relational responses that the REC model views as forming the basis for so-called *implicit attitudes*. Conversely, from the perspective of the REC model the extended relational responding that is needed to produce a response that coheres with one or more other relational response(s) in the person's behavioural repertoire provides the basis for so-called *explicit attitudes*. The behavioural effects captured by both self-report and implicit procedures are believed to reflect arbitrarily applicable relational responding. However, responding can be either brief and immediate *or* extended and elaborated -- depending on the properties of the measurement situation (e.g. whether or not participants are given time or opportunity to elaborate).

The REC model also offers an account for the dissociation that commonly emerges between implicit and explicit procedures, in terms of relational coherence. A relational network is said to cohere when all of the individual elements relate to each other in a manner that is consistent with the reinforcement history typically provided by the verbal community for such relational responding. According to RFT, the verbal community constantly reinforces coherence (and punishes incoherence) within relational networks, to the extent that relational coherence itself becomes a conditioned reinforcer for most verbal humans. Take, for example, the statement, "John is taller than Paul and Paul is taller than Ringo, but Ringo is taller than John." It is likely that you can recognise the incoherent nature of this simple relational network, and question its veracity[1].

This quest for relational coherence also relates to our own verbal behaviour. For example, responding to a picture of a black man as "dangerous" (with no additional

[1] It is worth noting that there may be some degree of conceptual overlap between relational coherence and the concept cognitive consistency (e.g. Festinger, 1957; Gawronski & Bodenhausen, 2006). The latter refers to the process of assessing the logical consistency between two or more propositions based on the assignment of truth values and the application of syllogistic rules and logical principles.

information) may not cohere readily with additional relational responses, which follow that initial response, such as "I am not a racist and I treat everyone equally". In this example, the individual has produced an incoherent relational network, and as a result additional relational responding may follow in an attempt to resolve the incoherence. Consequently the initial response may be considered to be "wrong", and therefore divergence between implicit and explicit attitudes would be observed. In other words, individuals may "reject" their immediate and brief relational responses (automatic evaluations) if they do not cohere with their more elaborate and extended relational responding. In certain contexts, however, relational elaboration may reduce or remove the incoherence within a network. For example, when the functions of the original stimulus are transformed the incoherence may be resolved. In the example above, the individual may conclude that the black man in the picture does actually look rather dangerous, which would thus cohere with the original brief and immediate relational response to the picture. In sum, brief and immediate evaluative responses may or may not cohere with subsequent relational responding -- when they do cohere, implicit and explicit measures will typically converge, but when they do not they will generally diverge.

The REC model, therefore, aims to account for the divergence in behavioural effects produced on implicit and explicit attitude procedures by appealing to the same process of arbitrarily applicable relational responding, but focusing on the extent to which such responding is brief and immediate or extended and coherent[1]. At this point, it is worth

[1] Strictly speaking the REC model is not a single process model given that it allows for the involvement of other behavioral processes other than relational framing (e.g. respondent conditioning and primary stimulus generalization). That said, the REC model broadly explains the difference between implicit and explicit attitude measurement procedures in terms of the elaboration and coherence involved in the single process of relational framing.

noting that when implicit cognition is viewed as relational, rather than purely associative, an alternative, non-associative measure of implicit attitudes becomes feasible. In this respect, the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes, Barnes-Holmes, Hayden, Milne, Power & Stewart, 2006) is a new methodology, which has recently been developed.

The IRAP: a non-associative measure of implicit attitudes. The IRAP is a computerised response latency procedure designed to target stimulus relations rather than mental associations in memory. Specifically, the task involves presenting relational terms (e.g., *Similar, Opposite, More than, Less than*) so that the properties of the relations among the relevant stimuli can be assessed. Similar to other response-latency methodologies, the IRAP involves asking participants to respond quickly and accurately in ways that are either consistent or inconsistent with their pre-experimentally established verbal relations. The rationale behind the IRAP is that responding should be faster on consistent (e.g. Love *Similar* to Pleasant) relative to inconsistent trials (e.g. Love *Opposite* Pleasant) because brief and immediate relational responding will coordinate more often with consistent overt responding. The response time differential between consistent and inconsistent trials (defined as the IRAP effect) is assumed to provide a non-relative index of the strength of the verbal or relational responses being assessed.

To illustrate this more clearly, consider an IRAP designed to index automatic racial attitudes towards Black and White people. On each IRAP trial one of two label stimuli "Black People are" or "White People are" is presented at the top of the computer screen with either a positive (e.g., "Good," "Peaceful," "Clever") or negative (e.g., "Bad,"

"Violent," "Stupid") target stimulus presented in the centre of the screen and participants are required to choose between one of two response options (e.g. "True" and "False") presented at the bottom left and right of the screen. During a block of consistent trials, a response defined as consistent with prevailing white verbal contingencies (e.g., choosing *True* given *White Person* and *Good*) clears the screen for 400ms and presents the next trial. If an inconsistent response is emitted (e.g., choosing *False* given *White Person* and *Good*) a red X appears immediately under the target stimulus. To remove the red X and continue to the 400ms inter-trial interval, participants are required to emit the consistent response. In contrast, during inconsistent blocks participants are required to make an inconsistent response in order to progress from one trial to the next (a consistent response produces the red X).

The IRAP typically consists of a minimum of two practice blocks and a fixed set of six test blocks. Each block presents the same number of trials, comprised of what are defined as four different *trial-types*. The trial-types are created by presenting each label stimulus with each of two sets of target words (see Figure 1, for a schematic representation of the IRAP). Given the previous example, a block of consistent trials thus requires the following pattern of responses: *White People are – Positive – True; White People are– Negative – False; Black People are – Positive – False; Black People are – Negative – True*. A block of inconsistent trials requires the opposite response pattern. The feedback contingencies are reversed across successive blocks of the IRAP, and thus participants are exposed to an alternating sequence of consistent and inconsistent blocks. For each block of IRAP trials participants are typically required to reach a standard of 80% correct

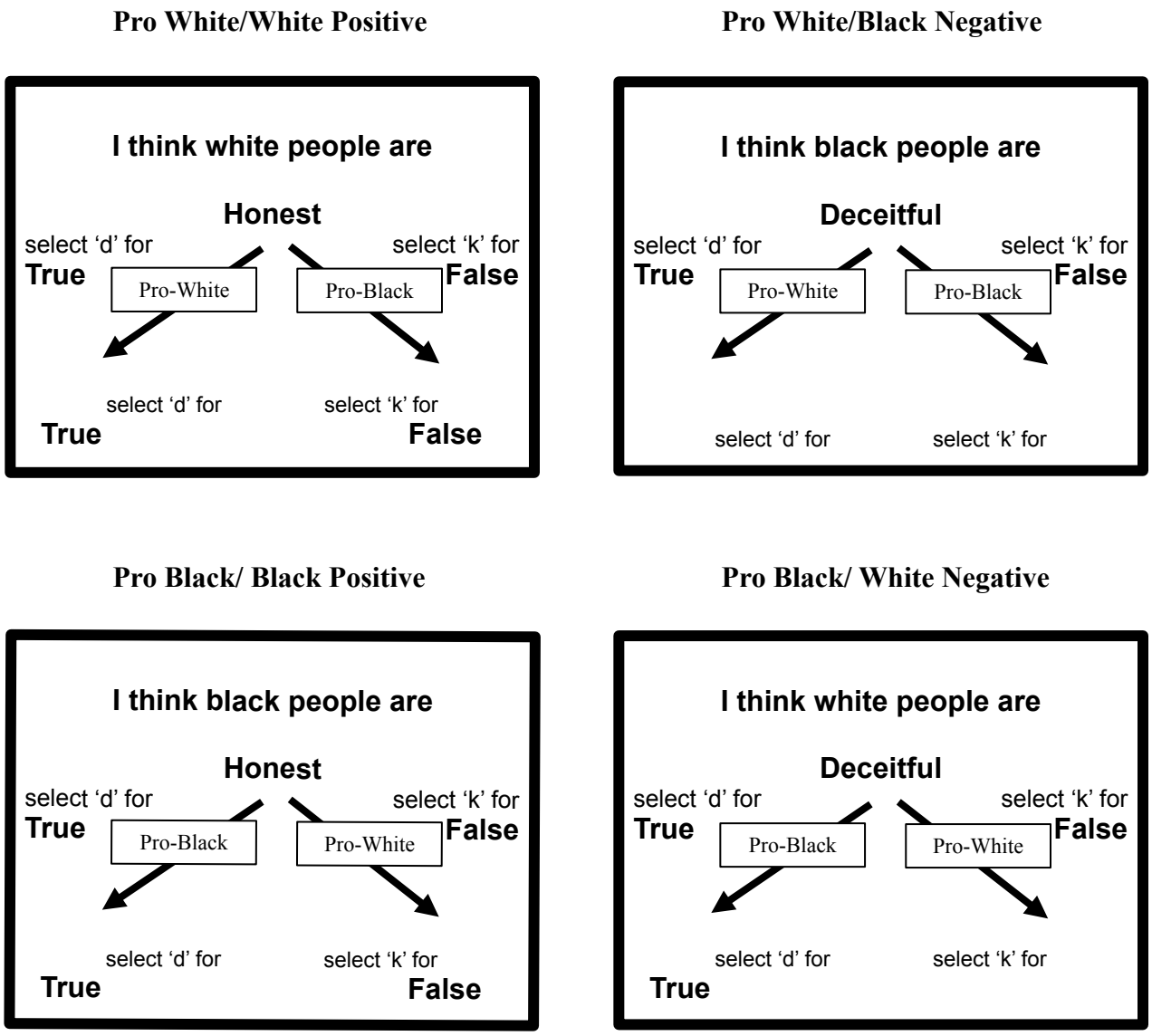


Figure 1. The four IRAP trial-types. The sample stimulus (e.g. 'I think Black People are' or 'I think White People are'), target word (positive or negative words e.g. honest, deceitful), and response options (True and False) appeared simultaneously on each trial. Arrows with superimposed text boxes indicate which responses were deemed pro-White or pro-Black (boxes and arrows did not appear on screen).

responses, and a median response time of less than 3000ms. Failure to maintain these criteria across successive test blocks results in the removal of data (see Barnes-Holmes et al, in press, for a more detailed overview of the procedure).

The IRAP differs from existing associative implicit measures, in that neither spatial nor temporal contiguity is manipulated across the task -- the presentation of the label and target stimuli remains unchanged throughout. However, the pattern of responding required by participants does change (responding *True* to *White* and *Good* in one block but responding *False* in another block), and thus the outcome of the measure is not readily attributable to the spatial and/or temporal association of stimuli within the procedure itself (the implications of this are discussed below).

The IRAP effect has now been replicated across a growing number of domains and studies have shown that the IRAP; (i) compares well with the IAT as a measure of individual differences (Barnes-Holmes, Murtagh, Barnes-Holmes, & Stewart, 2010; Barnes-Holmes, Waldron, Barnes-Holmes, & Stewart, 2010); (ii) demonstrates comparative levels of predictive validity to well-established procedures such as the IAT (Barnes-Holmes, Murtagh, Barnes-Holmes & Stewart, 2010; Barnes-Holmes, Waldron, Barnes-Holmes, & Stewart, 2009; Roddy et al., 2010), (iii) is not easily faked (McKenna, Barnes-Holmes, Barnes-Holmes, & Stewart, 2007); (iv) may be used as a measure of implicit self-esteem (Vahey, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009), implicit attitudes to work and leisure (Chan, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009), implicit ageism (Cullen, Barnes-Holmes, Barnes-Holmes Stewart, 2009), and deviant implicit attitudes in child sex offenders (Dawson, Barnes-Holmes, Gresswell, Hart & Gore, 2009); and (v) produces effects that clearly diverge from those obtained on explicit measures when targeting socially sensitive attitudes (Power, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009).

Critically, each of the studies outlined above required participants to respond directly to stimulus relations (or propositions), rather than to simple stimulus pairings or associations. Nevertheless, the IRAP produced behavioural effects that are typically defined as implicit attitude effects on associative procedures. From an associative perspective these findings could be perceived as counter-intuitive, in that the IRAP targets propositions and yet it reveals effects that are typically attributed to associations. It may eventually be possible to develop an associative account of the IRAP effects, however, at present, thinking relationally rather than associatively about implicit cognition has led to the development of a new methodology and data that stress possible limitations to a purely associative approach.

The REC model provides a non-associative account for the above mentioned IRAP effects. Put simply, according to the REC model, each IRAP trial involves asking participants to respond to the relationship between two stimuli. Each IRAP trial will therefore cause the participant to emit a brief and immediate relational response prior to pressing the appropriate computer key. The probability of this response will be determined by a combination of the participant's prior verbal and non-verbal learning history and current contextual variables. The most probable response will, by definition, be emitted first most often. Accordingly, on consistent IRAP trials the required key press will coordinate with the emitted response thus producing faster response latencies. Conversely, inconsistent IRAP trials require a key press that opposes the most immediate relational response emitted by the individual and, therefore, it occurs less quickly[1]. Thus across

[1] A potential behavioural explanation for the shorter latencies observed on consistent IRAP blocks is that relational coherence, as noted previously, is established by the verbal community as a conditional reinforcer. Thus brief and immediate relational responding coheres or coordinates more frequently than not with subsequent relevant responding in the day-to-day verbal behaviours of most individuals.

multiple trials the average latency for inconsistent blocks will be longer than for consistent blocks. It should be noted that this interpretation of the IRAP effect precludes any appeal to mediating mental constructs and instead formulates an explanation in terms of behavioural events that may occur either overtly or covertly.

The Current Research Programme

At the time of writing only one published study had used the IRAP to measure implicit racism (Barnes-Holmes, et al., 2010). In this study, participants were presented with the words "Safe" and "Dangerous" as label stimuli and pictures of white and black men holding guns as target stimuli. The results revealed an anti-black bias on the IRAP, but only when responding to the label-target combination, *Dangerous-Black.* That is, participants responded "True" more quickly than "False" when indicating that a Black man holding a gun was dangerous. Furthermore, this effect was only observed when participants were required to respond within 2000ms on each trial.

Although promising, this one study is somewhat preliminary and many issues remain to be addressed before one can conclude that the IRAP provides a valuable measure of implicit racial bias in Ireland. For example, one of the defining features of an implicit measure is its resilience to motivating factors in socially sensitive contexts. For example, implicit racial bias should be observed irrespective of whether or not participants are motivated to conceal their prejudice (Plant & Devine, 1998; Boysen, et al., 2006). Furthermore, levels of implicit racial bias should differ across relevant social groups. For instance, white participants may produce an anti-black bias on an implicit measure, but black participants may not (see Dasgupta, et al., 2000; Greenwald et al., 2002; Monteith,

et al., 2001; Livingston, 2002; Ottaway, et al., 2001).

The purpose of the research presented in the current thesis was to determine if the IRAP may be used as a reasonably reliable and valid measure of implicit racial bias in Ireland. To this end, the IRAP will be used to explore implicit racial prejudice across seven empirical studies focusing on: (a) its ability to detect participants' preferences for black and white people; (b) its malleability as a result of context manipulation effects; (c) the impact of response latency restrictions and personalisation; (d) known-group differences; (e) correlated neural activity; and (f) malleability as a result of intervention.

(b)*Malleability as a result of context manipulation effects*. In one of the first studies designed to assess the malleability of IRAP performance, Cullen, Barnes-Holmes and Barnes-Holmes (2009), employed the IRAP in an analogue of prior IAT-based work conducted by Dasgupta and Greenwald (2001), which examined the malleability of implicit anti-age bias. On each trial participants were presented with one of two sample stimuli ("Old People" or "Young People"), a target stimulus that was a positive (e.g., energetic, enthusiastic) or a negative adjective (e.g., tired, weary), and two response options ("Same" and "Opposite"). Thus, four trial-type response patterns were presented across this task; young-positive, young-negative, old-positive, and old-negative. Cullen et al. (2009) reported that exposure to images of positive old exemplars resulted in reduced levels of pro-young bias (smaller effects on the young-positive and young-negative trial-types) and reversed the anti-old effect (inverse effects on the old trial-types), whereas pro-young exemplars were found to have no effect on IRAP responses; in addition, they found that this effect persisted over 24 hours. Interestingly, exemplar exposure was not found to

influence explicit attitudes. Importantly, the IRAP was able to provide more detailed insight than the IAT on the influence of exemplars on age-related bias by revealing how the intervention affected particular trial-types. Specifically and as outlined above, the IRAP revealed that pro-old exemplars weakened the pro-young and reversed the anti-old biases. These findings indicate that relational responding as revealed by the IRAP may be affected by presenting participants with relevant exemplars beforehand and that the observed effects can be durable in the short-term. The IAT employed in the Dasgupta and Greenwald (2001) study could only identify reduced levels of implicit pro-young/anti-old bias overall whereas the IRAP could show the effect of exemplars on responding to young and old independently.

In a second IRAP investigation into the malleability of implicit responding, Barnes-Holmes, Murphy, Barnes-Holmes and Stewart (2010) examined the sensitivity of pro-white/anti- black responding as revealed by the IRAP to a public-private assessment manipulation. This study was an analogue of a previous IAT study conducted by Boysen, Vogel and Madon (2006) examining implicit homonegativity in public and private assessment contexts. In the public condition in Barnes-Holmes, Murphy et al. study: (1) the experimenter told the participant that she (the experimenter) would be able to see the levels of bias recorded on the IRAP; (2) the experimenter sat beside the participant while he or she completed the task; and (3) participants were required to tell the experimenter each of their answers on a series of self-report measures. Meanwhile, in the private condition: (1) participants performed the IRAP task alone; (2) they were informed that the

experimenter would collect but would not examine their IRAP scores; and (3) they filled out the self-report measures but were explicitly told not to record any personally identifiable information, because all responses were to remain confidential.

It was hypothesised that there would be less implicit stereotyping in the public than in the private condition; however, this was not quite the pattern observed. Instead, while participants in the private condition demonstrated significant pro-black attitudes, participants in the public context showed significant pro-white attitudes. To further investigate, therefore, additional empirical work was conducted examining IRAP performance in a public setting but using a 2000ms response latency criterion, which permitted comparison with the data from the initial investigation in the public setting. Results showed that participants responded with greater pro-white and anti-black bias than in the previous study, while self reported (explicit) attitudes indicated neutral or positive racial bias. These findings thus highlighted that relatively "fast" responding is needed on the IRAP when examining socially sensitive attitudes, as this increases the "implicitness" of the responses under investigation.

In summary, recent research has examined the malleability of implicit prejudicial responses on the IAT and IRAP by investigating the effect of contextual factors on the subsequent extent of negative, and socially sensitive, attitudes on these measures. The current work will extend these previous investigations in the context of a broader race-IRAP.

(c) *The impact of response latency restrictions and personalisation*. Previous IAT studies (see Payne, Govorun, & Arbuckle, 2008) have demonstrated the malleability of the

IRAP as a consequence of personalisation. For example, the use of phrases such as 'I Think" has been found reduce prejudicial responses on the IAT. In addition, as outlined above, Barnes-Holmes, Murphy et al., (2010) found that response latency restrictions were crucial and the current work sought to extend these findings.

(d) *Known-group differences*. Previous research using implicit measures has shown that white participants tend to show a relatively strong in-group pro-white bias (e.g., Noseck, et al., 2002; Pena, Sidanius, & Sawyer, 2004), the opposite pattern has not been observed for black participants. Rather, black participants tend to show a relatively weak out-group, pro-white bias (Nosek, et al., 2002; Pena, et al., 2004). Nevertheless, the difference in positive bias towards whites does differ significantly between the groups, with whites showing a stronger bias than blacks. Thus, insofar as the IRAP is functionally similar to other implicit measures, one might predict that both white and black participants will produce evidence of pro-white bias on the IRAP, although the white bias will be significantly stronger than the black bias.

(e) *Correlated neural activity*. Psychophysiological assessment tools such as ERPs (time-locked EEGs) have been suggested as viable and useful assessment techniques which can circumvent the limitations associated with self-report measures (Ito & Cacioppo, 2007). Furthermore, such measures may provide convergent validity for other less direct measures of biased responding. The very first IRAP study involved collecting both IRAP and electroencephalogram (EEG) data (Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, 2008), and the results showed different patterns of EEG activity across blocks of consistent versus inconsistent trials on the IRAP.

(f) *Malleability as a result of intervention*. A critical issue in the area of racial prejudice concerns the development of methods that may be used to modify or undermine such prejudice. There has been a growing increase in research in this area (Bargh, 1999; Blair, 2002; Cullen, et al., 2009). At the time of writing no published study had demonstrated the malleability of implicit racial bias as measured by the IRAP (cf. Barnes-Holmes, Murphy, et al., 2010). Furthermore, no published IRAP study had attempted to investigate the impact of educational or other types of interventions designed to undermine racial prejudice. Thus the current research programme sought to explore this issue.

Chapter 2 presents the first study of the thesis. This initial study directly explored the context sensitivity of the IRAP. Chapter 3 presents the second study of the thesis, which investigated the impact of response latency restrictions on IRAP performances. Chapter 4 presents Experiment 3, a study that further explored the impact of response latency restrictions on the IRAP. Chapter 5 presents the fourth study of the thesis, which explored the impact of personalising the labels on IRAP performances. Chapter 6 presents the fifth experiment, which involved a known-groups study exploring the predictive validity of the IRAP. Chapter 7 presents the sixth experiment, which investigated the relationship between neural activity and IRAP responses. Chapter 8 presents the final experiment in the current research programme, which investigated the malleability of the IRAP as a result of an acceptance- versus an education-based intervention. Chapter 9 concludes the thesis with a review and discussion of the main findings across the seven empirical studies.

CHAPTER 2

EXPERIMENT 1

CHAPTER 2

EXPERIMENT 1

As outlined previously, numerous studies using the well-known IAT have indicated that white participants tend to show a relatively strong pro-white/anti-black bias (see Dasgupta, et al., 2000; Greenwald et al., 2002; Monteith, et al., 2001; Livingston, 2002; Ottaway, et al., 2001) and that this pro-white/anti-black bias occurs for participants who explicitly state that they hold no racist attitudes. At the time of writing only one published study had used the IRAP to measure implicit racism (Barnes-Holmes, et al., 2010). This study employed only two words -- "Safe" and "Dangerous" as label stimuli and pictures of white and black men holding guns as target stimuli. Similar to the IAT literature, this preliminary study revealed an anti-black bias (although, only when responding to the label-target combination, *Dangerous-Black* and only under strict time constraints).

One criticism of the Barnes-Holmes et al. (2010) could be that the IRAP targeted only one specific dimension of racial prejudice, specifically safe versus dangerous. Furthermore, given the common portrayal of black males in the North American and British media as violent gun-carrying gang members, it could be argued that the resulting IRAP effects were hardly surprising because participants were asked to respond to pictures of black (and white) men holding guns. At the beginning of the research programme, therefore, it seemed important to determine if anti-black IRAP effects would be obtained again if participants were asked to respond to a range of negative-versus-positive attributes and to respond to statements rather than pictures. If the basic effect was replicated this would indicate that the IRAP may provide a general measure of racial bias, rather than one

that is restricted to a particular dimension.

Specifically, Experiment 1 sought to further develop the IRAP as a measure implicit racial stereotyping in Ireland, using a broader array of label stimuli and written statements rather than picture target stimuli. Similar to Barnes-Holmes, et al. (2010), Experiment 1 manipulated the assessment context in which the IRAP was completed. The aim of the experiment was to determine if manipulating the private versus public context of the assessment situation would impact upon the IRAP effects in a similar manner to that observed with the IAT in the Boysen, et al. (2006) study (i.e., a reduction in implicit in-group bias in the public relative to the private assessment context).

Method

Participants

Forty-two participants, 19 males and 23 females aged 18 to 37 years ($M = 25$), completed the experiment individually in the Department of Psychology at the National University of Ireland, Maynooth. All participants were white, Irish citizens and were randomly assigned to one of two assessment contexts — Public (21 participants) and Private (21 participants). No inducements were offered for participation in the study. All participants were experimentally naïve. Fifty-one individuals commenced the experiment, but the data from nine participants were excluded because they failed to achieve pre-determined performance criteria on the IRAP (described below) -- this level of attrition is not unusual.

Materials and Apparatus

Discrimination and Diversity Scales. All participants were given four explicit self-report measures to complete. The Discrimination (DS) and Diversity (DV) scales, created by Wittenbrink, Judd, and Park (1997), required participants to indicate on five point scales their agreement or disagreement with a total of 14 statements, 1 = strongly agree and 5 = strongly disagree. The DS scale consisted of ten statements concerning beliefs about discrimination within Irish society (e.g., "These days, reverse discrimination against Whites is as much a problem as discrimination against Blacks itself"). The DV scale consisted of four statements and targeted beliefs about the value of ethnic diversity within society (e.g. "There is a real danger that too much emphasis on cultural diversity will tear Ireland apart"). The questionnaires were scored such that 1 or 2 indicated negative racial stereotyping, 4 or 5 indicated positive racial stereotyping, and 3 indicated no stereotyping.

Semantic differential scales. Participants were required to complete 12 seven-point semantic differential scales, six for black people and six for white people. Each scale ranged from -3 to +3. These seven-point scales were anchored at either end by the following polar-opposite adjective pairs (taken from the IRAP): friendly – hostile, honest – deceitful, hardworking – lazy, peaceful – violent, good – bad and clever – stupid. To use the semantic differential scales as an explicit measure of racial stereotyping, the average ratings given for the White targets were subtracted from the average ratings given to the Black targets for each participant. Thus a positive score indicated pro-black stereotyping and a negative score indicated pro-white stereotyping.

Feeling thermometers. Two feeling thermometers assessed the favourability of

participants' explicit feelings about white and black people. Participants were asked to mark an appropriate position on a picture of a thermometer numerically labelled at 10º intervals from 0º (cold or unfavourable) to 99º (warm or favourable).

Motivation to conceal prejudice scales. The internal and external motivation to respond without prejudice scales (the IMS and EMS, respectively), created by Plant and Devine (1998), asked participants to indicate on nine-point scales their agreement or disagreement with a total of 10 statements, 1 = strongly disagree and 9 = strongly agree. The IMS scale consisted of five statements concerning beliefs about internal motivation to respond without prejudice (e.g., "I am personally motivated by my beliefs to be non-prejudiced to Black people"). The EMS scale consisted of five statements and targeted beliefs about external motivation to respond without prejudice (e.g. "I try to hide any negative thoughts about Black people in order to avoid negative reactions from others"). The questionnaires were scored such that high scores on either scale indicated a large degree of motivation to conceal prejudice.

Implicit Relational Assessment Procedure (IRAP). All participants completed the IRAP on a personal computer (Dell Pentium 4®). The IRAP software was used to present the stimuli and record participants' responses. Each IRAP trial presented one of two statements; "I think BLACK people are" or "I think WHITE people are". One of twelve target stimuli were also presented, and these consisted of six stereotypically positive words ("Friendly", "Honest", "Hardworking", "Peaceful", "Good", "Clever") or negative words ("Hostile", "Deceitful", "Lazy", "Violent", "Bad", "Stupid"). Finally, each trial presented two response options, "True" and "False". The program also presented the IRAP

instructions and a consent form.

Procedure

Public and Private contexts. Participants in the Public-context were given a form consisting of a "Public" statement, which they had to read and then indicate that they understood the information by providing a written summary:

> You are about to take a measure of racial prejudice on a computer. When you finish the test the computer will calculate the level of bias you have towards Black people on a scale from 0, meaning low bias, and 100, meaning the most bias possible. After I record your computer score, your bias will also be evaluated using some surveys.

This statement was used as a tool to elicit feelings of social desirability within the Public-context group, such that these participants may attempt to appear less racially biased on both the explicit measures (DS and DV scales, Semantic differential scales, Feeling thermometers) and the implicit measure (IRAP). The participants in the Private group were not given this form to fill out, and were told that the experimenter would collect but not examine their scores, with the implication that *individual* levels of racial stereotyping would remain unknown.

In both the Public and Private assessment situations, the experimenter sat adjacent to the participant and watched as he or she responded to the IRAP practice blocks. The experimenter then left the room while the participant completed the IRAP test blocks and did not return until the computer task was finished. For the explicit measures, participants in the Private-context were given the four scales in booklet form and told to fill them out

by circling the numbers that corresponded to their own feelings; they were also told not to mark the booklet in any other way (such as writing their name on it) because their answers were confidential. Public-context participants were given the booklet to read, and were also required to record their names on the coversheet because their answers were not confidential.

Implicit measure. The IRAP program began with a set of instructions, which described the task by illustrating the layout of the screen and explaining the response options (available from the first author upon request). The instructions informed participants that on each trial one of two statements, "I think BLACK people are" or "I think WHITE people are", would appear at the top of the screen along with a target word in the center of screen. Participants were also told that the response options "True" and "False" would appear at the bottom of the screen, and they were required to choose one of these options on each trial, by pressing either the 'd' or 'k' key; they were told that the left-right positions of these response options would switch randomly from trial-to-trial. The instructions also explained that the IRAP consisted of four different trial-types and illustrated examples of these were provided. In explaining these trial-types, participants were informed that sometimes they would be required to respond in a way that was consistent with their beliefs and at other times they would have to respond in a way that was inconsistent with their beliefs. Participants were assured that this was part of the experiment, and it was important for them to respond as quickly and accurately as possible on all trials of the IRAP (at no point was a participant informed which part of the experiment would be contradictory to their beliefs). The instructions also informed

participants that correct responses would allow them to progress to the next trial, but incorrect responses would produce a red 'X' in the middle of the screen, which could only be removed by pressing the correct key.

The IRAP task consisted of a minimum of two practice blocks and fixed set of six test blocks. Each block presented the same 24 trials, comprised of what are defined as four different trial-types (see Figure 1). The first block of the IRAP was designed to be consistent with pro-white stereotyping (e.g., *I think WHITE people are – Positive – True; I think BLACK people are – Positive – False; I think WHITE people are – Negative – False; I think BLACK people are– Negative – True*). The feedback contingencies alternated from block to block between pro-white and pro-black. Thus, in the second block of the IRAP correct responses were the opposite to the previous block (e.g., *I think WHITE people are – Positive – False; I think BLACK people are – Positive – True; I think WHITE people are – Negative – True; I think BLACK people are – Negative – False*). Before each new block began, the participants were informed that the previously correct and wrong answers would be reversed. The order in which IRAP blocks were presented was not counterbalanced across participants because previous research has found that this variable does not interact significantly with the critical IRAP effect (e.g., McKenna, et al., 2007; Power, et al. 2009; Vahey, et al. 2009).

Each IRAP block consisted of 24 trials, with each target stimulus presented once in the presence of each of the two statements. The trials were presented quasi-randomly with the constraint that none of the four trial-types could be presented twice in succession. The positioning of the two response options was also quasi-random in that they could not

appear in the same position three times in succession. For the first two practice blocks, participants were informed that it was a practice phase and errors were expected. Participants were required to reach a standard of >=80% correct responses, and a median response time of <=3000ms. These criteria were used to ensure that participants understood, and were complying with the IRAP instructions. If participants failed to achieve the two criteria for either of the two practice blocks, the required standard, and the standard of responding they had achieved, were presented on the screen. Participants were allowed three attempts (a total of six practice blocks) to achieve the practice criteria, and if they failed to do so, they were thanked, debriefed and their data were discarded (two participants were removed from the study on this basis). Participants who did achieve the practice criteria proceeded to the six test blocks.

The procedure for the first test block was similar to the first practice block, except that on-screen instructions informed participants that the next phase was a test and to "go quickly", although making "a few errors is okay". The second test block was similar to the second practice block, but with the modified instructions to go quickly. Test blocks 3 and 5 were the same as block 1, and test blocks 4 and 6 were the same as block 2. No performance criteria were applied during the test blocks in order to proceed, but if a participant's performance fell below 80% accuracy for any test block the data for that participant were discarded (seven participants were removed from the study on this basis). When all six test blocks had been completed participants reported to the researcher.

Explicit measures. Participants were given the four explicit measures to complete. The Discrimination (DS) and Diversity (DV) scales (see Appendix A), semantic

differential scales (see Appendix B), feeling thermometers (see Appendix C), and Internal and External Motivation to Conceal Prejudice scales (IMS and EMS respectively, see Appendix D). As noted earlier, the Private group completed the four scales by circling the appropriate numbers on the questionnaires whereas Public participants completed the four scales *and* recorded their names on the coversheet. The participants were then thanked, debriefed, and any questions were answered. All participants completed the experiment in a single session that lasted approximately 20-30 minutes.

Results and Discussion

Implicit Measure

Data preparation. The primary datum was response latency defined as the time in milliseconds that elapsed between the onset of a trial and a correct response emitted by a participant. To control for individual variations in speed of responding that may act as a possible confound when analyzing between group differences, the response latency data for each participant were transformed into *D*-IRAP scores (Barnes-Holmes, Murtagh, et al., 2010; Barnes-Holmes, Waldron, et al., 2009; Cullen & Barnes-Holmes, 2009; Vahey, et al., 2009) using an adaptation of the Greenwald, Nosek, and Banaji (2003) *D*-algorithm.

The steps involved in calculating the *D*-IRAP scores were as follows: (1) only response-latency data from the six test-blocks were used; (2) latencies above 10,000ms were removed from the dataset; (3) if the data from a participant contained more than 10% of test-block trials with latencies less than 300ms that participant was removed from the analyses; (4) twelve standard deviations for the four trial-types were calculated: four for the response-latencies from test-blocks 1 and 2, four from test-blocks 3 and 4, and a

further four from test-blocks 5 and 6; (5) 24 mean latencies were then calculated for the four trial types in each test-block; (6) difference scores for each of the four trial types were calculated, for each pair of test blocks, by subtracting the mean latency of the pro-white test-block from the mean latency of the corresponding pro-black test block; (7) each difference score was then 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; (8) four overall trial-type *D*-IRAP scores were then calculated by averaging the scores for each trial-type across the three pairs of test blocks; (9) an overall *D*-IRAP score was calculated by averaging all 12 trial-type *D*-IRAP scores from step 7.

The foregoing data transformation yields positive *D*-scores for positive bias, and negative scores for negative bias, towards *Whites*. In contrast, for the two *Black* trial-types negative *D*-scores indicate positive bias and positive scores indicate negative bias. In order to facilitate direct comparisons across the trial-types, the signs for the *Black* trial-type *D*-scores were reversed (i.e., + scores became – scores, and vice versa). Following this additional data transformation, positive *D*-scores now indicate positive bias towards *both Whites* and *Blacks* and negative scores indicate negative bias towards both groups (note, previously published IRAP studies have not included this final transformation). It should also be understood that the overall *D*-IRAP scores were calculated *before* reversing the signs for the two Black trial-types, and thus a positive overall *D* score indicates a pro-white/anti-black bias whereas a negative overall *D* score indicates a pro-black/anti-white bias.

Main analyses. A preliminary analysis of variance (ANOVA) indicated that there

were no main or interaction effects for the order in which pro-white versus pro-black blocks were presented (*ps* > .4), and thus this variable was removed from all subsequent analyses. The *D*-IRAP scores for the four trial-types in the Public and Private contexts are presented in Figure 2. All eight effects showed a positive bias. The strongest effects were produced on the *White-Positive* trial-type, and were similar across contexts. The *Black-Positive* trial-type also revealed a relatively strong effect for the Public but not for the Private context. Weaker effects were observed in both contexts for the two Negative trial-types, with the *Black-Negative* trial-type approaching zero in the Public context.

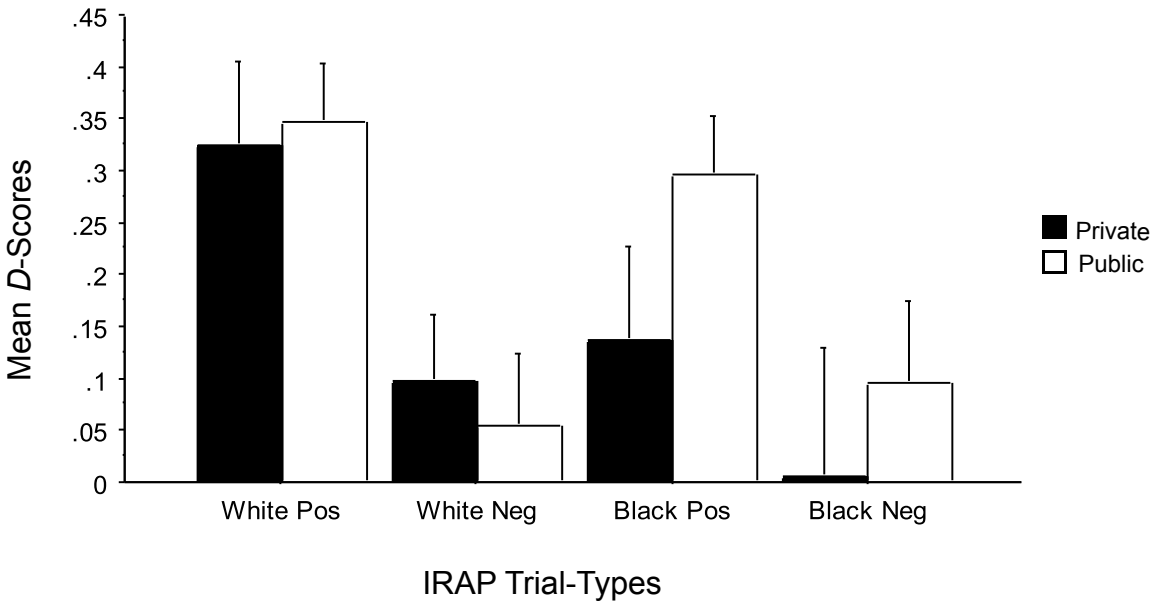


Figure 2. Mean *D*-IRAP trial-type scores, with standard error bars, for the Public and Private Assessment Situations.

A mixed repeated measures 2 x 4 analysis of variance (ANOVA) was conducted on the *D*-IRAP scores, with Private- and Public-contexts as the between-participant variable and trial-type as the within-participant variable. There was a significant main effect for trial-type, $F(3, 40) = 5.507$, $p < .001$, $\eta_p^2 = .12$, but no effect for context or interaction (*ps*

> .3). Fisher's PLSD post-hoc analyses revealed significant differences between *White-Positive* and *White-Negative* trial-types ($p < .01$) and between *Black-Positive* and *Black-Negative* trial-types ($p < .05$); no significant differences were obtained between the *Black-Positive* and *White-Positive* trial-types, or between the *Black-Negative* and *White-Negative* trial-types (*ps* > .1). That is, positive trial-types appeared to produce stronger IRAP effects than negative trial-types, but race had no significant effect on the IRAP performance.

Given that context had no significant main or interaction effects on the IRAP the data for Public and Private conditions were combined. The combined data for each trial-type were then subjected to one-sample *t*-tests to determine if the *D*-IRAP scores differed significantly from zero. The *White-Positive* effect was significant ($t = 6.93$, $df = 41$, $p <$ .0001), as was the *Black-Positive* effect ($t = 4.05$, $df = 41$, $p <$.0002). The effects for the two negative trial-types were non-significant (*ps* > .3).

Split-half reliability. To assess the internal consistency of the IRAP, five split-half reliability scores were calculated, one for each trial-type and one for the overall *D*-IRAP measure. In each case, two scores were calculated, one for odd trials and the second for even trials. These two scores were calculated in the same way as for the original *D*-IRAP scores, except that the algorithm was applied separately to all odd trials and to all even trials. The five split-half correlations between odd and even scores, applying Spearman-Brown corrections, proved to be weak and non-significant for the four individual trial-type scores: *White-Positive*, $r = .21$, $n = 42$, $p = .46$; *White-Negative*, $r = -.19$, $n = 42$, $p = .59$; *Black-Positive*, $r = .43$, $n = 42$, $p = .08$; *Black-Negative*, $r = .29$, $n = 42$, $p = .29$. The overall *D*-IRAP measure produced a weak to moderate and significant split-half

correlation, $r = .49$, $n = 42$, $p = .04$).

Explicit measures

Discrimination and diversity scales. The overall means for the DS scales revealed only a small difference between the Private ($M = 4.02$, $SD = .49$), and Public ($M = 3.92$, $SD = .56$) contexts, with both results revealing positive racial bias (i.e., mean scores above 3); a one-way ANOVA indicated that the difference was non-significant ($p > .5$). The overall means for the DV scales revealed a small difference between the Private ($M = 3.66$, $SD = .56$), and Public ($M = 3.24$, $SD = .58$) contexts, with both groups again showing positive racial bias. An ANOVA indicated that the effect was significantly more positive in Private than in Public, $F(1, 40) = 5.612$, $p = .02$, $\eta_p^2 = .12$.

Semantic differential scales. Four overall means were calculated for the semantic differential scales (Private/Black, $M = 5.00$, $SD = 5.5$; Public/Black, $M = 6.2$, $SD = 5.7$; Private/White, $M = 4.6$, $SD = 5.1$; Public/White, $M = 6.5$, $SD = 5.5$). Although more positive means were obtained in the Public context when rating both races, a 2x2 mixed repeated measures ANOVA revealed no significant effects ($ps > .3$).

Feeling thermometers. Four overall means were calculated, showing more positive means for the White relative to the Black scales in both contexts (Private/Black, $M = 70.9$, $SD = 14.4$; Public/Black, $M = 70.9$, $SD = 13.7$; Private/White, $M = 73.7$, $SD = 15.2$; Public/White, $M = 73.3$, $SD = 15.6$). A 2x2 mixed repeated measures ANOVA indicated that only the main effect for race was significant, $F(1, 40) = 4.762$, $p = .03$, $\eta_p^2 = .11$ (remaining $ps > .8$).

Motivation to conceal prejudice scales. Two overall means were calculated for

each motivation scale in each setting (Private/IMS, $M = 7.75$, $SD = .96$; Public/IMS, $M = 6.5$, $SD = 1.37$; Private/EMS, $M = 3.82$, $SD = 2.0$; Public/EMS, $M = 4.2$, $SD = 1.82$). The higher internal motivation in the Private context proved to be significant, $F(1, 40) = 11.716$, $p = .001$, $\eta_p^2 = .23$, but the higher external motivation in the Public context did not ($p > .4$)

Implicit-Explicit Correlations

Two correlation matrices of the implicit and explicit measures were calculated – one for the Public and one for the Private context. Each matrix thus involved correlating the four trial-type and overall *D*-IRAP scores with each of the eight explicit measures. Out of the 80 correlations only four were significant. The Public context yielded a significant negative correlation between the *White-Positive* trial-type and the black semantic differential scales (r -.43, $p < .05$), indicating that increased white bias on the IRAP predicted less positive explicit ratings for black. In addition, the Public context yielded two significant correlations that indicated a divergence between the implicit and explicit measures. Specifically, for the *Black-Negative* trial-type a greater pro-black bias predicted increased pro-white ratings on the semantic differential scales ($r = .484$, $p < .05$) and reduced pro-black ratings on the feeling thermometers ($r = -.454$, $p < .05$). Thus, the relationship between the *White-Positive* trial-type and the explicit measure appeared broadly consistent (i.e., pro-white predicted anti-black), but the relationship between the *Black-Negative* trial-type and the explicit measures did not (i.e., pro-black predicted pro-white and anti-black). Finally, the correlation matrix for the Private context yielded only one significant correlation, between the White-Positive trial-type and Internal Motivation

(r = -.579, p < .01), indicating that increased pro-white bias on the IRAP predicted reduced internal motivation to conceal racial prejudice.

Summary

The results from the IRAP failed to provide evidence for implicit anti-black stereotyping, in that all of the *D*-IRAP effects were positive and did not differ significantly across white versus black trial-types. Furthermore, the Public/Private manipulation failed to impact significantly on the IRAP measure. The results did indicate stronger effects for the two trial-types involving positive rather than negative target stimuli. Given the direction of the effects, it appears therefore that participants found it easier to confirm positive statements than to deny negative statements about white and black people.

In contrast to the IRAP, the explicit measures showed some sensitivity to both race and the Public/Private manipulation. Specifically, in the Private context racial diversity was endorsed more strongly, and internal motivation to conceal prejudice was higher, than in the Public context. Furthermore, the feeling thermometers yielded a small but significant pro-white bias, but no effect for context. Only three (out of 40) significant implicit-explicit correlations were obtained for the Public context, and these are difficult to interpret because one correlation showed convergence between the measures but the other two did not. For the Private context, only one (out of 40) implicit-explicit correlations were significant and this indicated that reduced internal motivation to conceal prejudice predicted stronger pro-white bias on the IRAP.

In conclusion, therefore, contrary to initial expectations the IRAP failed to produce evidence of racial bias and appeared largely insensitive to the Public/Private context

manipulation. Although the explicit measures showed some sensitivity to race and context, the effects were not clear cut. For example, participants endorsed greater racial diversity in the private rather than the Public context, but this effect may be explained by the fact that Private-context participants also reported higher levels of internal motivation to conceal prejudice. Furthermore, only a small number of significant implicit-explicit correlations were obtained and some of the effects appeared contradictory. Overall, therefore, the results of the current experiment, particularly with respect to the IRAP, were inconsistent with previous research. Consequently, in the next experiment two possibly important features of the IRAP were modified and participants were again exposed to a race IRAP in a Public or Private context.

CHAPTER 3

EXPERIMENT 2

CHAPTER 3

EXPERIMENT 2

Numerous published studies have shown a white in-group implicit bias using a range of measures (see Nosek, Smyth, Hansen, Devos, Lindner, Ranganath, Tucker Smith, Olson, Chugh, Greenwald, & Banaji, 2007, for a review), and yet no such effect was observed in the previous experiment using the IRAP. Shortly after Experiment 1 was conducted, however, a related IRAP study within the Maynooth laboratory did show an implicit pro-white/anti-black bias (Barnes-Holmes, Murphy, Barnes-Holmes & Stewart, 2010). Unlike the previous experiment, the study presented the sample words "Safe" and "Dangerous" with pictures of black and white men holding guns. Furthermore, a clear anti-black bias was observed only when participants were required to respond within 2000ms on each trial of the IRAP (rather than 3000). Indeed, Barnes-Holmes, et al. argued that reducing the response latency criterion served to increase the "automaticity" of the measure (see Moors & DeHouwer, 2006). Participants in Experiment 2 of the current study were thus required to respond within 2 seconds on each trial of the IRAP. Based on the assumption that personalizing an implicit measure may have unintended performance effects (Nosek & Hansen, 2008), a further modification was also made to the procedure by removing the phrase "I think" from the sample stimuli (i.e., only "Black People" and "White People" were presented as labels).

Method

Participants

Thirty-two participants, 15 males and 17 females aged 18 to 38 years (M = 24), completed the experiment individually in the Department of Psychology at the National University of Ireland, Maynooth. All participants were white, Irish citizens and were randomly assigned to one of two assessment contexts — Public (16 participants) and Private (16 participants). No inducements were offered for participation in the study. All participants were experimentally naïve. Forty-one individuals commenced the experiment, but the data from nine participants were excluded because they failed to achieve pre-determined performance criteria on the IRAP.

Materials and Apparatus

The apparatus and materials used in Experiment 2 were similar to those used in Experiment 1 (DS and DV scales, Semantic differential scales, Feeling thermometers, IMS and EMS scales). Note, however, the sample statements were shortened to *"BLACK people"* and *"WHITE people"* and the IRAP instructions were modified slightly to indicate that participants had to reach an average response latency of <= 2000ms across each practice block of the IRAP before they could proceed to the test blocks.

Procedure

The procedure in Experiment 2 was similar to that used for Experiment 1, except that the response latency practice criterion was reduced from 3000 to 2000ms, and all instructions and feedback were adjusted to reflect this change. Five participants failed to reach the practice criteria (i.e., >= 80% correct and a median response latency <=

2000ms), and thus did not proceed to the test blocks. The data for four additional participants were removed because their accuracy levels on one or more test blocks fell below 75% correct.

Results and Discussion

Implicit Measure

Data preparation. The response latencies were subjected to the same data preparation procedures as were employed for Experiment 1.

Main analyses. A preliminary analysis of variance (ANOVA) indicated that there was a significant main effect for order $F(3, 28) = 5.660$, $p < .02$, $\eta_p^2 = .17$, but no interaction effect with trial-type or context (*ps* > .58), and thus order was removed from subsequent analyses. The *D*-IRAP scores for the four trial-types in the Public and Private contexts are presented in Figure 3.

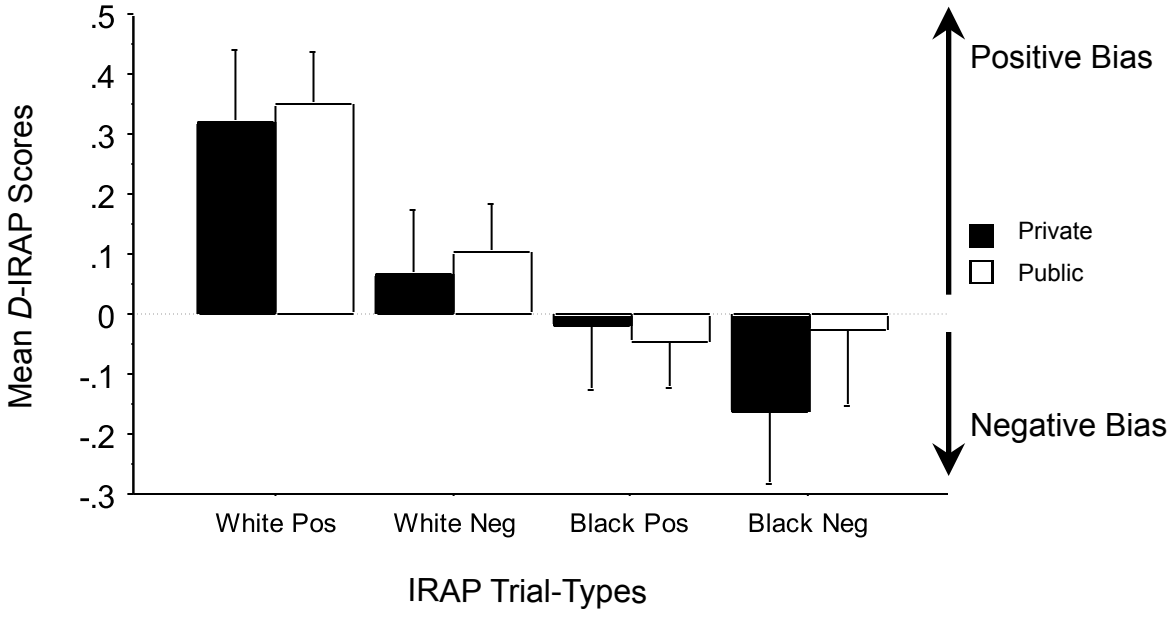


Figure 3. Mean White and Black *D*-IRAP trial-type scores, with standard error bars, for the Public and Private Assessment Situation.

The effects showed positive bias for the two white trial-types but negative bias for the two black trial-types. The *White-Positive* trial-type showed the strongest effect, although the *Black-Negative* trial-type produced a relatively strong negative bias in the Private context. Overall, context appeared to have limited impact on the *D*-IRAP effects. A mixed repeated measures 2 x 4 ANOVA was conducted on the *D*-IRAP scores, with Private- and Public-contexts as the between-participant variable and trial-type as the within-participant variable. There was a significant main effect for trial-type, $F(3, 30) = 5.822$, $p < .001$, $\eta_p^2 = .16$, but no effect for context or interaction (*ps* > .45). Fisher's PLSD post-hoc analyses indicated that the *White-Positive* trial-type produced significantly stronger positive bias than the other three trial-types (*ps* < .02), with no other significant differences (*ps* > .1).

Given that context had no significant main or interaction effects on the IRAP the data for Public and Private conditions were combined. The combined data for each trial-type were then subjected to one-sample *t*-tests to determine if the *D*-IRAP scores differed significantly from zero. The *White-Positive* effect was significant ($t = 4.586$, $df = 31$, $p <$ .0001), but the *White-Negative*, *Black-Positive* and *Black Negative* effects were not (*ps* > .2).

Split-half reliability. To assess the internal consistency of the IRAP, five split-half reliability scores were calculated (in the same way as for Experiment 1), and these were weak and non-significant for three of individual trial-type scores: *White-Positive*, $r = .198$, $n = 32$, $p = .55$; *White-Negative*, $r = .223$, $n = 32$, $p = .49$; and *Black-Positive*, $r = .05$, $n = 32$, $p = .89$. For *Black-Negative*, however, the effect was moderate to strong and

significant, $r = .627$, $n = 32$, $p = .007$. Finally, the overall *D*-IRAP measure produced a weak to moderate and significant split-half correlation, $r = .51$, $n = 32$, $p =. 05$.

Explicit Measures

Discrimination and diversity scales. The overall means for the DS scales revealed only a small difference between the Private ($M = 3.96$, $SD = .53$), and Public ($M = 4.16$, $SD = .43$) contexts, with both results revealing positive racial bias (i.e., mean scores above 3); a one-way ANOVA indicated that the difference was non-significant ($p > .2$). The overall means for the DV scales revealed a small difference between the Private ($M = 3.54$, $SD = .68$), and Public ($M = 3.20$, $SD = .54$) contexts, with both groups again showing positive racial bias. A one-way ANOVA indicated that the difference was also non-significant ($p > .1$).

Semantic differential scales. Four overall means were calculated for the semantic differential scales (Private/Black, $M = .7$, $SD = 1.02$; Public/Black, $M = 1.51$, $SD = 8.91$; Private/White, $M = .427$, $SD = .896$; Public/White, $M = 1.38$, $SD = .823$). More positive means were obtained in the Public context when rating both races and a 2x2 ANOVA revealed that this difference was significant $F(1, 30) = 8.560$, $p >.006$, $\eta_p^2 = .22$.

Feeling thermometers. Four overall means were calculated, showing more positive means for the White relative to the Black scales in both contexts (Private/Black, $M = 63.1$, $SD = 14.0$; Public/Black, $M = 71.2$, $SD = 15.8$; Private/White, $M = 64.3$, $SD = 12.0$; Public/White, $M = 74.3$, $SD = 16.6$). A 2x2 ANOVA indicated that the effect for context approached significance, $F(1, 30) = 3.495$, $p < .071$, $\eta_p^2 = .1$, but the effect for race and interaction did not ($ps > .28$).

Motivation to conceal prejudice scales. Two overall means were calculated for each motivation scale in each setting (Private/IMS, $M = 6.3$, $SD = 1.39$; Public/IMS, $M = 7.1$, $SD = 1.42$; Private/EMS, $M = 4.08$, $SD = 1.3$; Public/EMS, $M = 4.2$, $SD = 1.71$). Separate one-way ANOVA's for each scale revealed no significant differences (all $ps > .1$). Thus, unlike the previous Experiment, the Public-Private manipulation did not impact on motivation to conceal prejudice.

Implicit-Explicit Correlations

Two correlation matrices of the implicit and explicit measures were calculated – one for the Public and one for the Private context. Each matrix thus involved correlating the four trial-type and overall *D*-IRAP scores with each of the eight explicit measures. Out of the 80 correlations none were significant (all $ps > .15$).

Post-Hoc Analyses

In the current Experiment participants were required to respond within 2000ms on each trial of the IRAP, and were not permitted to continue to the test blocks if their median response latency on a practice block exceeded this criterion. Once participants started on the test blocks, however, they were not prevented from continuing if latency increased above the 2000ms criterion. It is possible, therefore, that latency may have floated above criterion for some participants during the test, leading perhaps to a reduction in the automaticity that the shorter latency criterion was designed to elicit. Consequently, the raw latency data for each of the participants were divided according the four trial-types and averaged across the consistent and across the inconsistent blocks, yielding eight mean latencies. If any of the eight latencies was greater than 2000ms, the data for that

participant were discarded. On this basis, the data for six participants were removed (one from the Private and 5 from the Public condition), and all of the previous analyses that were conducted on the IRAP data were repeated.

The 2x4 ANOVA again yielded a significant effect for trial-type, $F(3, 24) = 7.580$, $p < .0002$, $\eta_p^2 = .61$, with no effects for context or interaction (*ps* > .64). Interestingly, however, the Fisher's PLSD post-hoc tests indicated that the pro-white bias on the *White-Negative* trial-type was significantly different to the anti-black bias on the *Black-Negative* trial-type ($p = .01$); this difference was non-significant in the previous analyses. Furthermore, the previously significant difference between the *White-Positive* and *White-Negative* trial-types was no longer significant, and the non-significant difference between the *White-Negative* and *Black-Positive* trial-types now approached significance ($p = .06$). Overall, therefore, removing the data for participants who did not maintain the 2000ms practice latency criterion produced post-hoc effects suggestive of relatively stronger pro-white in-group and anti-black out-group implicit biases.

Given that context once again had no significant main or interaction effects the combined data for each trial-type were subjected to one-sample *t*-tests. Once again, the *White-Positive* effect was significant ($t = 4.495$, $df = 25$, $p < .0001$), and the two *Black* trial-type effects were not (ps > .14). In contrast to the previous analyses, however, the *White-Negative* effect was now significant ($t = 2.425$, $df = 25$, $p < .02$), suggesting again that removing "slow" test-block responders served to increase the implicit pro-white bias. Finally, analyses of the split-half reliabilities for the IRAP yielded similar results to the previous analyses.

Summary and Conclusion

As was the case in the previous experiment, the Private/Public context manipulation had no significant impact the IRAP effects. Unlike the previous experiment, however, IRAP effects indicative of pro-white in-group and anti-black out-group bias were observed. Furthermore, the in-group/out-group biases appeared to be strengthened when slow responding participants were removed from the data set. The split-half reliability of the IRAP was moderate and significant for the *Black-Negative* trial-type; in the previous experiment all of the reliabilities were weak and non-significant.

With regard to the explicit measures, the Public context increased the ratings for both races on the semantic differential scales and feeling thermometers (the latter approached significance). Contrary to the previous experiment none of the explicit measures produced significant effects indicative of racial bias, and furthermore context had no significant impact on motivation to conceal prejudice (in the previous experiment the Private context increased internal motivation). Finally, unlike Experiment 1, none of the implicit-explicit correlations were significant. At the current time, it is difficult to determine why these differences in the explicit measures emerged across the two experiments. However, given that the IRAP also produced unexpected results in the first experiment (i.e., no racial bias effects), it seems best simply to note the differences in the explicit measures at this point and to remain alert to this issue in subsequent experiments.

Overall, the current experiment produced IRAP effects indicative of in-group racial bias that appeared broadly consistent with the results of a related study (Barnes-Holmes, Murphy, et al., 2010). However, it remains unclear to what extent the different outcomes

for the two experiments were due to the reduced latency criterion or to the removal of "I think" from the sample statements. Experiment 3 was conducted to address this issue.

CHAPTER 4

EXPERIMENT 3

CHAPTER 4

EXPERIMENT 3

In Experiment 3 the phrase "I think" was reintroduced into the sample stimuli (i.e., as in Experiment 1, the current Experiment presented the phrases "I think Black People are" and "I think White People are"). Similar to Experiment 2, participants had to reach an average response latency of <= 2000ms across each practice block of the IRAP before they could proceed to the test blocks. Given that the public-private manipulation had no significant effects on the IRAP performances across the previous two experiments the "standard" private context was employed in the current experiment (i.e., participants were *not* told that the experimenter would know their level of racial bias). Although previous studies have found that the order in which IRAP blocks are presented does not moderate the IRAP effect, it was decided to check this variable once more in the context of the current research programme. Thus half of the participants were exposed to pro-white relations-first and the remaining participants were exposed to pro-black relations-first.

Method

Participants

Twenty-two participants, 12 males and 10 females aged 19 to 36 years (M = 26), completed the experiment individually in the Department of Psychology at the National University of Ireland, Maynooth. All participants were white, Irish citizens and all completed the experiment in a Private context. No inducements were offered for participation in the study. All participants were experimentally naïve. Thirty-two individuals commenced the experiment, but the data from ten participants were excluded

because they failed to achieve pre-determined performance criteria on the IRAP.

Materials and Apparatus

The apparatus and materials used in Experiment 3 were similar to those used in Experiment 2 (DS and DV scales, Semantic differential scales, Feeling thermometers, IMS and EMS scales), except that the sample statements were *"I think BLACK people are"* and *"I think WHITE people are"*.

Procedure

The procedure in Experiment 3 was similar to that used for Experiment 2 except that all participants completed Experiment 3 in a private setting and the order in which IRAP blocks were presented was counterbalanced across participants -- half of the participants were exposed to pro-white relations-first and the remaining participants were exposed to pro-black relations-first. Two participants failed to reach the practice criteria (i.e., >= 80% correct and a median response latency <= 2000ms), and thus did not proceed to the test blocks. The data for three additional participants were removed because their accuracy levels on one or more test blocks fell below 75% correct, and the data for five participants were removed because their mean response latencies were greater than 2000ms for one or more trial-types on one or more test-blocks.

Results and Discussion

Implicit Measure

Data preparation. The response latencies were subjected to the same data preparation procedures as were employed in the previous experiments.

Main analyses. The overall mean *D*-IRAP scores for the four trial-types, divided

by order, are presented in Figure 4. The effects showed positive bias for the two white trial-types and the *Black-Positive* trial-type, but negative bias for the *Black-Negative* trial-type; the *White-Positive* trial-type showed the strongest effect. The order in which IRAP blocks were presented appeared to have little effect on each of the four trial-types.

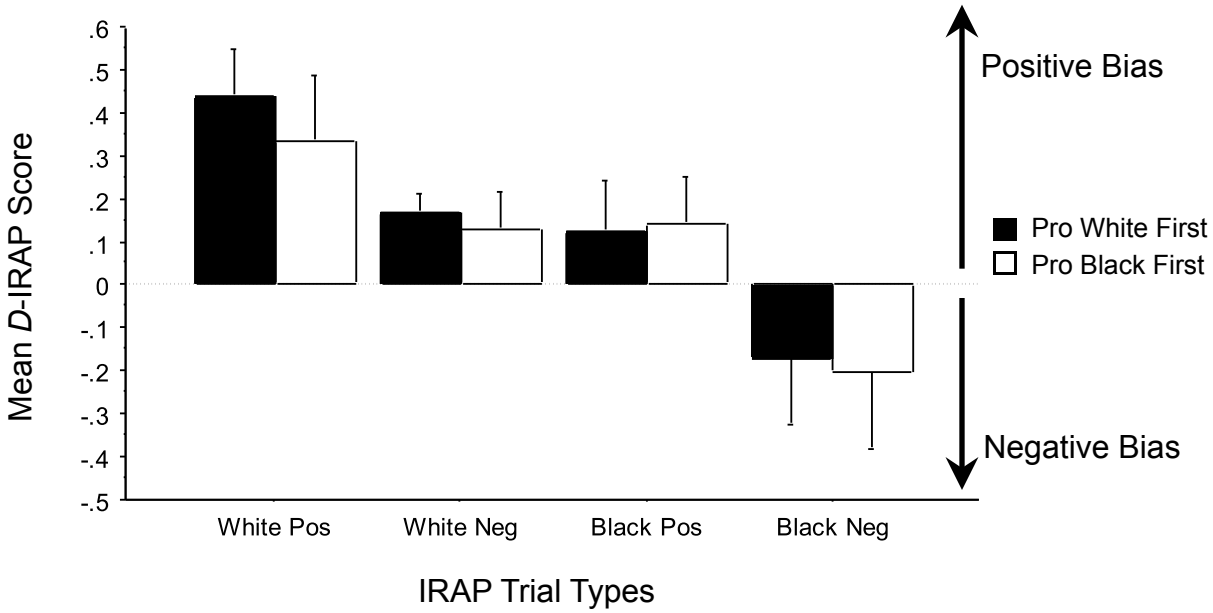


Figure 4. The mean *D*-IRAP scores, with standard error bars, for the four trial-types in the pro-white first and pro-black first conditions.

A mixed repeated measures ANOVA indicated that there was a significant main effect for trial-type, F (3, 20) = 7.128, p = .0004, η_p^2 = .26, but no main or interaction effects for order (*ps* > .6). Fisher's PLSD post-hoc analyses indicated that the *White-Positive* trial-type produced significantly stronger positive bias than the two black trial-types (*ps* < .04), and that the *Black-Negative* trial-type produced significantly stronger negative bias than the *White-Negative* and *Black-Positive* trial-types (*ps* < .01).

The combined data (across order) for each trial-type were then subjected to one-

sample *t*-tests to determine if the *D*-IRAP scores differed significantly from zero. The *White-Positive* effect was significant ($t = 4.257$, $df = 21$, $p = .0004$), as was the *White-Negative* effect ($t = 3.070$, $df = 21$, $p < .005$), but the *Black-Positive* and *Black-Negative* effects were not ($ps > .1$).

Split-half reliability. To assess the internal consistency of the IRAP, five split-half reliability scores were calculated (in the same way as for Experiments 1 and 2), and these were weak and non-significant for two individual trial-type scores: *White-Positive*, $r = .362$, $n = 22$, $p = .33$, and *Black-Positive*, $r = .455$, $n = 22$, $p = .19$. For *Black-Negative*, the effect was moderate and approached significance, $r = .558$, $n = 22$, $p = .07$. The correlation for *White-Negative* was negative, thus violating reliability model assumptions. Finally, the overall *D*-IRAP measure produced a weak to moderate but non-significant split-half correlation, $r = .438$, $n = 22$, $p = .21$.

Explicit Measures

Discrimination and diversity scales. The overall means for both the DS and DV scales revealed a positive racial bias ($M = 4.09$, $SD = .47$, and $M = 3.55$, $SD = .6$, respectively).

Semantic differential scales. Two overall means were calculated for the semantic differential scales, one for Black and one for White (Black, $M = 5.09$, $SD = 5.50$; White, $M = 4.40$, $SD = 5.19$). A more positive mean was obtained for Black relative to White, but this difference was non-significant ($p > .36$).

Feeling thermometers. Two overall means were calculated, showing more positive means for White relative to Black (Black, $M = 70.4$, $SD = 14.5$; White, $M = 71.7$, $SD =$

15.7), but this difference was non-significant ($p > .26$).

Motivation to conceal prejudice scales. Two overall means were calculated, one for the IMS and one for the EMS (IMS, $M = 7.81$, $SD = .89$; EMS, $M = 3.85$, $SD = 1.95$), and these showed broadly similar levels of motivation to the previous two experiments.

Implicit-Explicit Correlations

A correlation matrix of the implicit and explicit measures was calculated. This involved correlating the four trial-type and overall *D*-IRAP scores with each of the eight explicit measures. Out of the 40 correlations only three were significant (all other $ps > .1$). The *Black-Negative* trial-type was correlated negatively with both the black feeling thermometer ($r = -.45$, $p < .03$) and the white feeling thermometer ($r = -.45$, $p < .03$), indicating that greater negativity towards black people on the IRAP predicted increased positivity towards both black and white people on the two thermometers. Finally, the *Black-Negative* trial-type was correlated positively with the diversity scales ($r = .48$, $p < .01$), indicating that reduced negativity towards black people on the IRAP predicted greater endorsement of racial diversity in Ireland.

Post-Hoc Analyses

The current experiment produced IRAP effects indicative of in-group racial bias similar to those observed in Experiment 2. Unlike the previous experiment, however, the phrase "I Think" was included in the sample statements. In order to determine if the absence-versus-presence of the phrase impacted upon the IRAP performances, the data from Experiments 2 and 3 were subjected to a 2x4 mixed repeated ANOVA, with sample phrase as the between-participant variable and trial-type as the repeated measure (to be

consistent with Experiment 3, only the data from the Private condition in Experiment 2 were included, and the data for one participant were removed because the mean response latencies were greater than 2000ms for one or more trial-types on one or more test-blocks). The ANOVA failed to yield significant main or interaction effects for sample-phrase (*ps* > .28), thus indicating that personalizing the sample with "I think" did not have a significant effect on IRAP performance.

Summary and Conclusion

Similar to the previous experiment, the IRAP effects were indicative of pro-white in-group and anti-black out-group bias. Once again, split-half reliability was moderate only for the *Black-Negative* trial-type, but on this occasion it failed to reach significance. The order in which the IRAP blocks were presented did not appear to moderate the IRAP effects.

Consistent with Experiment 2, none of the explicit measures produced significant effects indicative of racial bias, and motivation to conceal prejudice was similar to previous experiments. Unlike Experiment 2, three of the forty implicit-explicit correlations were significant, and each of them involved the *Black-Negative* trial-type. Two of the correlations indicated that greater negativity towards black people on the IRAP predicted increased positivity towards both black and white people on the two feeling thermometers. Although apparently contradictory, this result may have emerged because at least some white participants attempted to match their out-group thermometer ratings of black people with their in-group ratings of white people. Certainly, the Motivation to conceal prejudice scales indicated that participants were internally motivated in this direction (i.e., to appear

non-prejudiced). On balance, however, the third significant correlation indicated that reduced negativity towards black people on the *Black-Negative* trial-type predicted greater endorsement of racial diversity in Ireland. It is difficult to explain this correlation in terms of participants simply matching a black rating to a white rating because all the statements on the DV scales pertain to black people. These data thus suggest an IRAP performance may correlate with one or more explicit measures of racial prejudice, although motivation to conceal prejudice may play an important moderating role in that implicit-explicit relationship. This issue is addressed directly in the next Experiment.

Overall, the current experiment again produced IRAP effects indicative of in-group racial bias that appeared broadly consistent with the results of a related study (Barnes-Holmes, Murphy et al., 2010). Furthermore, a comparison of Experiments 2 and 3 suggested that the presence-versus-absence of the phrase "I think" had no significant impact on the IRAP effects, although this latter conclusion is based on a post-hoc analysis. In the next experiment, therefore, the "I think" variable was targeted directly. In addition, it was noted that 5 participants in Experiment 3 were removed from the data set because their response latencies during the test blocks "drifted" over 2000ms on one or more trial-types. A modification to the IRAP software was thus introduced at this point in the research program. Specifically, the warning message "Too Slow!" was presented on any trial (practice or test) whenever a participant did not respond within 2000ms.

CHAPTER 5

EXPERIMENT 4

CHAPTER 5

EXPERIMENT 4

As noted above, the "I think" variable was manipulated directly in Experiment 4, and the warning message "Too Slow!" was presented if a participant failed to respond within the specified latency criterion (2000ms). The current experiment also aimed to investigate an issue that has been the focus of previous studies on implicit attitudes. Specifically, a number of studies have found that the relationship between implicit and explicit measures is moderated by participants' motivation to conceal the attitude under investigation (e.g., Payne, Govorun, & Arbuckle, 2006). That is, implicit measures may predict explicit measures when participants lack motivation to conceal the relevant attitude, but the predictive relationship is absent when motivation is high. A series of regression analyses will thus be used to determine if the relationship between the IRAP and explicit measures is also moderated by such motivation. Insofar as the IRAP overlaps, to some extent, with other implicit measures the IRAP may correlate with explicit measures for those participants who are low in motivation to conceal racial prejudice, but will not correlate for those participants who are highly motivated to conceal prejudice.

Method

Participants

Thirty-six participants, 17 males and 19 females aged 18 to 38 years (M = 24), completed the experiment individually in the Department of Psychology at the National University of Ireland, Maynooth. All participants were white, Irish citizens and were randomly assigned to one of two assessment contexts — "I think" (18 participants) and No

"I think" (18 participants). Within each assessment context, half of the participants were exposed to pro-white relations-first and half were exposed to pro-black relations-first. No inducements were offered for participation in the study. All participants were experimentally naïve. Forty-one individuals commenced the experiment, but the data from five participants were excluded because they failed to achieve or maintain performance criteria on the IRAP (described below).

Materials and Apparatus

The apparatus and materials used in Experiment 4 were similar to those used in Experiment 3 (IRAP, DS and DV scales, Semantic differential scales, Feeling thermometers, IMS and EMS scales). Note, however, the two IRAP sample statements for half of the participants were shortened to *"BLACK people"* and *"WHITE people"*. In addition response latency feedback was introduced. That is, if a participant took longer than 2000ms on any IRAP trial, the phrase "Too Slow!" was presented on the screen. The IRAP instructions were modified slightly to indicate these changes.

Procedure

The procedure in Experiment 4 was similar to that used for Experiment 3, except that the response latency feedback was introduced. Specifically, if a participant failed to emit a response (correct or incorrect) within 2000ms on any trial, the message "Too Slow!" appeared in the lower bottom centre of the screen, and remained there until a response was emitted. If the response was correct the screen cleared and the program progressed to the 400ms inter-trial interval; if the response was incorrect only the "Too Slow!" message was removed and the red X appeared (to continue to the inter-trial interval

a correct response was required). The pre-experimental instructions were adjusted to reflect the presentation of the "Too Slow!" feedback.

Four participants failed to reach the practice criteria (i.e., >= 80% correct and a median response latency <= 2000ms), and thus did not proceed to the test blocks. The data for one additional participant were removed because accuracy levels on one or more test blocks fell below 80% correct.

Results and Discussion

Implicit Measure

Data preparation. The response latencies were subjected to the same data preparation procedures as were employed for Experiment 3.

Main analyses. A preliminary mixed repeated measures 2x2x4 ANOVA indicated that there were no main or interaction effects for order or for "I think" *(ps* > .53). Four planned comparisons between-participant ANOVAs were conducted to insure that the "I think" variable had no impact on any of the individual trial-types and each of these was non-significant (*ps* > .22). At this point, therefore, the order and "I think" variables were removed from subsequent analyses.

The overall mean *D*-IRAP scores for the four trial-types are presented in Figure 5. The results showed positive bias for the two white trial-types, an almost neutral effect for *Black-Positive,* and negative bias for *Black-Negative*. A one-way repeated measures ANOVA revealed a significant main effect for trial-type, $F(3, 35) = 21.244$, $p < .0001$, $\eta_p^2 = .37$. Fisher's PLSD post-hoc analyses indicated that the *White-Positive* trial-type produced significantly stronger positive bias than the other three trial-types (*ps* < .006) and

revealed significant differences between the *White-Negative* and *Black-Negative* trial-types ($p < .0001$), and between the *Black-Positive* and *Black-Negative* trial-types; the difference between the *White-Negative* and *Black-Positive* trial-types approached significance ($p > .07$). One-sample *t*-tests indicated that the *White-Positive* effect was significantly different from zero ($t = 9.135$, $df = 35$, $p < .0001$), as were the *White-Negative* ($t = 2.270$, $df = 35$, $p < .02$) and the *Black-Negative effects* ($t = -4.511$, $df = 35$, $p < .0001$), but the *Black-Positive* effect was not ($p > .7$).

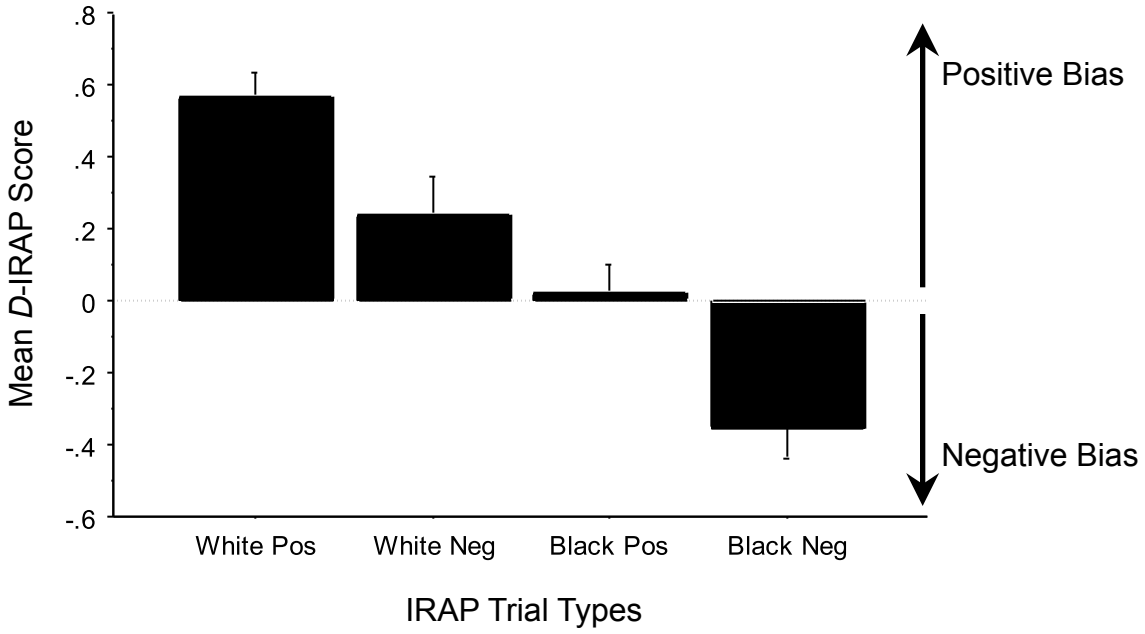


Figure 5. The mean *D*-IRAP scores, with standard error bars, for the four trial-types.

Split-half reliability. To assess the internal consistency of the IRAP, five split-half reliability scores were calculated (in the same way as for the previous Experiments) and these were moderate to strong and significant for three of individual trial-type scores: *White-Positive*, $r = .591$, $n = 36$, $p < .01$; *Black-Positive*, $r = .505$, $n = 36$, $p < .04$; and *Black-Negative*, $r = .739$, $n = 36$, $p < .0001$. For *White-Negative*, however, the effect was

weak and non-significant, $r = .128$, $n = 36$, $p > .7$. Finally, the overall *D*-IRAP measure produced a moderate to strong and significant split-half correlation, $r = .774$, $n = 36$, $p < .0001$.

Explicit Measures

Discrimination and diversity scales. The overall means for the DS scales revealed only a small difference between the "I think" ($M = 3.75$, $SD = .67$), and No "I think" ($M = 3.84$, $SD = .54$) sample stimuli, with both results revealing positive racial bias (i.e., mean scores above 3); a one-way ANOVA indicated that the difference was non-significant ($p > .6$).The overall means for the DV scales revealed another small difference between the "I think" ($M = 3.46$, $SD = .73$), and No "I think" ($M = 3.42$, $SD = .54$) samples, with both groups again showing positive racial bias. A one-way ANOVA indicated that the difference was also non-significant ($p > .8$).

Semantic differential scales. Four overall means were calculated for the semantic differential scales ("I think"/Black, $M = .79$, $SD = 1.14$; No "I think"/Black, $M = .9$, $SD = 8.6$; "I think"/White, $M = .87$, $SD = 1.14$; No "I think"/White, $M = .63$, $SD = .81$); all four means revealed a positive bias (> 0). A 2x2 mixed repeated measures ANOVA yielded no significant main or interaction effects ($ps > .14$).

Feeling thermometers. Four overall means were calculated, showing more positive means for the White relative to the Black scales given the "I think" samples (Black, $M = 66.5$, $SD = 19.8$; White, $M = 74.3$, $SD = 21.3$), but the opposite pattern given the No "I think" context (Black, $M = 70$, $SD = 11.9$; White, $M = 68.9$, $SD = 12.3$). A 2x2 mixed repeated measures ANOVA yielded a significant main effect for feeling thermometer, $F(1,$

34) = 4.403, $p < .05$, $\eta_p^2 = .1$, and a significant interaction, $F(1, 34) = 7.827$, $p < .01$, $\eta_p^2 = .2$. Four simple effects tests yielded one significant difference; participants who were presented with the "I think" samples in the IRAP produced thermometer ratings that were significantly greater for White than for Black ($p < .01$; remaining *ps* > .35).

Motivation to conceal prejudice scales. Two overall means were calculated for each motivation scale in each setting ("I think"/IMS, $M = 6.6$, $SD = 1.5$; No "I think" / IMS, $M = 6.5$, $SD = 1.3$; "I think"/EMS, $M = 4.8$, $SD = 1.5$; No "I think"/EMS, $M = 4.07$, $SD = 1.6$). Separate one-way ANOVA's for each scale revealed no significant differences (*ps* > .15).

Implicit-Explicit Correlations

A correlation matrix of the implicit and explicit measures was calculated. This involved correlating the four trial-type and overall *D*-IRAP scores with each of the eight explicit measures. Out of the 40 correlations none were significant (a*ll ps* > .1).

The Moderating Impact of Motivation to Conceal Prejudice

As noted previously, it was predicted that there should be some indication that implicit attitudes coincide with explicit attitudes for individuals who are not motivated to conceal racial prejudice. In contrast, individuals who are highly motivated in this regard should show equivalent implicit attitudes on the IRAP but under-report their racial prejudice.

IRAP trial-type regression analyses. A single motivation score for each participant was first obtained by averaging the combined IMS and EMS scores. Single feeling thermometer and semantic differential scores were obtained by subtracting the white from

black ratings. All variables were then standardized and a series of two-step multiple regressions were conducted. On the first step, one of the explicit attitude measures (feeling thermometer, semantic differential, DV or DS) was entered as the dependent variable with one of the four IRAP trial-types (*White-Positive, White-Negative, Black-Positive, Black-Negative*) and the single motivation score as predictors. In the second step, the IRAP trial-type x motivation interaction was entered. A total of 16 separate regression analyses were thus conducted, and one of these yielded a significant two-way interaction effect. Specifically, the *Black-Negative* trial-type interacted with motivation in predicting participants' ratings on the semantic differentials, $b = .393$, $t = 1.989$, $p < .05$. The second step added a significant increment in variance explained $\Delta R^2 = .11$, $F(3, 35) = 11.44$, $p = .05$. Regression lines relating the *Black-Negative* IRAP performance and semantic differential ratings are displayed at one standard deviation above and below the mean on motivation to conceal prejudice (Figure 6).

The figure illustrates that the *Black-Negative* IRAP performance was strongly related to the semantic differential rating for those individuals who were unmotivated to conceal their prejudice ($b = .41$). In contrast, the relationship was much weaker and in the opposite direction for those who were motivated to conceal prejudice ($b = -.19$). The interaction is driven by the fact that highly motivated participants tended to under-report their racial prejudice on the semantic differentials, but did not perform differently on the *Black-Negative* IRAP trial-type, as indicated by the lack of correlation between the IRAP and motivation to conceal prejudice scores reported above.

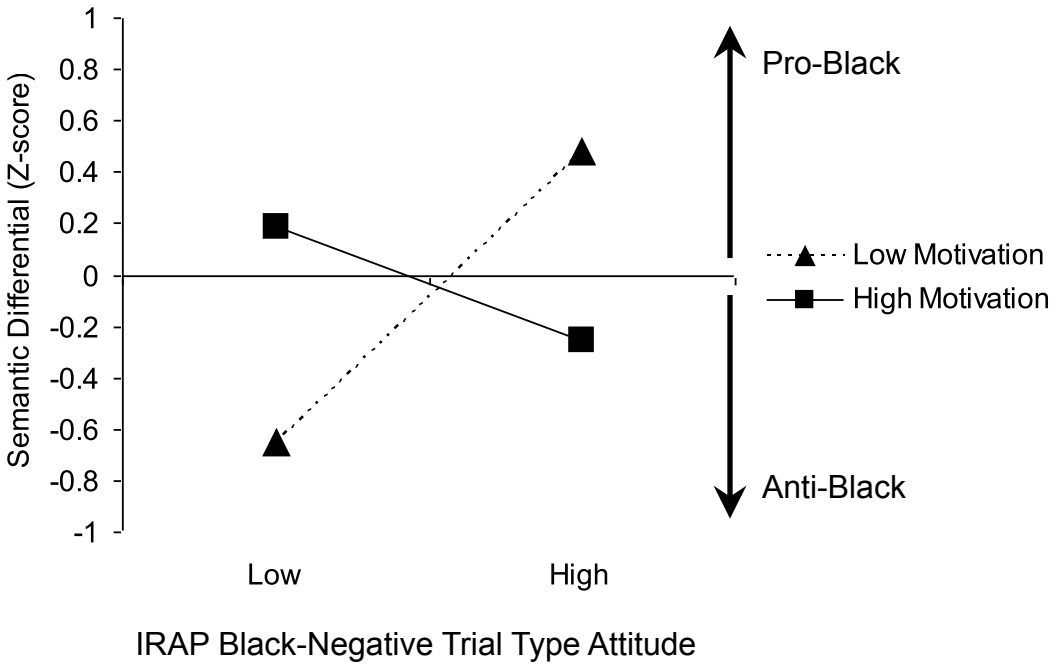


Figure 6. Regression interaction plot with lines relating *Black-Negative* IRAP performance and semantic differential ratings, displayed at one standard deviation above and below the mean on motivation to conceal prejudice.

Overall D-IRAP regression analyses. The analytic strategy employed with the individual IRAP trial-type scores was also used with the overall *D*-IRAP score. That is, four separate regression analyses were conducted. As noted in the Results section of Experiment 1 overall *D*-IRAP scores are calculated *before* reversing the signs for the two Black trial-types, and thus a positive overall *D* score indicates a pro-white/anti-black bias whereas a negative overall *D* score indicates a pro-black/anti-white bias. To facilitate graphical comparison with the feeling thermometer and semantic differential measures, black ratings were subtracted from white ratings, thus yielding positive scores for pro-white/anti-black and negative scores for pro-black/anti-white.

For each analysis one of the explicit measures was entered as the dependent

variable, and on the first step the overall *D*-IRAP score and motivation were entered as predictors. On the second step, the IRAP x motivation interaction was entered, and one of the four regression analyses yielded a significant two-way interaction effect. The IRAP again interacted with motivation in predicting participants' ratings on the semantic differentials, $b = -.404$, $t = 1.9$, $p < .05$. The second step added a significant increment in variance explained $\Delta R^2 = .13$, $F(3, 35) = .522$, $p = .05$. The interaction plot presented in Figure 7 again indicates that the IRAP performance was strongly related to the semantic differential rating for unmotivated individuals ($b = .3$), but the relationship was weaker and in the opposite direction for motivated participants ($b = -.12$).

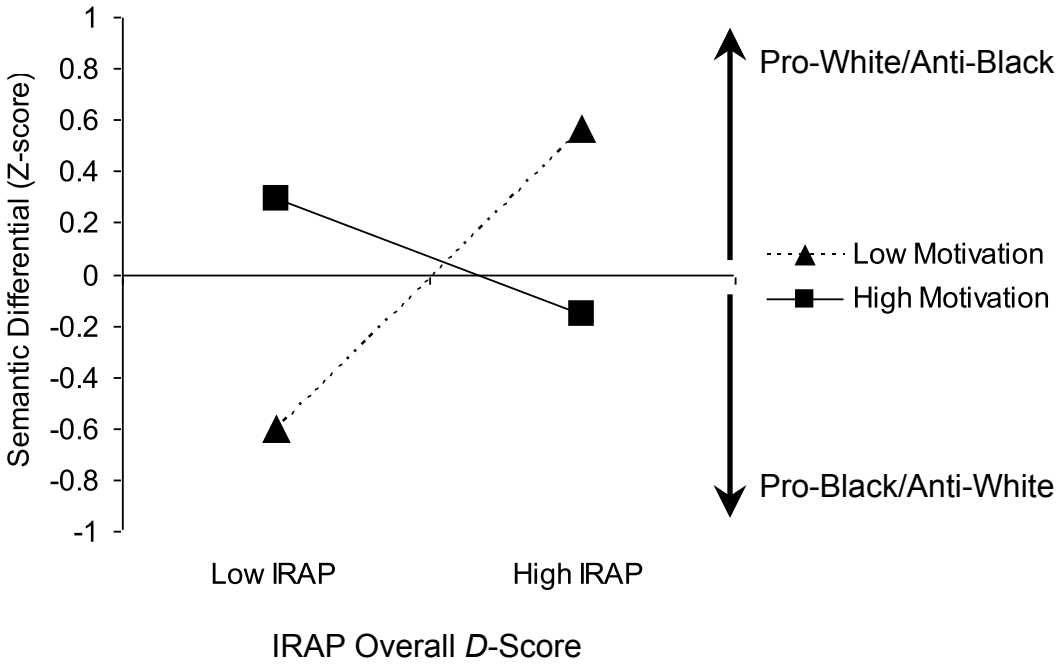


Figure 7. Regression interaction plot with lines relating overall *D*-IRAP performance and semantic differential ratings, displayed at one standard deviation above and below the mean on motivation to conceal prejudice.

Summary and Conclusion

Similar to the previous experiment, the IRAP effects were indicative of pro-white in-group and anti-black out-group bias. Consistent with the post-hoc comparison made between Experiments 2 and 3, the presence-versus-absence of the phrase "I think" had no significant impact on the IRAP effects. Unlike any of the previous experiments, the anti-black bias observed on the *Black-Negative* trial-type proved to be significantly different from zero. Furthermore, in contrast to earlier experiments split-half reliabilities were generally moderate to strong and significant. Overall, these findings suggest that reducing the latency criterion to 2000ms and introducing the "Too Slow!" latency feedback message served to increase the IRAP effects and their internal reliabilities.

Experiments 2 and 3 did not produce significant effects indicative of racial bias for any of the explicit measures, and this result was observed again in the current experiment for the DS and DV scales, and for the semantic differentials, for both types of IRAP ("I think" versus No "I think"). For the feeling thermometers, however, a significant pro-white/anti-black bias was observed but only for those participants who had previously completed the "I think" IRAP. The opposite pattern was observed for the No "I think" IRAP participants, but the effects were non-significant. This result is difficult to interpret at the current time, but it does suggest that although the two different samples did not impact on the IRAP performance itself they may have influenced subsequent ratings on the feeling thermometers. Future research may explore this issue further.

Consistent with Experiment 2, none of the forty implicit-explicit correlations were significant. As noted previously, however, Experiment 3 yielded three significant

correlations involving the *Black-Negative* trial-type, and it was suggested that motivation to conceal prejudice may play an important moderating role in this respect. Evidence to support this conclusion was obtained in the current experiment in which overall IRAP and *Black-Negative* trial-type performances were found to be strongly related to an explicit measure for unmotivated but not motivated individuals.

At this point in the research programme an IRAP had been developed that showed relatively strong in-group and out-group racial bias, reasonable levels of internal reliability, and did not appear to be influenced greatly by increased personalization (i.e., using the phrase "I think"). Furthermore, participants who were motivated to conceal prejudice tended to under-report their racial prejudice on the semantic differentials, but did not perform differently on the IRAP. This pattern of findings is generally consistent with the conclusion that the IRAP possesses at least some of the properties of an implicit measure. Nevertheless, an important test of the validity of the current IRAP would involve conducting a known-groups analysis (De Houwer, & De Bruycker, 2007, Barnes-Holmes, Murtagh, et al., 2010, Barnes-Holmes, Waldron, et al., 2009). Specifically, would performance on the current IRAP differ significantly between groups of black and white participants resident in Ireland? Insofar as the IRAP is a valid measure of racial bias one might indeed expect clear differences. Testing this prediction was the primary purpose of the next experiment.

CHAPTER 6

EXPERIMENT 5

CHAPTER 6

EXPERIMENT 5

The primary purpose of the current experiment was to test the prediction that performance on the IRAP will differ significantly between groups of black and white participants resident in Ireland. Although previous research using implicit measures has shown that white participants tend to show a relatively strong in-group pro-white bias (e.g., Nosek, et al., 2002; Pena, Sidanius, & Sawyer, 2004), the opposite pattern has not been observed for black participants. Rather, black participants tend to show a relatively weak out-group, pro-white bias (Nosek, et al., 2002; Pena, et al. 2004). Nevertheless, the difference in positive bias towards whites does differ significantly between the groups, with whites showing a stronger bias than blacks. Thus, insofar as the IRAP is functionally similar to other implicit measures, one might predict that both white and black participants will produce evidence of pro-white bias on the IRAP, although the white bias will be significantly stronger than the black bias.

On balance, Experiment 5 will be the first study of implicit attitudes to be conducted with black participants in Ireland, and thus a precise prediction is difficult. Ireland has a very short history of significant black immigration with past censuses showing, for example, that the number of black African nationals living in Ireland increased almost ten-fold from 4,867 in 1996 to 42,764 in 2006 (http://www.cso.ie/census/default.htm). As such, Ireland presents an unusual social and cultural context, relative to countries in which black minorities have resided for decades if not centuries. Furthermore, many black residents in Ireland came seeking asylum from various forms of persecution in

their indigenous countries, and thus may not be directly comparable to previous samples of black participants employed in non-Irish studies of implicit racial bias. Given this rather unusual historical context there are insufficient grounds on which to predict that black Irish residents will respond with the pro-white bias observed in previous studies. It does seem reasonable, however, to predict that black participants will respond differently from white participants on at least some of the four IRAP trial-types. Experiment 5 tested this prediction.

Method

Participants

Considerable difficulty was encountered in recruiting black Irish residents as participants for the study, largely due to language difficulties. Eventually, a sample of twenty-two black individuals attending adult education classes in an inner-city school agreed to participate. Sixteen of these participants aged 17 to 26 years (M = 22), completed the experiment individually. All participants were born in Nigeria but had been resident in Ireland for at least 5 years. All participants were experimentally naïve. Twenty-two individuals commenced the experiment, but the data from six participants were excluded because they failed to achieve or maintain performance criteria on the IRAP (described below). The data from the eighteen white Irish participants who completed the "I think" condition of Experiment 4 were used to compare with the data from the black participants.

Materials and Apparatus

The apparatus and materials used in Experiment 5 were similar to those used in

Experiment 4 (IRAP, DS and DV scales, Semantic differential scales, Feeling thermometers). However, the IMS and EMS scales were not employed because questions pertaining to motivation to conceal prejudice among a black minority living in Ireland were not deemed relevant, or even meaningful, in the context of the current study. The two "I think" IRAP sample statements were employed for all participants in Experiment 5.

Procedure

The procedure in Experiment 5 was similar to that used for Experiment 4. Six participants failed to reach the practice criteria (i.e., >= 80% correct and a median response latency <= 2000ms), and thus did not proceed to the test blocks.

Results and Discussion

Implicit Measure

Data preparation. The response latencies were subjected to the same data preparation procedures as were employed for Experiments 1 through 4.

Trial-type analyses. The *D*-IRAP scores for the four trial-types for black participants are presented in Figure 8, along with the data from the eighteen white participants who completed the identical "I think" condition of Experiment 4. The black participants showed positive bias across all four trial-types. The white participants also showed positive bias across the two white trial-types and the *Black-Positive* trial-type but relatively strong negative bias for the *Black-Negative* trial-type; the positive bias for the *Black-Positive* trial-type was relatively weak compared to the bias observed for the black participants.

A mixed repeated measures 2 x 4 ANOVA was conducted on the *D*-IRAP scores,

with race of participant as the between-participant variable and trial-type as the within-participant variable. There was a significant main effect for trial-type, $F(3, 32) = 6.31$, $p < .0006$, $\eta_p^2 = .16$, and for race of participant $F(1, 32) = 11.9$, $p < .001$, $\eta_p^2 = .27$, and a significant interaction, $F(3, 32) = 7.65$ $p < .0001$, $\eta_p^2 = .19$. Between-group post-hoc analyses (Fisher's PLSD) revealed significant differences between black and white participants' performances on the two black trial-types (*ps* < .02), but not on the white trial-types (*ps* > .2).

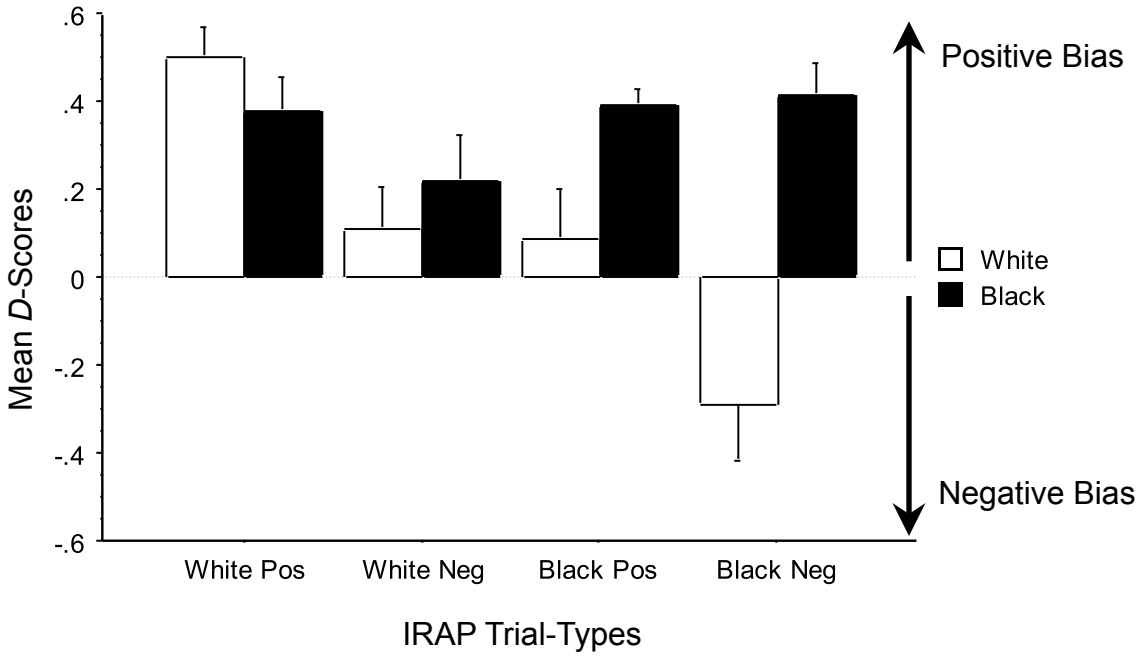


Figure 8. The mean *D*-IRAP scores, with standard error bars, for the four trial-types for Black and White participants.

A within-group post-hoc ANOVA for the black participants revealed no significant trial-type effect ($p > .1$), but an ANOVA for the white participants was significant, $F(3, 17) = 9.92$, $p < .0001$, $\eta_p^2 = .36$. Fisher's PLSD post-hoc analyses for the latter ANOVA

indicated that the *White-Positive* trial-type produced significantly stronger positive bias than the other three trial-types (*ps* < .009); furthermore, the *Black-Negative* trial-type was significantly different, and in a negative direction, compared to the *White-Negative* and *Black-Positive* trial-types (*p* < .01).

Eight one-sample *t*-tests indicated that three of the four trial-type effects for the black participants were significantly different from zero (*ps* < .001); the *White-Negative* effect approached significance (*p* = .06). For the white participants, the *White-Positive* effect was significant (*p* <.0001), as was the *Black-Negative* effect (*p* <.03), but the remaining two effects were not (*ps* > .2).

In summary, the black participants showed positive racial bias for both the in- and out-groups that did not differ significantly across trial-types. The white participants also showed positive bias towards white people, but in stark contrast to the black participants they showed relatively weak positive or strongly negative bias towards black people.

Split-half reliability. To assess the internal consistency of the IRAP, five split-half reliability scores were calculated for white participants (in the same way as for the previous Experiments) and these were weak and non-significant for *White-Positive*, *r* = . 362, *n* = 18, *p* = .3; weak, negative and non-significant for *White-Negative*, *r* = -.2, *n* = 18, *p* = .72; weak to moderate and non-significant for *Black-Positive*, *r* = .525, *n* = 18, *p* = .1; and moderate to strong and significant for *Black-Negative*, *r* = -.786, *n* = 18, *p* = .002. Finally, the overall *D*-IRAP measure for white participants produced a strong and significant split-half correlation, *r* = .803, *n* = 18, *p* < .001. Five split-half reliability scores were also calculated (in the same way) for black participants and these were weak and

non-significant for *White-Positive*, $r = .007$, $n = 16$, $p = .99$; moderate to strong and significant for *White-Negative*, $r = .795$, $n = 16$, $p = .004$; weak, negative and non-significant for *Black-Positive*, $r = -.577$, $n = 16$, $p = .4$; and *Black-Negative*, $r = -.593$, $n = 16$, $p = .4$. The overall *D*-IRAP measure for black participants produced a weak and non-significant split-half correlation, $r = -.331$, $n = 16$, $p = .6$. In summary, the *D*-IRAP measures showed good internal reliability for white participants for *Black-Negative* and Overall *D*-scores, and for black participants for the *White-Negative* trial-type.

Explicit Measures

Discrimination and diversity scales. The overall means for the DS scales revealed a difference between white participants ($M = 3.76$, $SD = .67$), and black participants ($M = 3.31$, $SD = .27$), with both revealing positive racial bias (i.e., mean scores above 3); a one-way ANOVA indicated that white participants responses were significantly more positive than those of black participants, $F(1, 32) = 6.129$, $p < .01$, $\eta_p^2 = .16$. The overall means for the DV scales again revealed positive bias for both white ($M = 3.46$, $SD = .73$), and black participants ($M = 3.73$, $SD = .8$). A one-way ANOVA was non-significant ($p > .3$).

Semantic differential scales. Four overall means were calculated for the semantic differential scales (white participants/Black, $M = .8$, $SD = 1.14$; black participants/Black, $M = 1.6$, $SD = .6$; white participants/White, $M = .87$, $SD = 1.14$; black participants/White, $M = 1.23$, $SD = .55$), with all four means revealing a positive bias (> 0). A 2x2 mixed repeated measures ANOVA yielded a significant main effect for race of participant $F(1, 32) = 4.32$, $p < .04$, $\eta_p^2 = .12$, but no other main or interaction effects ($ps > .09$). Follow up tests revealed that black participants rated black people more positively than white

participants rated black people, $F(1, 32) = 6.768$, $p < .01$, $\eta_p^2 = .17$; although black participants rated white people more positively than white participants rated white people, this difference was not significant ($p > .26$).

Feeling thermometers. Four overall means were calculated, showing more positive means for White relative to Black ratings for white participants (Black, $M = 66.5$, $SD = 19.8$; White, $M = 74.3$, $SD = 21.3$), but the opposite pattern for black participants (Black, $M = 77.5$, $SD = 12.4$; White, $M = 74.4$, $SD = 14.1$). A 2x2 mixed repeated measures ANOVA yielded no significant main effects ($ps > .1$), but a significant interaction, $F(1, 32) = 13.125$, $p < .001$, $\eta_p^2 = .29$. Two between-participant follow-up ANOVAs yielded one effect that approached significance; black participants rated black people more positively than white participants rated black people, $F(1, 32) = 3.620$, $p < .07$, $\eta_p^2 = .1$; the rating of white people by black and white participants did not differ significantly ($p > .9$). Two within-participant follow-up ANOVAs indicated that white participants rated white people significantly more positively than they rated black people $F(1, 17) = 9.686$, $p < .006$, $\eta_p^2 = .36$, and black participants rated black people more positively than they rated white people, but only at a level that approached significance, $F(1, 15) = 4.310$, $p < .06$, $\eta_p^2 = .2$.

Implicit-Explicit Correlations

A correlation matrix of the implicit and explicit measures was calculated across black and white participants. This involved correlating the four trial-type and overall *D*-IRAP scores with each of the six explicit measures. Out of the 30 correlations, six were significant and two approached significance (all other $ps > .1$): *White-Positive* trial-type with black feeling thermometer ($r = -.35$, $p < .04$); *Black-Positive* trial-type with both

black semantic differential (r = .39, p < .02) and black feeling thermometer (r = .411, p < .01); *Black-Negative* trial-type with both black semantic differential (r = .336, p < .05) and black feeling thermometer (r = .343, p < .04); overall *D*-IRAP score with black feeling thermometer (r = -.38, p < .02); *White-Negative* trial-type with diversity scale (r = -.33, p = .06); and overall *D*-IRAP with black semantic differential (r = -.32, p = .07). For each of the eight correlations the IRAP effect was consistent with the explicit measure. For example, increased pro-white bias on the *White-Positive* trial type predicted lower ratings on the black feeling thermometer, whereas increased pro-black bias on the *Black-Positive* trial-type predicted higher ratings on this thermometer. Note also that a negative overall *D*-IRAP score indicates a pro-black/anti-white bias, and thus the negative correlation with the black semantic differential is consistent with the other correlations.

Predictive Validity

A series of hierarchical logistic regression analyses were conducted to determine if one or more of the IRAP measures increased the predictive validity each of the six explicit measures. For illustrative purposes, consider the first regression analysis reported in Table 1. The DS was entered as a predictor or race (i.e., white or black participant) in the first step of the model, and this proved to be weak but significant, B = 1.82, p = .03, accounting for 13% of the variance. The *White-Positive D*-IRAP scores were entered in the second step of the model, and this produced virtually no increment in predictive validity, B = 1.35, p = .32, accounting for 15% of the variance (R^2 change = .02). A further four separate models were then created in which the Discrimination Scale was entered as the first step and the remaining IRAP measures were entered as second steps. The *Black-Positive,*

Black-Negative, and overall *D*-IRAP measures significantly increased the predictive validity of the DS, with the *Black-Negative* measure yielding the largest increment (R^2 change = .41). The same general strategy was then applied to the remaining five explicit measures (see Table 1), and a similar pattern of results was obtained for these except that the *Black-Positive* measure did not significantly increase predictive validity for the black semantic differential and black feeling thermometer.

In short, the Black-Negative and Overall D-IRAP measures each significantly increased the predictive validity of each of the six explicit measures. The Black-Negative measure in particular produced large increases in the percentage of variance accounted for, adding between 36 to 44 percent to the explicit measures.

Table 1

Summary of Hierarchical Logistical Regression analysis for the variables predicting race of participants (N = 34).

Step 1 Explicit Measure				Step 2 Explicit + Implicit Measures				
Predictor Variables	*B*	R^2	*p*	Predictor Variables	*B*	R^2	*p*	R^2 *Change*
Discrimination Scale	1.82	.13	.03*	Discrimination Scale +				
				White-Pos *D*-IRAP	1.35	.15	.32	.02
				White-Neg *D*-IRAP	0.07	.13	.94	0
				Black-Pos *D*-IRAP	3.09	.27	.03*	.14
				Black-Neg *D*-IRAP	7.53	.54	.02*	.41
				Overall *D*-IRAP	6.25	.34	.03*	.21
Diversity Scale	0.49	.02	.30	Diversity Scale +				
				White-Pos *D*-IRAP	1.48	.06	.25	.04
				White-Neg *D*-IRAP	0.78	.04	.39	.02
				Black-Pos *D*-IRAP	2.67	.15	.04*	.13
				Black-Neg *D*-IRAP	6.50	.45	.02*	.43
				Overall *D*-IRAP	4.57	.19	.02*	.17
Semantic Differential (SD) Black	0.97	.14	.02*	SD Black +				
				White-Pos *D*-IRAP	1.18	.15	.38	.01
				White-Neg *D*-IRAP	0.64	.14	.52	0
				Black-Pos *D*-IRAP	2.04	.20	.13	.06
				Black-Neg *D*-IRAP	7.42	.50	.02*	.36
				Overall *D*-IRAP	4.35	.26	.04*	.12
Semantic Differential (SD) White	0.47	.03	.26	SD White +				
				White-Pos *D*-IRAP	1.37	.05	.30	.02
				White-Neg *D*-IRAP	0.59	.04	.53	.01
				Black-Pos *D*-IRAP	2.65	.15	.05*	.12
				Black-Neg *D*-IRAP	6.56	.45	.02*	.42
				Overall *D*-IRAP	4.97	.21	.02*	.18
Feeling Thermometer (FT) Black	0.04	.08	.08	FT Black +				
				White-Pos *D*-IRAP	1.91	.09	.50	.01
				White-Neg *D*-IRAP	0.69	.09	.49	.01
				Black-Pos *D*-IRAP	2.22	.15	.10	.07
				Black-Neg *D*-IRAP	6.69	.45	.02*	.37
				Overall *D*-IRAP	4.16	.21	.04*	.13
Feeling Thermometer (FT) White	0.00	.00	.99	FT White +				
				White-Pos *D*-IRAP	1.65	.04	.22	.04
				White-Neg *D*-IRAP	0.68	.01	.44	.01
				Black-Pos *D*-IRAP	2.72	.13	.04*	.13
				Black-Neg *D*-IRAP	6.63	.44	.01*	.44
				Overall *D*-IRAP	4.39	.17	.02*	.17

*$p < .05$

Discriminant Analysis

A series of discriminant analyses were performed in order to determine the extent to which each of the IRAP and explicit measures predicted whether a participant was black or white. For illustrative purposes, consider the first discriminant analysis reported in Table 2. The value of the discriminant function for the White-Positive IRAP measure was not significantly different for black and white participants, $\chi 2$ (1, 32) = 1.41, p = .23, with the overall function successfully predicting outcome for 67.6% of cases, with accurate predictions being made for 62.5% of the black group, and 72.2% of the white group. This indicated a 37.5% false negative misclassification of the black group, and a 27.8% false positive classification of the white group. The remaining discriminant analyses indicated that three of the IRAP measures (Black-Positive, Black-Negative, and Overall D-IRAP) and two of the explicit measures (DS and black semantic differential) were significant predictors (the black feeling thermometer approached significance). The best predictor of group status was the Black-Negative IRAP measure, predicting outcome for 82.4 percent of cases.

Table 2

Summary of Discriminant Analyses for the Variables Predicting Race of Participants ($N = 34$).

Measure	χ^2	*df*	*p*	Race	Predicted Percentage of Group Membership		Overall Percentage Correct Classification
					Black	White	
White-Pos *D*-IRAP	1.41	1, 32	.23	Black White	62.5 27.8	37.5 72.2	67.6
White-Neg *D*-IRAP	.56	1, 32	.46	Black White	37.5 22.2	62.5 77.8	58.8
Black-Pos *D*-IRAP	5.31	1, 32	.02*	Black White	68.8 33.3	31.3 66.7	67.6
Black-Neg *D*-IRAP	16.38	1, 32	.00*	Black White	93.8 27.8	6.3 72.2	82.4
Overall *D*-IRAP	7.23	1, 32	.01*	Black White	68.8 38.9	31.3 61.1	64.7
Discrimination Scale	5.52	1, 32	.02*	Black White	87.5 38.9	12.5 61.1	73.5
Diversity Scale	1.06	1, 32	.30	Black White	43.8 50.0	56.3 50.0	47.1
Semantic Differential Black	6.04	1, 32	.01*	Black White	68.8 33.3	31.3 66.7	67.6
Semantic Differential White	1.26	1, 32	.26	Black White	75.0 44.4	25.0 55.6	64.7
Feeling Thermometer Black	3.38	1, 32	.07	Black White	56.3 33.3	43.8 66.7	61.8
Feeling Thermometer White	.00	1, 32	.99	Black White	50.0 66.7	50.0 33.3	41.2

Summary and Conclusion

The primary purpose of the current Experiment was to conduct a "known-groups" analysis of the race-IRAP developed across the previous four experiments. Specifically, would performance on the IRAP differ significantly between groups of black and white participants resident in Ireland? The results showed that black participants showed positive racial bias for both the in- and out-groups that did not differ significantly across trial-types. The white participants also showed positive bias towards white people, but in stark contrast to the black participants they showed relatively weak positive or strongly negative bias towards black people. The internal reliability of the *D*-IRAP measures were reasonably robust for white participants for *Black-Negative* and Overall *D*-scores, and for black participants for the *White-Negative* trial-type.

The explicit measures yielded mixed results. The DS indicated that white participants were significantly more pro-black than black participants, but no such difference was obtained on the DV measure. For the semantic differentials, the black participants were more positive about both races than the white participants. The feeling thermometers revealed an interaction between race of participant and in and out-group ratings. Specifically, black participants rated black people more positively than white participants rated black people; ratings of white people were similar across participants. Finally, both groups rated the in-group more positively than the out-group.

The implicit and explicit measures correlated on only 6 of the thirty correlations, with two others approaching significance. In each case, the IRAP effect was consistent with the explicit measure. A series of hierarchical logistic regression analyses indicated

that the Overall *D*-IRAP, and the Black-Negative trial-type in particular, substantively increased the predictive validity of each of the six explicit measures. A series of discriminant analyses indicated that three of the IRAP measures and two of the explicit measures each predicted group status, with the *Black-Negative* IRAP trial-type being the best predictor.

Overall, these findings show that the current IRAP revealed in-group/out-group bias for the white participants but not for the black participants. The explicit measures produced mixed results, but correlated in a limited number of cases with the IRAP data. Finally, one or more of the IRAP measures provided increased predictive validity over the explicit measures, and provided the best prediction of group status.

At this point in the research programme an IRAP had been developed that yielded clear evidence of implicit racial bias among white Irish participants. The very first IRAP study involved collecting both response latency and electroencephalogram (EEG) data (Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, 2008), and the results showed different patterns of EEG activity across blocks of consistent versus inconsistent trials on the IRAP. At the time of writing, EEG data had not been collected in another IRAP study and thus in the next experiment EEGs were recorded while participants were exposed to the race-IRAP that had been developed across the previous experiments.

CHAPTER 7

EXPERIMENT 6

CHAPTER 7

EXPERIMENT 6

Psychophysiological assessment tools have been suggested as viable and useful techniques for the assessment of prejudiced emotional responses. Such tools have the advantages of circumventing limitations of self-report measures (Guglielmi, 1999) and of eliminating possible sources of bias in questionnaires more generally such as modification of responses for reasons of social desirability. Several relevant techniques have pervaded the literature (e.g., functional magnetic resonance imaging [*f*MRI], event-related potentials [ERPs], electromyography [EMG], startle eye-blink responses, and autonomic responses; see Guglielmi, 1999; Ito & Cacioppo, 2007 for reviews). As discussed, the very first IRAP study also measured ERPs, and the results showed different patterns of activity across blocks of consistent and inconsistent IRAP trials.

Experiment 6 sought to extend these findings, and thus recordings were taken from multiple EEG signals, while participants completed the race-IRAP, and these signals were then transformed into event-related potentials (ERPs; e.g., Kutas, 1993; Kutas & Hilliard, 1984). This method of recording neural activity is relatively noninvasive and inexpensive, and allows researchers to investigate the neurophysiological processes underlying functions such as perception, semantic relations, and reasoning (see Barnes-Holmes, Staunton, et al. 2005; Barnes-Holmes, Regan, et al., 2005, for examples of ERP research within the behavior-analytic tradition).

Generating ERP data involves time-locking the EEG signals to a particular series of events and then averaging the signals across trials. The process of averaging allows the

researcher to distinguish the brain's normal background activity from the activity produced by the stimuli presented in the experiment. In effect, each EEG signal for a particular set of stimuli is collated and averaged to produce a single waveform for each site, and then these waveforms are averaged across participants to provide "grand average" waveforms that provide group-based measures of the effect of the targeted stimulus or stimuli.

There is a range of waveforms associated with ERP measures. Some ERPs, for example, are thought to be correlated with specific cognitive processes, such as differentiating different auditory stimuli from one another or understanding words. These ERPs commonly occur at around 300 or 400ms after stimulus onset. The use of ERP measures with the race-IRAP in the current study was entirely exploratory, and thus no specific predictions were made pertaining to the ERP waveforms that might emerge. One ERP measure however that seemed particularly pertinent to the IRAP is the N400, a late negative waveform (see Holcomb & Anderson, 1993; Kounios & Holcomb, 1992). The N400 is usually produced when participants are required to respond to stimuli that are unexpected, unrelated, or wrongly paired in some sense (known as low *cloze-probability)*. Presenting pairs of words that are semantically unrelated, for example, tends to produce an N400, whilst words from the same semantic categories do not. Insofar as pro-black/anti-white trials on the race-IRAP require "incorrect" or "wrongly paired" responses, a more negative waveform may emerge for these trials relative to pro-white/ anti-black trials. Indeed, this is the general pattern of results obtained in the only study that has measured EEG signals while participants completed an IRAP (Barnes-Holmes, et al., 2008). On balance, the previous study was conducted using verbal relations that would not be

deemed socially sensitive (e.g., Pleasant – Holiday – Similar) and a practice latency criterion of 3000ms was applied. Given that the current study will employ socially sensitive verbal relations (e.g., Black – Stupid – True) and a 2000ms response latency criterion, it is quite possible that different EEG results will emerge.

In Experiment 6, separate ERP waveforms, recorded across a range of sites, for blocks of pro-white/anti-black IRAP trials were collected. Similarly, waveforms were also collected for blocks of anti-white/pro-black trials. A comparison could thus be made between the ERP waveforms associated with these two types of IRAP trials.

Method

Participants

Sixteen participants, 8 male and 8 female, agreed to participate. Ages ranged from 18 to 33 years. Data from seven participants were excluded due to excessive noise in the EEG data (explained below). Participants were given a local record-store voucher worth 10 euros upon completing the study.

Apparatus and materials

The entire experiment was conducted in an electrically shielded room in the human neuroscience laboratory in the Department of Psychology at NUI, Maynooth. The stimuli and materials used with the race-IRAP were identical to those of Experiment 5. To record EEG signals during the IRAP task, a Brain Amp, magnetic resonance (MR) compatible (Class IIa, Type BF) with approved control software (Brain Vision Recorder 1.0), and electrode cap (BrainCap/ BrainCap MR) were used. Two Dell personal computers (Pentium 4) were employed for the experiment. One computer controlled the Brain Amp,

and a second the IRAP. The ERPs data were analyzed using approved analysis software (Brain Vision Analyser 1.0). Hardware and software were manufactured and supplied by Brain Products GmbH, Munich, Germany.

Procedure

The IRAP was identical to that of Experiment 5. Participants were first attached to the Brain Amp and were then exposed to the entire IRAP. Each session, consisting of electrode placement and then the IRAP task, lasted on average 1 hr and 15 mins. Only the ERPs data from the six test blocks were analyzed. Evoked potentials were recorded and analyzed from 32 sintered AG/AG-CI scalp electrodes positioned according to the international 10-20 system. The 32 sites chosen for recording were Fp1, Fp2, F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, C4, T8, TP9, TP7, CP3, CPz, CP4, TP8, TP10, P7 P3, Pz, P4, P8, O1, Oz and O2. The central vertex electrode was used as reference and the FPz as ground. Amplifier resolution was 0.1 μV (range, ±3.2768 mV) and the bandwidth set between 0.5 and 62.5 Hz, with a sampling rate of 250 Hz. The notch filter was set at 50 Hz. All electrode impedances were at or below 5 kΩ. The EEG was collected continuously and edited off-line.

Results and Discussion

Implicit Measure

Data preparation. The response latencies were subjected to the same data preparation procedures as were employed for Experiment 5.

Main analyses. The overall mean *D*-IRAP scores for the four trial-types are presented in Figure 9. The results showed positive bias for the two white trial-types, and

negative bias for the two black trial-types. A one-way repeated measures ANOVA revealed a significant main effect for trial-type, $F(3, 8) = 88.906$, $p < .0001$, $\eta_p^2 = .92$. Fisher's PLSD post-hoc analyses indicated that the two white trial-types produced significantly stronger positive bias than the two black trial-types (*ps* < .0001). The two white trial types did not differ significantly from each other (*ps* > .9), and neither did the two black trial-types (*ps* > .4). One-sample *t*-tests indicated that each of the four trial-type effects differed significantly from zero (all *ps* < .0006).

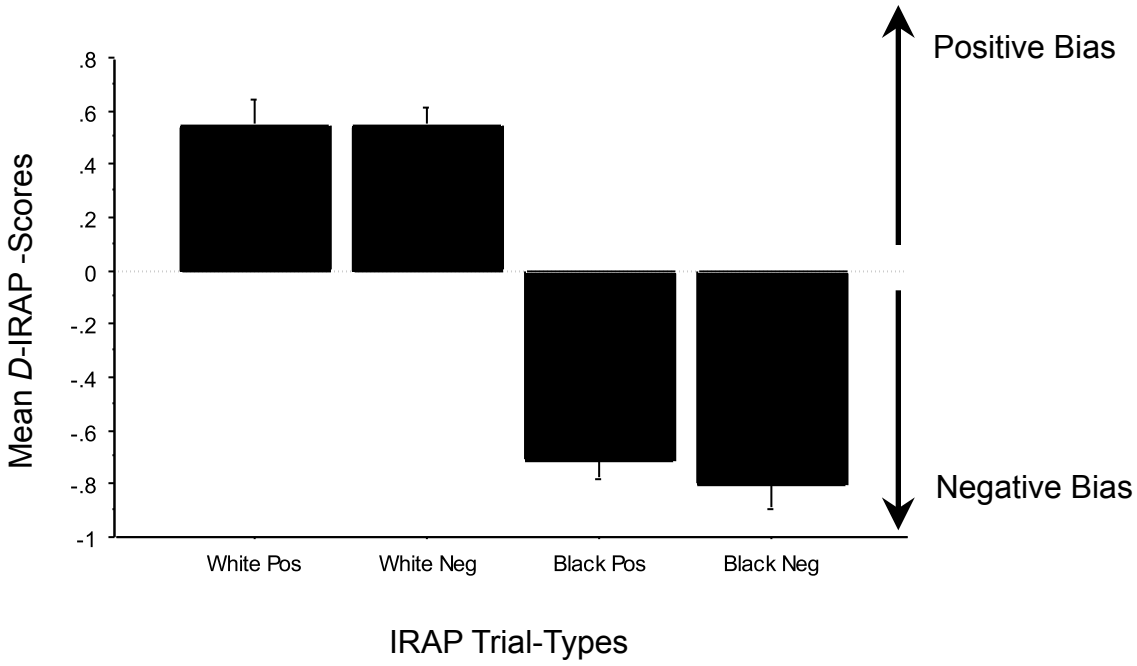


Figure 9. The mean *D*-IRAP scores, with standard error bars, for the four trial-types.

Split-half reliability. To assess the internal consistency of the IRAP, five split-half reliability scores were calculated (in the same way as for the previous Experiments). Three of the reliability estimates (for *White-Positive, White-Negative* and *Black-Positive*) were

negative, thus violating reliability model assumptions. The split-half correlation for the *Black-Negative* trial-type, however, was moderate (r = .613) but not significant. The overall *D*-IRAP measure was weak (r = .164) and also non-significant.

Summary and Conclusion. Similar to previous experiments, the IRAP effects were indicative of pro-white in-group and anti-black out-group bias. The split-half reliability measures were all non-significant, but once again internal reliability was strongest for the *Black-Negative* trial-type. It is also worth noting that this was the first experiment to record a significant anti-black effect on the *Black-Positive* trial-type.

ERPs Data

The continuous EEG signals for each of 16 participants were filtered (0.53 Hz, time constant = 0.3 s, 24 dB/octave roll-off) and then segmented. The segments were divided into 900ms epochs commencing 100ms before onset of the stimuli on each trial (overlapping segments were removed). Vertical and horizontal ocular artifacts were then corrected, and any segments on which EEG or electro-ocular activity exceeded ±75 μV were rejected (the data from 7 participants were removed from subsequent analyses because no segments were artifact free). The remaining segments were then baseline corrected (using the 100ms pre-stimulus interval). Finally, to reduce noise for the ERPs analyses, the data for the three pro-white/anti-black test blocks were collapsed, as were the data for the three pro-black/anti-white test blocks (for ease of communication, these two types of test block will be referred to as pro-white and pro-black, respectively).

The grand average waveforms for each of the 6 frontal electrode sites (Fp1, Fp2, F7, F3, F4, and F8) for pro-white (light lines) versus pro-black (dark lines) blocks are

presented in Figure 10. No differences in evoked potentials between pro-white and pro-black trials were detectable at any of the other sites and thus, in accordance with common practice (e.g., Weisbrod, Keifer, Winkler, Maier, Hill, Roesch-Ely et al., 1999) these data are not reported. Visual inspection of the waveforms from the six sites indicated little evidence of differential activity between the pro-white and pro-black blocks until approximately 200ms after stimulus onset. Thereafter, the two waveforms separated with the pro-black blocks producing greater positivity than the pro-white blocks. The waveforms for sites F3 and F4 tended to converge again around 500ms, whereas the waveforms for the remained sites did not.

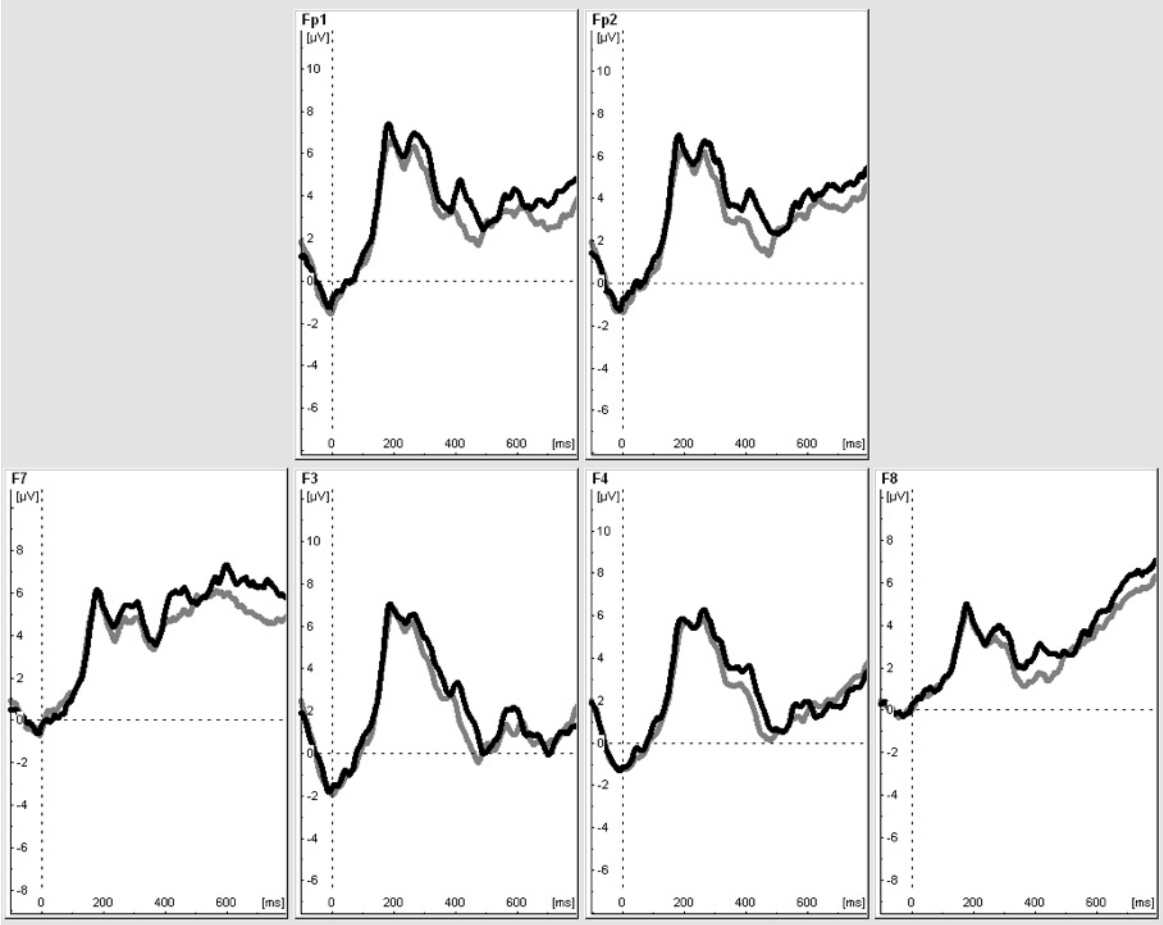


Figure 10. The grand average waveforms for each of the 6 frontal electrode sites (Fp1, Fp2, F7, F3, F4, and F8) for pro-white (light lines) versus pro-black (dark lines) blocks.

The area dimensions (μV x ms) for each ERP waveform (in the temporal region 300–800ms) for each participant were calculated, yielding either positive or negative values with respect to the 0 μV level. For the purposes of statistical analysis average area dimensions were calculated across the three left sites (Fp1, F7, F3) and across the three right sites (Fp2, F8, F4) for pro-white and pro-black waveforms.

The data were entered into a 2x2 repeated measures ANOVA with laterality (left versus right) and IRAP (pro-white versus pro-black) as variables. The main effect for laterality proved to be significant, F (1, 8) = 7.37, p = .03, η_p^2 = .48, as did the effect for the IRAP, F (1, 8) = 7.48, p = .02, η_p^2 = 48; the interaction was non-significant (p > .6). Follow-up paired t-tests for each of the six sites revealed significant differences between pro-white and pro-black waveforms at Fp1, Fp2, F7, and F8 (all ps < .03).

Summary and Conclusions

Experiment 6 replicated the IRAP effects obtained in previous experiments revealing pro-white in-group and anti-black out-group bias. However, this was the first experiment to record a significant anti-black effect on the *Black-Positive* trial-type. Similar to previous experiments internal reliability was strongest for the Black-Negative trial-type. At the present time, it is unclear why an anti-black effect was observed for the Black-Positive trial-type, but given the relatively low n compared to previous experiments, it would be unwise to conclude that this effect that would be replicated reliably.

The EEG recordings revealed that the ERP grandaverage waveforms for the pro-black trials were more positive than for the pro-white trials across six of the frontal sites between 300-800ms. Insofar as pro-black responding for white participants is considered

history-inconsistent and pro-white responding history-consistent, the current experiment produced completely opposite effects to those reported in the only other IRAP study that employed EEG as a measure (Barnes-Holmes, et al., 2008). Specifically, waveforms associated with relational responding that was deemed inconsistent with the participants' prior history were more negative than those waveforms associated with history-consistent responding. In addition, the previous study also reported significant differences between the waveforms for sites in the central and parietal areas; these were not observed in the current experiment.

At the present time, it remains unclear why these differences emerged in the EEG measures across the two studies. As noted earlier, however, the previous study employed stimuli that were not deemed socially sensitive, and used a response-latency criterion of 3000ms (rather than 2000ms.). Furthermore, participants in the earlier study were not required to remain within the latency criterion during the test blocks (this was required in the current study). Clearly, therefore, further research will be required to determine the variables responsible for the different ERP patterns observed across the two studies. Nevertheless, the current findings do indicate that EEG signals may be used to differentiate between two different types of IRAP trial, even when socially-sensitive stimuli are employed.

Thus far, the research reported in the current thesis has involved developing an IRAP that may be used to measure implicit racial bias. Both response latencies and EEG patterns have been shown to be sensitive dependent measures. However, a critical issue in the area of racial prejudice concerns the development of methods that may be used to

modify or undermine such prejudice. The final study reported in the current thesis focused on this issue.

CHAPTER 8

EXPERIMENT 7

CHAPTER 8

EXPERIMENT 7

The development of methods that may be used to modify or undermine prejudice remains a critical issue. However, attempts to modify biases uncovered by other so-called implicit measures (e.g., the IAT) have yielded disappointing results. For example, in studies which employed interventions including an education and empathy manipulation and counter-conditioning procedures neither Gapinksi, Schwartz and Brownell (2006) nor Teachman, Gapinski, Brownell, Rawlins and Jeyaram, (2003) achieved a successful reduction in bias. A recent promising pilot study by Lillis and Hayes (2007) compared two approaches to reducing racial and ethnic prejudice: a protocol based on acceptance and commitment therapy (ACT) and an education-based protocol drawn from a well-known textbook on the psychology of racial differences. In Experiment 7 we sought to compare the effects of both of these approaches on IRAP and explicit performances.

Implicit attitudes were originally believed to be relatively fixed and uncontrollable and research on prejudice reduction has traditionally focused on changing explicit attitudes (Bargh, 1999; Devine, 1989). Researchers have argued that an awareness of one's bias and motivation to change it are necessary for prejudice reduction to be successful (Allport, 1954; Devine, Monteith, Zuwerink, & Elliot, 1991; Myrdal, 1944). However, there has been a growing increase in research investigating the malleability of implicit attitudes (Bargh, 1999; Blair, 2002; Cullen, et al., 2009). Cullen et al (2009, p.592) discuss a number of recent studies which suggest that implicit attitudes can be influenced by (a) expectancies (Blair & Banaji, 1996), (b) practice or training (Kawakami, Dovidio, Moll,

Hermsen, & Russin, 2000), (c) automatic motives (Moskowitz, Salomon, & Taylor, 2000), and (d) motivation to respond without prejudice (Lepore & Brown, 1997).

Some researchers have begun to study the malleability of implicit attitudes through exemplar training. This involves presenting participants with a series of exemplars which are designed to influence their attitudes toward a specific target (e.g., Dasgupta & Greenwald, 2001; Lowery, et al., 2001). In Experiment 1 of their exemplar study, Dasgupta and Greenwald (2001) presented participants with pictures of either admired Black and disliked White individuals, or disliked Black and admired White individuals. Participants then completed an IAT directly following exemplar exposure and again 24 hours later (without re-exposure to the exemplars). Explicit attitude measures were also administered across the two sessions. Findings indicated that exposure to admired Black and disliked White exemplars significantly weakened implicit pro-White preferences for 24 h, but did not affect explicit attitudes. This basic effect was replicated in a second experiment, but with implicit ageism as the target attitude.

As discussed previously, in a more recent study, Cullen et al. (2009) conducted a partial replication of Experiment 2 from Dasgupta and Greenwald (2001), but using the IRAP rather than the IAT. The study showed that when participants were presented with positive examples of old people and negative examples of young people, implicit negative bias towards old people was significantly reduced. Similar to the Dasgupta and Greenwald study, the explicit measures were largely unaffected.

At the time of writing no published study had demonstrated the malleability of implicit racial bias as measured by the IRAP (cf. Barnes-Holmes, Murphy, et al., 2010).

Furthermore, no published IRAP study had attempted to investigate the impact of educational or other types of interventions designed to undermine racial prejudice. Thus, in the final study of the current research programme two types of intervention were employed to determine if they would reduce the levels of negative implicit racial bias observed across many of the previous experiments.

For the purposes of the current study, the two interventions selected were taken from a recent report that sought to explore the effectiveness of a psychological acceptance-based protocol versus a traditional prejudice awareness education training programme. Given that the current research was exploratory, no specific predictions were made concerning the impact of these two different interventions.

Method

Participants

Twenty four participants, 13 male and 11 female, who had completed Experiment 4 or 5 agreed to participate, and their responses to both the implicit and explicit measures from those experiments were employed as pre-intervention baseline data. Ages ranged from 19 to 32 years. No inducements were offered for participation in the study.

Materials and Apparatus

The apparatus and materials used in Experiment 7 were similar to those used in Experiment 5 (IRAP, DS and DV scales, Semantic differential scales, Feeling thermometers, IMS and EMS scales). Based on work by Brochu and Morrison (2007) participants in Experiment 7 were also presented with Behavioral Intention Questionnaires (BIQs, see Appendix E), which included twelve photographs; six depicting females (three

black and three white), and six depicting males (three black and three white). For each photograph, the participant was required to answer five questions assessing the extent to which they would interact with the pictured person. Each question involved a 7 point rating scale (1 = *very unlikely* to 7 = *very likely)*. Therefore, scores could range from 5 to 35 for each pictured person, with higher scores indicating greater likelihood of interaction with the target. In line with Lillis and Hayes (2009), participants were randomly assigned to either an acceptance or a psycho-education based brief protocol. See Appendix F for the materials employed in each protocol.

Procedure

The IRAP and explicit measures were the same as those employed in Experiment 5, except the BIQ was also employed. There were three phases in the current Experiment. In Phase 1, participants completed the BIQ individually. In Phase 2 participants were randomly assigned to either the acceptance or psycho-education based intervention, which lasted from 30-45 minutes (see below).

Prejudice awareness training (education). The education based protocol was adapted from Lillis and Hayes (2007). The material included is based on a widely used textbook of multicultural psychology by Sue and Sue (2003). Chapter 11, which specifically addresses characteristics of African Americans, was used. Participants were presented with material directly from the textbook outlining characteristics of this minority ethnic group. It emphasized group strengths and common stereotypes and included information about the importance of recognizing and correcting one's own biases, becoming more aware of and open to different cultures and identifying the uniqueness of

each individual. This approach also highlighted the moral aspects of prejudice and the negative impact that behavioral expressions of prejudice have on others (Lillis & Hayes, 2007). As in the Lillis and Hayes (2007) study, participants were required to examine the truth of their own prejudicial thoughts and to consider the ways in which these thoughts may impact on their behavior toward people from minority ethnic groups.

Participants assigned to the Education protocol were presented with detailed information from Chapter 11 in Sue and Sue (2003) pertaining to the discrepancy between the treatment of White and Black people in America. Participants read how African Americans experience nearly three times as much poverty as White Americans and are twice as likely to experience unemployment. The text highlighted the disadvantaged status of African Americans as well as the impact of racism and poverty on their lives and the opportunities they are presented with. They were informed that one third of African American men in their 20s are in jail, on probation, or on parole and that this rate grew by over a third between 1990 and 1995.

Other equally bleak statistics were presented, from the finding that the lifespan of African Americans is five to seven years shorter than that of White Americans to the fact that despite comparative insurance cover, compared to white patients, African American patients are less likely to undergo corrective surgeries or major therapeutic procedures. As part of this protocol, participants completed a number of exercises. The first exercise asked participants to “Please write down your immediate reactions to this information.”

Further exercises asked participants to “Now, please write down what you think it would be like to be a black person living in Ireland” and to “Please write down any ideas

you might have for overcoming current or potential race-related difficulties in the Irish context."

Acceptance and commitment training. The mindfulness and acceptance based protocol was based on Lillis and Hayes (2007) and the current study also aimed to encourage participants to: (a) become mindfully aware of their own prejudicial thoughts, feelings and reactions, (b) accept those thoughts and feelings as the natural result of learning and using language in a prejudicial society, (c) notice the automatic processes of evaluation and judgment more generally, and (d) orient to positive actions consistent with one's own values regarding how to treat other human beings.

Again, as in the Lillis and Hayes (2007) study participants completed a number of exercises. The first exercise asked participants to complete statements in writing, such as "Most Black people tend to . . ." and "Some racial slurs I know are . . . ." Participants were asked to notice the different thoughts that came up while they completed the task. Participants were then directed to complete common phrases such as "Blondes have more . . ." and "There's no place like . . . ." and were then asked to notice how automatic these thoughts were. The next exercise requested students to say and to memorize the numbers 1, 2, 3. The text then asked "What are the numbers?" and attention was drawn to how difficult it was to get "the numbers" out of their heads. Participants were asked to recognize how easy it was for a specific thought to be "put in their head" and to consider how difficult it is to remove a thought.

Other exercises adapted from Lillis and Hayes (2007) challenged participants to question the value in trying to change ones thoughts or feelings. Another exercise asked

participants to imagine themselves in various interpersonal scenarios with people whose race or ethnic identity changed (at work, alone on the street, etc.) and to notice their emotional reactions. Participants were reminded of how easily prejudicial thoughts, attitudes, and feelings can emerge but the text highlighted that these thoughts need not affect behavior. Participants were encouraged to mindfully acknowledge the presence of prejudicial thoughts and feelings without attempting to alter them, and to focus on behaving in a manner consistent with their values (Lillis and Hayes, 2007).

In the third and final phase of the experiment participants completed the IRAP and the explicit measures (including the BIQs), and then were thanked and fully debriefed.

Results

Implicit Measure

Data preparation. The response latencies were subjected to the same data preparation procedures as were employed for Experiments 1 through 6.

Preliminary analyses. Fourteen one-way ANOVAs indicated that there were no significant differences at baseline on either the implicit or the explicit measures between participants randomly assigned to either intervention (*ps* > .17).

Trial-type analyses. The *D*-IRAP scores were entered in a mixed repeated measures 2 x 2 x 4 ANOVA, with intervention (ACT versus Education) as the between-participant variable and trial-type and pre- and post-intervention as the within-participant variables. There was a significant main effect for trial-type, $F(3, 22) = 45.96$, $p < .0001$, $\eta_p^2 = .68$, but not for pre-post or intervention (*ps* > .09). There was, however, a significant interaction between trial-type and pre-post $F(3, 22) = 2.76$, $p = .049$, $\eta_p^2 = .11$. In effect,

the interventions appeared to impact significantly upon the IRAP trial-type effects, but this was not moderated by the type of intervention (i.e., the effects of ACT and Education did not differ significantly). Consequently, the data were collapsed across the interventions and the results are presented in Figure 11.

At baseline, participants showed positive bias across the two white trial-types but negative bias for the two black trial-types. Post intervention, participants showed positive bias across three of the four trial-types; the two white trial-types and the *Black-Positive* trial-type. The negative bias on the *Black-Negative* trial-type was maintained, but was weaker compared to baseline. The *Black-Positive* trial-type whilst slightly negative at baseline became positive post intervention. The positive bias for the *White-Positive* trial-

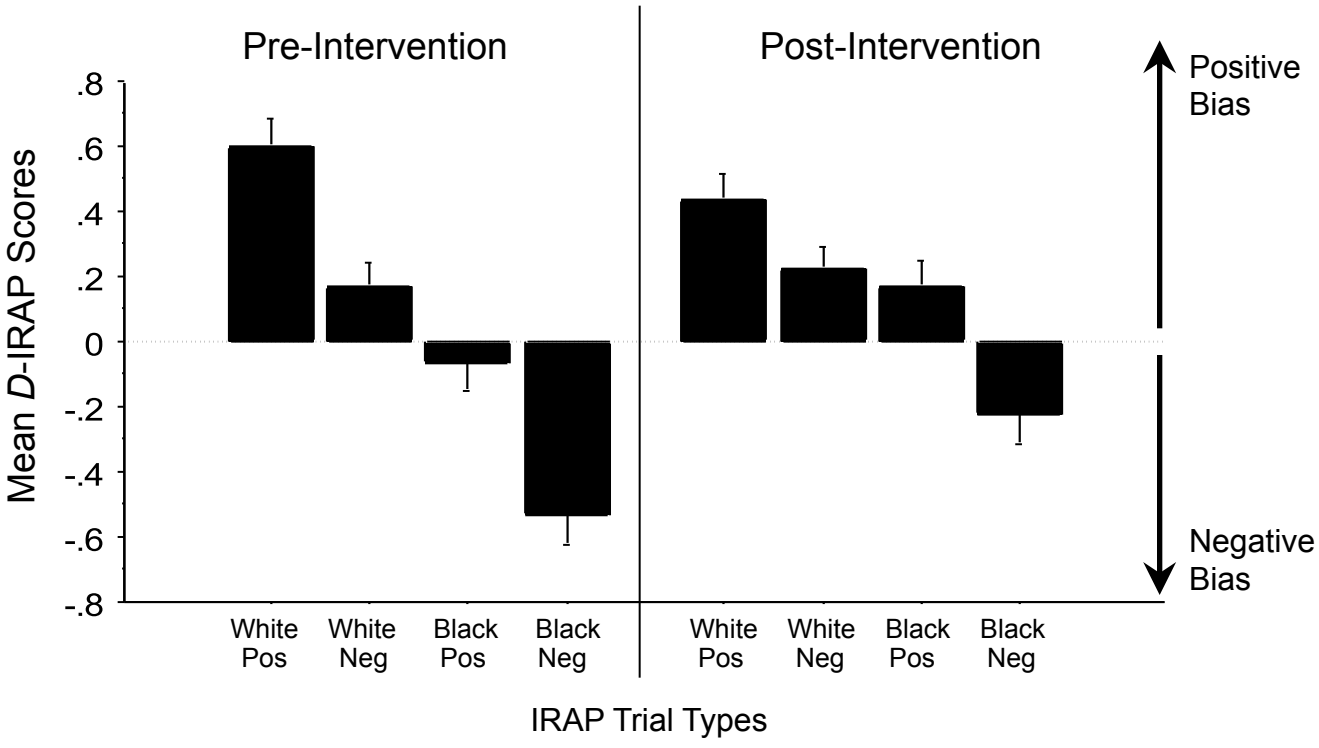


Figure 11. Mean *D*-IRAP trial-type scores, with standard error bars, pre and post intervention.

type was also reduced post intervention compared to baseline. In effect, for three of the trial-types, implicit pro-white and anti-black bias was reduced following the interventions.

To explore the impact of the interventions on each trial-type, four paired *t*-tests were used to determine if the changes from pre to post were significant. The changes for the white trial-type were not (*ps* >.2) whereas the changes for the black trial-types were marginally significant (*ps* < .055). Eight one-sample *t*-tests indicated that three of the four trial-type effects were significantly different from zero at baseline; *White-Positive* (t = 7.252, df = 23, p <.0001), *White-Negative* (t = 2.230, df = 23, p <.036), and *Black-Negative* (t = -5.946, df = 23, p <.0001), but the *Black-Positive* effect was not (p > .4). Post intervention, each of the four trial-type effects were significantly different from zero; *White-Positive* (t = 5.565, df = 23, p <.0001), *White-Negative* (t = 3.144, df = 23, p <.036), *Black-Positive* (t = 2.093, df = 23, p <.047), and *Black-Negative* (t = -2.369, df = 23, p <. 026).

In summary, at baseline participants showed positive racial bias towards white people but showed weak or strongly negative bias towards black people. Post intervention, participants again showed positive racial bias towards white people but also slightly more positive bias toward black people on the *Black-Positive* trial-type, and slightly less negative anti-black bias on the *Black-Negative* trial-type.

Explicit Measures

A total of 10 explicit measures were employed in the current study (DS, DV, SD white, SD black, FT white, FT black, IMS, EMS, BIQ white, BIQ black). The mean ratings and standard deviations for each measure for both pre and post intervention are

presented in Table 3. Each explicit measure was entered into a 2x2 mixed repeated measures ANOVA with pre/post as the repeated measure and intervention as the between-participant variable (note that preliminary analyses had established that there were no significant differences on any of these measures at baseline).

Table 3

Mean ratings and standard deviations for each explicit measure for both pre and post intervention

	ACT- based Intervention				Education-based Intervention			
	Pre		Post		Pre		Post	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Discrimination Scale	3.74	.592	4.08	.38	3.92	.65	3.89	.45
Diversity Scale	3.35	.538	3.60	.55	3.63	.79	3.58	.44
Semantic Differential Black	.94	1.00	1.25	.76	.83	.98	1.1	-.6
Semantic Differential White	.58	1.00	1.1	.74	.79	1.05	1.1	.47
Feeling Thermometer Black	68.33	12.67	80.00	12.06	66.58	18.09	82.50	11.38
Feeling Thermometer White	73.33	14.98	80.83	12.40	69.08	20.52	83.33	9.85
Internal Motivation to Conceal Prejudice	6.03	1.35	7.82	.89	6.32	1.21	7.35	1.41
External Motivation to Conceal Prejudice	3.83	1.91	4.43	1.67	4.87	1.63	4.13	2.1
Behavioural Intention Questionnaire Black	21.74	2.80	22.83	2.97	21.13	2.46	21.75	2.48
Behavioural Intention Questionnaire White	21.94	2.72	22.91	2.75	21.63	2.38	22.37	2.28

Discrimination and diversity scales. No significant differences were obtained from the ANOVAs for these two measures (*ps* > .26).

Semantic differential scales. The ANOVA for the rating of White people yielded a significant main effect for pre/post, $F(1, 22) = 4.72$, $p < .04$, $\eta_p^2 = .18$ with no other significant effects (p > .5). A similar pattern was obtained for the rating of Black people,

but the effect only approached significance $F(1, 22) = 2.98$, $p = .09$, $\eta_p^2 = .12$ (remaining *ps* > .56). In summary, participants in both intervention groups rated both White and Black people more positively on the SD scales post intervention.

Feeling thermometers. Both ANOVAs yielded significant main effects for pre/post (White $F(1, 22) = 9.65$, $p < .005$, $\eta_p^2 = .3$; Black $F(1, 22) = 17.29$, $p < .0004$, $\eta_p^2 = .44$), with no other main or interaction effects (*ps* > .34). Once again, participants in both groups indicated more positive reactions to both White and Black people following the interventions.

Motivation to conceal prejudice scales. The ANOVA for internal motivation yielded a significant main effect for pre/post, $F(1, 22) = 12.19$, $p < .002$, $\eta_p^2 = .36$ (all remaining *ps* for this ANOVA and the ANOVA for external motivation > .26). These results indicate that both intervention groups became more internally motivated to conceal prejudice following the interventions.

Behavioral Intention Questionnaires. The ANOVA for the BIQ for white people yielded a significant main effect for pre-post, $F(1, 22) = 249.84$ $p < .0001$, $\eta_p^2 = .92$ (remaining *ps* > .06), indicating that behavioural intentions to interact with white people increased following both interventions. The ANOVA for the BIQ for black people also yielded a main effect for pre-post, $F(1, 22) = 76.07$, $p < .0001$, $\eta_p^2 = .78$ but an interaction with intervention was also recorded, $F(1, 22) = 5.63$, $p < .03$, $\eta_p^2 = .2$. Two one-way repeated measures ANOVAs indicated that both intervention groups showed significant increases in behavioural intentions towards black people, ACT, $F(1, 11) = 37.48$, $p < .0001$, $\eta_p^2 = .76$; Education, $F(1, 11) = 56.32$, $p < .0001$, $\eta_p^2 = .83$. Thus, both

interventions increased behavioural intentions, but the effect was larger for the Education group.

Implicit-Explicit Correlations

Two correlations matrices of the implicit and explicit measures were calculated; one matrix for pre- and a second for post-intervention. For each matrix the four trial-type and overall *D*-IRAP scores were correlated with each of the ten explicit measures. Out of the 50 correlations for the pre-intervention matrix, two were significant (all other *ps* > .06). Increased external motivation predicted greater positive bias towards white people (on the *White-Positive* trial-type; $r = .455$, $p < .03$), whereas increased internal motivation predicted greater negative bias towards black people (on the *Black-Negative* trial-type; $r = -.424$, $p < .04$). Out of the 50 correlations for the second matrix only one was significant (all other *ps* > .1). Specifically, higher ratings on the Diversity scale predicted greater positive bias towards white people (on the *White-Positive* trial-type; $r = .424$, $p < .04$).

Summary and Conclusions

Both interventions impacted significantly on the *D*-IRAP performances. There was very limited evidence of a reduction in the pro-white bias, but close to significant reductions in the anti-black bias from pre- to post-intervention. Most of the explicit measures showed significant changes from pre- to post-intervention, except for the Discrimination and Diversity scales. Both the semantic differential scales and feeling thermometers showed more positive ratings for both black and white people post-intervention. With respect to the motivation scales, internal motivation increased significantly from pre to post, but external motivation did not. Both interventions also

increased behavioural intentions towards white people. A similar increase was also observed when rating black people, but the effect was larger for the Education group. The implicit-explicit correlations yielded only three significant effects out of a possible 100. Pre-intervention, greater external motivation predicted greater positive bias towards white people, but increased internal motivation predicted greater negative bias towards black people. Post-intervention, higher ratings on the Diversity scale predicted greater positive bias towards white people.

Overall, the findings indicated that both implicit and explicit attitudes changed as a result of the two interventions, with little evidence that one intervention was more effective than the other. Interestingly, the change that occurred for the implicit measures indicated a reduction in negative bias towards black people, with little change in the positive bias recorded for white people. In contrast, the explicit measures generally showed changes in responses to both black and white people. Thus the two types of measure appeared to differ in terms of the changes observed following the two interventions. This differential impact on the IRAP and the explicit measures is consistent with previous research that has documented the independent malleability of implicit and explicit attitudes (Cullen, et al., 2010; Barnes-Holmes, Murphy et al., 2010; Boysen et al., 2006).

CHAPTER 9

GENERAL DISCUSSION

CHAPTER 9

GENERAL DISCUSSION

The aim of the current programme of research was to determine if the IRAP, a recently developed methodology for the assessment of implicit cognition, is a useful measure of implicit racial bias in the Irish context. Over the course of a series of experiments, the research has refined the IRAP and has also examined its relationship with various alternative attitudinal indices, including self-report measures and measures of behavioural intentions. In addition, the fifth experiment, presented in Chapter 6, explored the predictive validity of the IRAP using known-groups, while Experiment 6, presented in Chapter 7, investigated the relationship between neural activity and IRAP responses. In the final experiment, the malleability of IRAP performances, as a result of acceptance and education-based interventions, was investigated.

In this final chapter of the thesis, the major findings of the six empirical investigations conducted will be summarised and the wider implications of the research will be discussed.

Overview of the Findings

Experiment 1 (Chapter 2), the first empirical investigation of the current programme of research, manipulated the assessment context in which the IRAP was completed. The aim of the experiment was to determine if manipulating the private versus public context of the assessment situation would impact upon the IRAP effects in a similar manner to that observed with the IAT in the Boysen, et al. (2006) study (i.e., a reduction in implicit in-group bias in the public relative to the private assessment context). Results

showed, contrary to initial expectations, that the IRAP failed to produce evidence of racial bias and appeared largely insensitive to the Public/Private context manipulation. Although the explicit measures showed some sensitivity to race and context, the effects were not clear cut. Furthermore, only a small number of significant implicit-explicit correlations were obtained and some of the effects appeared contradictory. Overall, therefore, the results of Experiment 1 (particularly with respect to the IRAP), were inconsistent with previous research (see Nosek, et al. 2007, for a review).

The irregularity of these results, however, was consistent with another study that was conducted shortly after the current research programme began. Specifically, Barnes-Holmes, Murphy et al. (2010) reported an implicit pro-white/anti-black bias *but* only when participants were required to respond within 2000ms on each trial of the IRAP (rather than 3000). Consequently, in Experiment 2 (Chapter 3) of the current research programme participants were required to respond within 2 seconds on each trial of the IRAP. In addition, based on the assumption that personalizing an implicit measure may have unintended performance effects (Nosek & Hansen, 2008), a further modification was also made to the procedure by removing the phrase "I think" from the sample stimuli (i.e., only "Black People" and "White People" were presented as labels). Results showed that participants in Experiment 2 did not produce significant effects indicative of racial bias for any of the explicit measures and furthermore, none of the forty implicit-explicit correlations were significant. Participants did however produce IRAP effects indicative of in-group racial bias which were broadly consistent with the results of Barnes-Holmes, Murphy, et al., 2010). Nevertheless, it remained unclear to what extent the different

outcomes for the two experiments were due to the reduced latency criterion or to the removal of "I think" from the sample statements.

In order to address this issue, Experiment 3 (Chapter 4) reintroduced the phrase "I think" into the sample stimuli (i.e., as in Experiment 1), and like Experiment 2, participants in Experiment 3 were required to respond within 2000ms on each trial of the IRAP. Experiment 3 again yielded IRAP effects indicative of in-group racial bias, but failed to produce significant effects indicative of racial bias for any of the explicit measures. Three of the forty implicit-explicit correlations were significant and each of these involved the *Black-Negative* trial-type. A post-hoc comparison of Experiments 2 and 3 suggested that the presence-versus-absence of the phrase "I think" had no significant impact on the IRAP effects.

To further explore the impact of personalising the IRAP, Experiment 4 (Chapter 5) directly targeted the "I think" variable. In addition, in order to reduce attrition rates, a modification to the IRAP software was introduced. Specifically, the warning message "Too Slow!" was presented on any trial (practice or test) whenever a participant did not respond within the pre-set 2000ms latency criterion. Experiment 4 also aimed to investigate the moderating effect that motivation to conceal prejudice has on the implicit – explicit relationship, an issue that has been the focus of previous studies on implicit attitudes (e.g., Payne, Govorun, & Arbuckle, 2006). Results showed that similar to Experiment 3, the IRAP effects were indicative of pro-white in-group and anti-black out-group bias. In addition, consistent with the post-hoc comparison made between Experiments 2 and 3, the presence-versus-absence of the phrase "I think" did not impact

significantly on the IRAP effects.

Critically, in contrast to the previous experiments, the anti-black bias observed on the *Black-Negative* trial-type proved to be significantly different from zero. Furthermore, unlike the earlier experiments, split-half reliabilities were generally moderate to strong and significant. Overall, these findings suggest that reducing the latency criterion to 2000ms and introducing the "Too Slow!" latency feedback message served to increase the IRAP effects and their internal reliabilities. Interestingly, Experiment 4 did produce significant effects indicative of racial bias on one of the explicit measures (the feeling thermometers), but only for those participants who completed the personalised IRAP (i.e., with "I think" labels). These results were in contrast with the earlier reported experiments and were difficult to interpret. They may suggest, however, that although manipulating the "I think" variable did not impact on the IRAP performance it may have influenced subsequent ratings on the feeling thermometers. Finally, a series of regression analyses indicated that participants who were motivated to conceal prejudice tended to under-report their racial prejudice on the semantic differentials, but did not perform differently on the IRAP from those participants who were not motivated in this regard. Overall, the pattern of findings obtained across Experiments 2, 3, and 4 was deemed to be generally consistent with the conclusion that the IRAP possesses at least some of the properties of an implicit measure.

At this point in the research programme an IRAP had been developed that showed relatively strong in-group and out-group racial bias, reasonable levels of internal reliability, and did not appear to be influenced greatly by increased personalization. Furthermore, participants who were motivated to conceal prejudice tended to under-report

their racial prejudice, but did not perform differently on the IRAP. In order to further validate the utility of the race-IRAP developed across the previous four experiments, Experiment 5 (Chapter 6) involved a known-groups analysis. The aim of Experiment 5, therefore, was to test the prediction that performance on the current IRAP would differ significantly between groups of black and white participants resident in Ireland.

The results of Experiment 5 revealed in-group/out-group bias for the white participants but not for the black participants. Specifically, the white participants showed positive bias towards white people, *but* relatively weak positive or strongly negative bias towards black people whereas the black participants showed positive racial bias for both the in- and out-groups that did not differ significantly across trial-types. The internal reliabilities of the *D*-IRAP measures were reasonably robust for white participants for *Black-Negative* and Overall *D*-scores, and for black participants for the *White-Negative* trial-type.

The explicit measures again produced mixed results and only six of the thirty implicit-explicit correlations proved to be significant. A series of hierarchical logistic regression analyses indicated that the Overall *D*-IRAP, and the *Black-Negative* trial-type in particular, significantly increased the predictive validity of each of the six explicit measures. In addition, a series of discriminant analyses indicated that three of the IRAP measures and two of the explicit measures each predicted group status, with the *Black-Negative* IRAP trial-type being the best predictor.

At this point in the research programme an IRAP had been developed that yielded clear evidence of implicit racial bias among white Irish participants, but a lack of such bias

among black participants resident in Ireland. In further testing the IRAP as a measure of implicit racial bias, it was decided to record an additional measure of IRAP performance to that of response latency. The first IRAP study (Barnes-Holmes, et al., 2008) involved collecting both response latency and electroencephalogram (EEG) data and the results showed different patterns of EEG activity across blocks of consistent versus inconsistent trials on the IRAP. At the time of writing, EEG data had not been collected in another IRAP study and thus in the next experiment EEGs were recorded while participants were exposed to the race-IRAP that had been developed thus far. It was also noted that the Barnes-Holmes, et al. study was conducted using verbal relations that would not be deemed socially sensitive (e.g., Pleasant – Holiday – Similar) and a practice latency criterion of 3000ms was applied. Given that the race-IRAP employed socially sensitive verbal relations (e.g., Black – Stupid – True) and a 2000ms response latency criterion, it was unclear whether different EEG results would emerge.

Experiment 6 (Chapter 7), replicated the IRAP effects obtained in previous experiments revealing pro-white in-group and anti-black out-group bias. However, this was the first experiment to record a significant anti-black effect on the *Black-Positive* trial-type. Similar to previous experiments, internal reliability was strongest for the *Black-Negative* trial-type. At the present time, it is unclear why an anti-black effect was observed for the *Black-Positive* trial-type, but given the relatively low *n* compared to previous experiments, it would be unwise to conclude that this effect would be replicated reliably.

The EEG recordings revealed that the ERP grandaverage waveforms for the pro-black trials were more positive than for the pro-white trials across six of the frontal sites

between 300-800ms. Insofar as pro-black responding for white participants is considered history-inconsistent and pro-white responding history-consistent, the current experiment produced completely opposite effects to those reported in the only other IRAP study that employed EEG as a measure (Barnes-Holmes, et al., 2008). Specifically, waveforms associated with relational responding that was deemed inconsistent with the participants' prior history were more negative than those waveforms associated with history-consistent responding. In addition, the previous study also reported significant differences between the waveforms for sites in the central and parietal areas; these were not observed in the current experiment.

At the present time, it remains unclear why these differences emerged in the EEG measures across the two studies. As noted above, however, the previous study employed stimuli that were not deemed socially sensitive, and used a response-latency criterion of 3000ms (rather than 2000ms.). Furthermore, participants in the earlier study were not required to remain within the latency criterion during the test blocks (this was required in the current study). Clearly, therefore, further research will be required to determine the variables responsible for the different ERP patterns observed across the two studies. Nevertheless, the current findings do indicate that EEG signals may be used to differentiate between two different types of IRAP trial, even when socially-sensitive stimuli are employed.

Thus far, the research reported in the current thesis involved developing an IRAP that may be used to measure implicit racial bias. Furthermore, both response latencies and EEG patterns were shown to be sensitive dependent measures. However, a critical issue in

the area of racial prejudice concerns the development of methods that may be used to modify or undermine such prejudice. The final study reported in the current thesis focused on this issue. Experiment 7 (Chapter 8), therefore, sought to compare two approaches to reducing racial and ethnic prejudice: a protocol based on acceptance and commitment therapy (ACT) and an educational based protocol drawn from a textbook on the psychology of racial differences. Both of these approaches were based on the work of Lillis and Hayes (2007).

The findings indicated that both implicit and explicit attitudes changed as a result of the two interventions, with little evidence that one intervention was more effective than the other. Interestingly, the change that occurred for the implicit measures indicated a reduction in negative bias towards black people, with little change in the positive bias recorded for white people. In contrast, the explicit measures generally showed changes in responses to both black and white people. Thus the two types of measure appeared to differ in terms of the changes observed following the two interventions. This differential impact on the IRAP and the explicit measures is consistent with previous research that has documented the independent malleability of implicit and explicit attitudes (Cullen, et al., 2010; Barnes-Holmes, Murphy et al., 2010; Boysen et al., 2006).

Overall, the seven experiments reported in the current thesis lead to the development and refinement of an IRAP that could be used to measure implicit racial bias. Support for the reliability of the measure was provided when the same overall pattern of pro-white and anti-black bias among white participants was observed across all but the first experiment, which employed a 3000ms latency criterion. Measures of internal

reliability were also found to be relatively robust, especially for a response-time measure. Interestingly, internal reliability was highest for the trial-type that typically indicated racial bias (the *Black-Negative* trial-type). The known-groups study provided strong support for the validity of the race IRAP because it clearly discriminated between black and white participants, and increased predictive validity over the explicit measures. When EEG was employed as an additional dependent measure of IRAP performance, significant differences between responding on pro-white/anti-black versus pro-black/anti-white blocks was observed, thus providing additional support for the reliability and validity of the measure. Finally, consistent with previous studies on the malleability of implicit measures, performance on the race IRAP shifted in a predicted direction as a result of two interventions that were designed to reduce racial prejudice. These data again provide support for the validity of the measure. In sum, the research reported in the current thesis has yielded an IRAP that could be used in subsequent research in the study of implicit racial bias.

Wider Implications of the Research

The REC model. Reducing the response latency criterion from 3000 to 2000ms, and introducing trial-by-trial temporal feedback (across Experiments 1-4), increased implicit racial bias on the IRAP. This result is consistent with the findings of Barnes-Holmes, Murphy, et al. (2010), and also with the REC model (Barnes-Holmes, et al., in press). As described in the Introductory chapter, the model assumes that specific IRAP trials may produce an immediate and relatively brief relational response before the participant actually presses a response key. The probability of this initial response will

often be determined by the verbal and nonverbal histories of the participant and by current contextual variables. By definition, the most probable immediate response will be emitted first most often, therefore any IRAP trial that requires a key press that coordinates with that immediate response will be emitted relatively quickly; however, if an IRAP trial requires a key press that opposes the immediate relational response, it may be emitted less quickly. Accordingly, across multiple trials, the average latency for inconsistent blocks will be longer than for consistent trials. In short, the IRAP effect is based on immediate relational responding, which is made apparent to the researcher when the behavioural system is put under pressure to respond quickly and accurately.

Given that pressure to respond quickly was greatest in Experiment 2 onwards, the results indicate that the immediate relational responses White–Positive–True and Black–Negative–True predominated (for white participants). According to the REC model, such response patterns would likely emerge from exposure to some of the verbal and nonverbal contingencies that operate for white individuals who have grown up and live in Ireland (e.g., the common portrayal of black males in the North American and British media as violent gun-carrying gang members). In attempting to explain why such contingencies had little if any impact on self-reports, the REC model assumes that responses to these measures likely reflected relatively elaborate and coherent relational responding. In other words, when asked to express an attitude or belief on a particular issue, it is likely that an individual will produce a relational response that coheres with one or more other relational responses in his or her behavioural repertoire (see Barnes-Holmes, Hayes, & Dymond, 2001). Imagine, for example, that a participant produced the same ratings for black and

white men on the semantic differentials. Such relational responses would likely cohere with *other* relevant relational networks, such as "The only difference between these pictures is skin colour" and "Racism is wrong." The important point to note here is that explicit measures are typically not completed under high time pressure, and thus participants have ample time to engage in the extended relational responding that is needed to produce a response that coheres with other relational responses. When exposed to a time-pressured IRAP, however, participants are not afforded the opportunity to elaborate because there is insufficient time, on a trial-by-trial basis, to engage in the additional and sometimes complex relational activity that serves to generate a relationally coherent response.

In summary, therefore, the REC model assumes that the IRAP effect, when produced under sufficient time pressure, is driven largely by immediate and relatively brief relational responses, whereas explicit measures reflect extended and coherent relational networks. The core of the REC model explanation for the impact of increased time pressure on the divergence between implicit and explicit measures rests on two assumptions. Firstly, immediate or automatic evaluative responses may or may not cohere with subsequent relational responding. When they cohere, implicit and explicit measures will typically converge, but when they do not, the measures will typically diverge. In other words, it is assumed that participants usually "reject" their immediate and brief relational responses (or automatic evaluations) if they do not cohere with their more elaborate and extended relational responding. Secondly, the REC model predicts that the divergence between implicit and explicit "socially sensitive" attitudes should increase with greater

time pressure on the IRAP, because participants have less time to engage in elaborated relational responding. In effect, as time pressure increases on the IRAP, the "contaminating" effects of elaborated relational responding on response latencies decrease.

Note, however, that the REC model does not predict that decreasing time pressure on the IRAP will necessarily produce increasing convergence with explicit measures. As time pressure decreases, it is difficult to predict exactly what variables will impact upon response latency, and thus the potential utility of the measure is lost. Indeed, the current findings support this conclusion because the internal reliability of the IRAP was absent at the level of the individual trial-type when the response latency criterion was set at the upper value of 3000ms. When the criterion was set at 2000ms, internal reliability tended to be moderate to strong and significant for the overall IRAP effect and for the *Black-Negative* trial-type; for the black participants, significant internal reliability was recorded for the *White-Negative* trial-type. At the present time it remains unclear why internal reliability differed across different IRAP trial-types, but the relative relational response strengths targeted by each of the trial-types may be involved here. That is, perhaps stronger or more probable relational responses tend to yield greater internal reliability simply because those responses vary less than weaker ones. Indeed, the idea that different trial-types target relational responses of different strength is relevant to another feature of the current findings discussed below.

As noted previously, the pattern of results for white participants across Experiments 2-7 produced strong IRAP effects for the *White-Positive* and *Black–Negative* trial types. According to the REC model, therefore, the data from these experiments

indicate that frames of coordination (i.e., the verbal relation of equivalence or similarity) between "White" and "Positive" and between "Black" and "Negative" were relatively strong, but frames of distinction (i.e., the verbal relation of difference) between ""White" and "Negative" and between "Black" and "Positive" for the most part, were not (the word *strong* is used here simply to denote a high probability in *immediate* relational responding). The REC model assumes that such differences in relational response strengths may be attributed, at least in part, to the verbal and nonverbal contingencies surrounding racial stereotyping. For example, common verbal practices would typically summarize such stereotyping as "white is good" and "black is bad," rather than "white is *not* bad" and "black is *not* good." In other words, two elements of a relational network may well cohere, as in "X is good" and "X is not bad," but the relative strengths or weaknesses of the two elements will be influenced to some degree by other variables, such as differences in frequency of exposure to the two parts of the network. The current results are therefore readily explained by the REC model, although testing the model systematically will have to await further empirical inquiry.

It is worth noting that the specific pattern of responding observed in Experiments 2, 3, 4, 5 and 7 for the *Black–Negative* but not the *Black-Positive* trial type has been observed previously in IRAP studies on race (Barnes-Holmes, Murphy, et al., 2010) and homonegativity (Cullen & Barnes-Holmes, 2008). These effects appear consistent with recent evidence that indicates the influence of a "negativity bias" in attitude formation (cf. Kunda, 1999). Specifically, when negatively valenced stimuli are presented with "Black" or "Gay," this serves to activate an implicit anti-black or anti-gay bias, respectively, which

is not observed when positively valenced stimuli are presented. On balance, procedural variables specific to the IRAP may be involved here. For example, the stereotyping effect for the *Black-Negative* trial type required responding "True" more quickly than "False," but the opposite was required for the *Black-Positive* trial type. It is possible, therefore, that a bias toward responding "True" over "False," per se, interacted with the socially loaded stimulus relations presented in the IRAP. If such a response bias does play a role, however, the source of that bias needs to be explained. As suggested previously, the impact of common verbal practices, which tend to confirm negative rather than deny positive stereotypes, is a possibly important variable.

Future research. The current research programme raises at least two areas of study that will need further attention. First, the extent to which the IRAP predicts actual behaviour in the natural environment needs to be determined. Second, the duration of any change in an IRAP performance following an intervention needs to be established. A preliminary attempt was made to address the first issue in the final experiment reported in the current thesis. Specifically, participants were asked to complete Behavioural Intentions Questionnaires (BIQs) for white and black people both before and after an intervention. However, none of the IRAP effects, or changes in those effects from pre- to post-intervention, correlated with the BIQs. Interestingly, this finding contrasts with a previous IRAP study on implicit body-size bias in which the measure *did* correlate with a BIQ (Roddy, et al., 2010). At the present time it remains unclear why this correlational effect was not replicated in the current research. Perhaps when participants are asked to report their behavioural intentions towards pictures of black and white people social desirability

is more salient than when the stimuli are *white* overweight and normal-weight individuals. Insofar as this is the case, then the race BIQs may have functioned more as standard explicit measures of racial bias rather than valid measures of behavioural intentions. In any case, given that previous research has repeatedly shown that implicit measures do predict behaviours (see Greenwald, Poehlmann, Uhlmann, & Banaji, 2009, for a review), future studies are needed to determine to what extent the IRAP predicts racially-biased behaviour in the natural environment.

The second issue that needs to be addressed in future work was also highlighted in the final experiment of the current thesis. Specifically, the findings indicated that both implicit and explicit attitudes changed as a result of two interventions. However, no attempt was made to determine the relative persistence of these changes. In the only other IRAP study that investigated the malleability of implicit attitudes by a direct intervention, a follow-up measure was conducted 24 hours later and the changes in implicit attitudes were maintained, but with some suggestion that they were reverting back to the original pattern (Cullen, et al., 2009). An earlier study using the IAT also reported a similar effect (Dasgupta & Greenwald, 2001). At the present time, however, it remains unclear exactly how long changes on implicit attitude measures will last following a relevant intervention. Obviously, this is an important area of research in terms of developing effective psycho-social interventions for racial bias, or prejudice more generally (e.g., Lillis & Hayes, 2007). In particular, it will be important to determine to what extent changes in measures of implicit and explicit attitudes predict, either together or independently, changes in specific types of racially-biased behaviours following a relevant intervention.

Conclusion

The programme of research reported in the current thesis lead to the development and refinement of an IRAP that could be used to measure implicit racial bias. The same overall pattern of pro-white and anti-black IRAP effects was produced by white participants across most of the experiments, and internal reliability was also relatively robust. The validity of the race IRAP was supported because it only correlated with the explicit measures for those participants who were low in motivation to conceal prejudice. Furthermore, it clearly discriminated between black and white participants, and critically it increased predictive validity over a range of explicit measures. The recording of electroencephalograms provided additional support for the reliability and validity of the measure. Finally, performance on the IRAP shifted in a predicted direction following two interventions that were designed to reduce racial prejudice, thus providing further support for its validity. The research reported in the current thesis has thus produced an IRAP that could be used in subsequent research in the study of implicit racial bias. In particular, this research will need to focus on the extent to which IRAP performances predict actual behaviour in the natural environment, and on the relative persistence in changes in IRAP performances following interventions designed to undermine social prejudice.

REFERENCES

REFERENCES

Allport, G. W. (1954). *The nature of prejudice,* Reading, MA: Addison-Wesley, 23, 107-117.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*, 179-211.

Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology, 52*, 27-58.

Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behaviour. In D. Albarracin, B. T. Johnson, & M. P. Zanna (Eds.), *Handbook of attitude and attitude change* (pp.173-214). Mahwah, NJ: Erlbaum.

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

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

Barnes, D., & Hampson, P. (1993). Stimulus equivalence and connectionism: Implications for behaviour analysis and cognitive science. *The Psychological Record, 43*, 617-638.

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

Barnes, D., McCullagh, P., & Keenan, M. (1990). Equivalence class formation in non-hearing impaired children and hearing impaired children. *Analysis of Verbal Behavior, 8*, 1-11.

Barnes-Holmes, D., Barnes-Holmes, Y., & Cullinan, V. (2000). Relational frame theory and Skinner's Verbal Behavior: A possible synthesis. *The Behavior Analyst, 23*, 69-84.

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

Barnes-Holmes, D., Barnes-Holmes, Y., Smeets, P. M., Cullinan, V., & Leader, G. (2004). Relational frame theory and stimulus equivalence: Conceptual and procedural issues. *International Journal of Psychology & Psychological Therapy, 4*, 161-193.

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (in press). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) Model. *The Psychological Record*.

Barnes-Holmes, D., Hayden, E., Barnes-Holmes, and Stewart, I. (2008). The Implicit Relational Assessment Procedure (IRAP) as a response-time and event-related-potentials methodology for testing natural verbal relations: A preliminary study. *The Psychological Record, 58*, 497-516.

Barnes-Holmes, D., Hayes, S. C., & Dymond, S. (2001). Self and self-directed rules. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational Frame Theory: A post-Skinnerian account of human language and cognition* (pp. 119-140). New York: Plenum.

Barnes-Homes, D., Murphy, A., Barnes-Holmes, Y., & Stewart, I. (2010). The Implicit Relational Assessment Procedure (IRAP): Exploring the impact of private versus public contexts and the response latency criterion on pro-white and anti-black stereotyping among white Irish individuals. *The Psychological Record, 60*, 57-66.

Barnes-Holmes, D., Murtagh, L., Barnes-Holmes, Y., & Stewart, I. (2010). Using the Implicit Association Test and the Implicit Relational Assessment Procedure to measure attitudes towards meat and vegetables in vegetarians and meat-eaters. *The Psychological Record, 60*, 287-306.

Barnes-Holmes, D., Regan, D., Barnes-Holmes, Y., Comins , S., Walsh, D., Stewart, I., et al . (2005). Relating derived relations as a model of analogical reasoning: Reaction times and event related potentials. *Journal of the Experimental Analysis of Behavior. 84*, 435–452.

Barnes-Holmes, D., Staunton, C., Barnes-Holmes, Y., Whelan, R., Stewart, I., Commins, S., et al. (2004). Interfacing relational frame theory with cognitive neuroscience: Semantic priming, the implicit association test, and event related potentials. *International Journal of Psychology and Psychological Therapy, 4*, 215–240.

Barnes-Holmes, D., Waldron, D., Barnes-Holmes, Y., & Stewart, I. (2009). Testing the validity of the Implicit Relational Assessment Procedure (IRAP) and the Implicit Association Test: Measuring attitudes towards Dublin and country-life in Ireland. *The Psychological Record, 59*, 389-40.

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

Baron, R. A., Byrne, D., & Watson, G. (2004). *Exploring Social Psychology (4th ed.).* Canada: Pearsons.

Baum, W. M. (1994). *Understanding behaviorism: Science, behavior, and culture.* New York: Harper-Collins.

Beale, D. A., & Manstead, A. S. R. (1991). Predicting mothers' intentions to limit frequency of infants'sugar intake: Testing the theory of planned behavior. *Journal of Applied Social Psychology, 21*, 409-431.

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 1-62). San Diego, CA: Academic Press.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*, 242-261.

Blair, I. V., & Banaji, M. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology, 70*, 1142-1163.

Boysen, G. A., Vogel, D. L., & Madon, S. (2006). A public versus private administration of the Implicit Association Test. *European Journal of Social Psychology 36*, 845-856.

Breckler, S. J. (1984). Empirical validation of affect, behaviour, and cognition as distinct components of attitude. *Journal of Personality & Social Psychology, 47*, 1191-1205.

Brochu, P. M., & Morrison, M. A. (2007). Implicit and explicit prejudice toward overweight and average weight men and women: Testing their correspondence and relation to behavioral intentions. *Journal of Social Psychology, 147*, 681-706.

Campbell, D. T. (1950). The indirect assessment of social attitudes. *Psychological Bulletin, 47*, 15-38.

Central Statistics Office Ireland (2006). Retrieved December 10, 2009, from http:// www.cso.ie/census/default.htm.

Chaiken, S., & Stangor, C. (1987). Attitudes and attitude change. *Annual Review of Psychology, 38*, 575- 630.

Chambliss, H. O., Finley, C. E., & Blair, S. N. (2004). Attitudes toward obese individuals among exercise science students. *Medicine and Science in Sports and Exercise, 36*, 468–474.

Chan, G., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). Implicit attitudes to work and leisure among North American and Irish individuals: A preliminary study. *International Journal of Psychology & Psychological Therapy, 9*, 317-334.

Cowley, J., Green, G., & Braunling-McMorrow, D. (1992). Using stimulus equivalence procedures to teach name-face matching to adults with brain injuries. *Journal of Applied Behaviour Analysis, 25*, 461-475.

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

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science, 15*, 806-813.

Dambrun, M., & Guimond, S. (2004). Implicit and explicit measures of prejudice and stereotyping: Do they assess the same underlying knowledge structure? *European Journal of Social Psychology, 34*, 663-676.

Dasgupta, N. G., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality & Social Psychology, 81*, 800-814.

Dasgupta, N. G, 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.

Dawes, R. M. (1972). *Fundamentals of attitude measurement.* New York: Wiley.

Dawson, D. L., Barnes-Holmes, D., Greswell, D. M., Hart, A. J. P., & Gore, N. J. (2009). Assessing the implicit beliefs of sexual offenders using the Implicit Relational Assessment Procedure: A first study. *Sexual Abuse: A Journal of Research and Treatment, 21*, 57-75.

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

De Houwer, J. (2003b). The Extrinsic Affective Simon Task. *Experimental Psychology, 50*, 77-85.

De Houwer, J. (2006). What are implicit measures and why are we using them? In R. W. Wiers & A. W. Stacy (Eds.), *The handbook of implicit cognition and addiction* (pp. 11-28). CA: Sage Publishers.

De Houwer, J., & De Bruycker, E. (2007). The Implicit Association Test outperforms the Extrinsic Affective Simon Task as an implicit measure of interindividual differences in attitudes. *British Journal of Social Psychology, 46*, 401-421.

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

Devany, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behaviour, 46*, 243-257.

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

Devine, P. G., Monteith, M., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology, 60*, 817-830.

Dickins, D. W., Singh, K. D., Roberts, N., Burns, P., Downes, J. J., Jimmieson, P., & Bentall, R. P. (2001). An fMRI study of stimulus equivalence. *Neuroreport, 12*, 405-411.

Dugdale, N., & Lowe, F. C. (2000). Testing for symmetry in the conditional discriminations of language trained chimpanzees. *Journal of the Experimental Analysis of Behavior, 73*, 5-22.

Eagly, A. H., & Chaiken, S. (1993). *The Psychology of Attitudes*. Orlando, FL: Harcourt College Publishers.

Eagly, A. H., and Chaiken, S. (2007). The advantages of an inclusive definition of attitude. *Social Cognition, 25*, 582-602.

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion, 15*, 115-141.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297-327.

Fazio, R. H., & Petty, R. E. (2008). *Attitude, their structure, function and consequences.* New York: Psychology Press.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality & Social Psychology, 50*, 229-238.

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

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behaviour: An introduction to theory and research*. Reading MA: Addison-Wesley.

Gapinski, K. D., Schwartz, M. B., & Brownell, K. D. (2006). Can television change anti-fat attitudes and behaviour? *Journal of Applied BioBehavioural Research, 11,* 1-28.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*, 692-731.

Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science, 2*, 181-193.

Gemar, M. C., Segal, Z. V., Sagrati, S., & Kennedy, S. J. (2001). Mood-induced changes on the Implicit Association Test in recovered depressed patients. *Journal of Abnormal Psychology, 110*, 282- 289.

Grant, L., & Evans, A. (1994). *Principles of behavior analysis*. New York: Harper Collins.

Greenwald, A. G. (1989). Why attitudes are important: Defining attitude and attitude theory 20 years later. In A. R. Pratkanis, S. J. Breckler & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 429-440). Hillsdale, NJ: Erlbaum.

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*(1), 3-25.

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

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality & Social Psychology, 85*, 198-216.

Greenwald, A. G., Poehlmann, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality & Social Psychology, 97*, 17-41.

Grey, I., & Barnes, D. (1996). Stimulus equivalence and attitudes. *The Psychological Record, 46*, 243- 270.

Guglielmi, R. S. (1999). Psychophysiological assessment of prejudice: Past research, current status, and future directions. *Personality & Social Psychology Review, 3*, 123-157.

Hayes, S. C. (1989). Nonhumans have not yet shown stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 51*, 385-392.

Hayes, S. C., & Barnes-Holmes, D. (2004). Relational operants: processes and implications: A response to Palmer's review of relational frame theory. *Journal of the Experimental Analysis of Behavior, 82*, 213-224.

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., & Hayes, L. J. (1989). The verbal action of the listener as a basis for rule-governance. In S. C. Hayes (Ed.), *Rule-governed behavior: Cognition, contingencies, and instructional control* (pp. 153-190). New York: Plenum.

Hayes, S. C., & Wilson, K. G. (1993). Some applied implications of a contemporary behavior-analytic account of verbal events. *The Behavior Analyst, 16*, 283-301.

Holcomb, P. J., & Anderson, J. E., (1993). Cross-modal semantic priming: A time-course analysis using event-related potentials. *Language and Cognitive Processes, 8*, 327–411.

Holland, R. W., Verplanken, B., & Van Kippenberg, A. (2002). On the nature of attitude-behaviour relations: The strong guide, the weak follow. *European Journal of Social Psychology, 32*, 869- 876.

Ito, T. A., & Cacioppo, J. T. (2007). Attitudes as mental and neural states of readiness: Using physiological measures to study implicit attitudes. In B. Wittenbrink & N. Schwarz, *Implicit Measures of Attitudes* (pp.125-158). New York: Guilford Press.

Katz, D., & Stotland, E. (1959). A preliminary statement to a theory of attitude structure and change. In S. Koch (Ed.), *Psychology: A study of science* (Vol. 3, pp. 423-475). New York: Mc Graw-Hill.

Kawakami, K., Dovidio, J.F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology. 78*, 871-888.

Kendall, S. B. (1983). Tests for mediated transfer in pigeons. *The Psychological Record, 33*, 245-256.

Kounios , S. A., & Holcomb , P. J. (1992). Structure and process in semantic memory: Evidence from event-related potentials and reaction times. *Journal of Experimental Psychology: General, 121*, 460–480.

Krech, D., & Crutchfield, R. S. (1948). *Theory and problems of social psychology*. New York: McGraw- Hill.

Krosnick, J. A. (1998). Attitude importance and attitude change. *Journal of Experimental Social Psychology, 24*, 240-255.

Kunda, Z. (1999). *Social cognition: Making sense of people*. Cambridge, MA: MIT Press.

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for simple-word and sentence-context effects. *Language and Cognitive Processes, 8*, 533–578.

Kutas, M., & Hiliard , S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 1161–1163.

LaPiere, R. T. (1934). Attitudes vs. actions. *Social Forces, 13*, 230-237.

Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology, 72*, 275-287.

Leslie, J. C., & O'Reilly, M. F. (1999). *Behavior Analysis: Foundations and applications to psychology*. New York: Psychology Press.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 5-53.

Lillis, J., & Hayes, S. C. (2007). Applying acceptance, mindfulness, and values to the reduction of prejudice: A pilot study. *Behavior Modification, 31*(4), 389-411.

Livingston, R. W. (2002). The role of perceived negativity in the moderation of African Americans' implicit and explicit racial attitudes. *Journal of Experimental Social Psychology, 38*, 405–413.

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality & Social Psychology, 81*, 842-855.

Maison, D., Greenwald, A. G., & Bruin, R. H. (2004). Predictive validity of the Implicit Association Test in studies of brands, consumer attitudes, and behaviour. *Journal of Consumer Psychology, 14*, 405-415.

McKenna, I., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2007). Testing the fake-ability of the Implicit Relational Assessment Procedure (IRAP): The first study. *International Journal of Psychology and Psychological Therapy, 7*, 253-268.

Montieth, M. J., Voils, C. I., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial biases. *Social Cognition, 19*, 395–417.

Moors A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*, 297–326.

Moskowitz, G. B., Salomon, A. R., & Taylor, C. M. (2000). Preconsciously controlling stereotyping: Implicitly activated egalitarian goals prevent the activation of stereotypes. *Social Cognition, 18*(2), 151–177.

Myrdal, Gunnar (1944). *An American dilemma: The negro problem and modern democracy*. New York: Harper & Bros.

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

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

Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics, 6*, 101-115.

Nosek, B. A., & Hansen, J. J. (2008). The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation. *Cognition and Emotion, 22*, 553-594.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Tucker Smith, C., Olson, K. R., Chugh, D., Greenwald, A. G., and Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18*, 36-88.

Nye, R. D. (1975). *Three view of man: Perspectives from Sigmund Freud, B.F. Skinner, and Carl Rogers*. CA: Brooks-Cole.

O'Brien, K. S., Hunter, J. A., Halberstadt, J., & Anderson, J. (2007). Body image and explicit and implicit anti-fat attitudes: The mediating role of physical appearance comparisons. *Body Image,* 4, 249- 256.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776-783.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Chicago: Urbana.

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

Paulhus, D.L. (1984). Two-component model of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609.

Payne, K.B., Govorun, O., & Arbuckle, N.L. (2008). Automatic attitudes and alcohol: Does implicit liking predict drinking? *Cognition & Emotion, 22*(2), 238-271.

Peña, Y., Sidanius, J., & Sawyer, M. (2004). Racial democracy in the Americas: A Latin and North American comparison. *Journal of Cross-Cultural Psychology, 35*, 749-767.

Petty, R. E., Fazio, R. H., & Brinol, P. (2008). *The new implicit measures: An overview*. New York: Psychology Press.

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

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

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention grabbing power of negative social information. *Journal of Personality and Social Psychology, 61*, 380-391.

Raja, S., & Stokes, J. P. (1998). Assessing attitudes toward lesbians and gay men: The Modern Homophobia Scale. *Journal of Gay, Lesbian, and Bisexual Identity, 3*, 113-134.

Rasinski, K. A. (1989). The effect of question wording on public support for government spending. *Public Opinion Quarterly, 53*, 388-394.

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

Richeson, J. A., & Ambady, N. (2003). Effects of situational power on automatic racial prejudice. *Journal of Experimental Social Psychology, 39*, 177–183.

Roche, B., & Barnes, D. (1996). Arbitrarily applicable relational responding and sexual categorization: A critical test of the derived difference relation. *The Psychological Record, 46*, 451-475.

Roddy, S., Stewart, I., & Barnes-Holmes, D. (2010). Anti-fat, pro-slim, or both? Using two reaction time based measures to assess implicit attitudes to the slim and overweight. *Journal of Health Psychology, 15*, 416-425.

Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioual components of attitudes. In C. I. Hovland & M. J. Rosenberg (Eds.), *Attitude organization and change: An analysis of consistency among attitude components* (pp. 1-14). New Haven, CT: Yale University Press.

Rudman, L. A., Feinberg, J., & Fairchild, K. (2002). Minority members' implicit attitudes: Automatic ingroup bias as a function of group status. *Social Cognition, 20*, 294-320.

Rudman, L. A., Greenwald, A. G., & McGhee, D. E. (2001). Implicit self-concept and evaluative implicit gender stereotypes: Self and in-group share desirable traits. *Personality & Social Psychology Bulletin, 27*, 1164-1178.

Rust, J., & Golombok, S. (1999). *Modern psychometrics: The science of psychological assessment* (2nd ed.). New York: Routledge.

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

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behaviour, 37*, 5-22.

Sue, D. W., & Sue, D. (2003) *Counselling the culturally diverse* (4th ed.). New York: John Wiley & Sons.

Teachman, B. A., & Brownell, K. D. (2001). Implicit anti-fat bias among health professionals: Is anyone immune? *International Journal of Obesity, 25*, 1525-1531.

Teachman, B. A, Gapinski, K. D., Brownell, K. D., Rawlins, M., & Jeyaram, S. (2003). Demonstrations of implicit anti-fat bias. *Health Psychology, 22*, 68-78.

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

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, 33*, 529-554.

Triandis, H. C. (1971). *Attitude and attitude change*. New York: Wiley.

Vahey, N. A., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). A first test of the Implicit Relational Assessment Procedure (IRAP) as a measure of self-esteem: Irish prisoner groups and university students. *The Psychological Record, 59*, 371-388.

Watt, A., Keenan, M., Barnes, D., & Cairns, E. (1991). Social categorization and stimulus equivalence. *The Psychological Record, 41*, 33-50.

Weisbrod, M., Keifer, M., Winkler, S., Maier, S., Hill, R., Roesch-Ely, D., et al. (1999). Electrophysiological correlates of direct versus indirect semantic priming in normal volunteers. *Cognitive Brain Research, 8*, 289–298.

Wicker, A. W. (1969). Attitude versus actions: The relationship of verbal and overt behavioural responses to attitude objects. *Journal of Social Issues, 25*, 41-78.

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

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

APPENDICES

APPENDICES

Appendix A
The Discrimination (DS) and Diversity Scales (DV)

Please read each of the following statements and rate them as honestly as you can by circling the number that you feel is most appropriate on each scale.

Appendix A continued
The Discrimination (DS) and Diversity Scales (DV)

DS

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

Strongly Agree — Strongly Disagree

1 2 3 4 5

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

Strongly Agree — Strongly Disagree

1 2 3 4 5

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

Strongly Agree — Strongly Disagree

1 2 3 4 5

4. These days, reverse discrimination against Whites is as much a problem as discrimination against Blacks itself.

Strongly Agree — Strongly Disagree

1 2 3 4 5

5. More and more, Blacks use accusations of racism for their own advantage.

Strongly Agree — Strongly Disagree

1 2 3 4 5

Appendix A continued
The Discrimination (DS) and Diversity Scales (DV)

6. Blacks are ultimately responsible for the state of race relations in this country.

7. Discrimination against Blacks is not a problem in Ireland.

8. A primary reason that ethnic minorities tend to stay in lower paying jobs is that they lack the motivation required for moving up.

9. Many ethnic minorities do not understand how hard one has to work to achieve success.

10. In Ireland people are not judged by their skin colour.

Appendix A continued
The Discrimination (DS) and Diversity Scales (DV)

DV

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

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

3.- Whites will need to learn about Black culture if positive interethnic relations are to be achieved.

4. The establishment and maintenance of all-Black groups and coalitions prevents successful racial integration.

Appendix B
The Semantic Differential Scales

The purpose of these scales is to find out your attitudes to people and so you are requested to rate how you feel. Please use your first impression and try not to figure out the "right answer" or the answer that makes most sense. Please work quickly by marking an 'X' in th3e place where you feel is most appropriate. All of your responses will remain anonymous and confidential

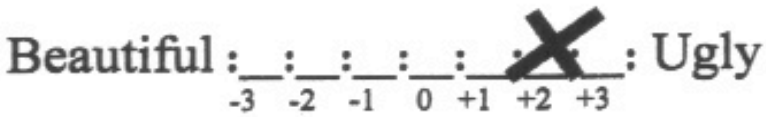
Appendix B continued
The Semantic Differential Scales

The following is a Semantic Differential Scale,

Your task is to mark the position on the scale (as shown in the example below) with an 'X' where you feel most applies.

EXAMPLE

Spider



If you have any questions please ask now.
If not, please turn the page and start immediately.

Appendix B continued
The Semantic Differential Scales

Black People

Friendly :__:__:__:__:__:__:__: Hostile
-3 -2 -1 0 +1 +2 +3

Black People

Honest :__:__:__:__:__:__:__: Deceitful
-3 -2 -1 0 +1 +2 +3

Black People

Lazy :__:__:__:__:__:__:__: Hardworking
-3 -2 -1 0 +1 +2 +3

Black People

Peaceful :__:__:__:__:__:__:__: Violent
-3 -2 -1 0 +1 +2 +3

Black People

Bad :__:__:__:__:__:__:__: Good
-3 -2 -1 0 +1 +2 +3

Black People

Stupid :__:__:__:__:__:__:__: Clever
-3 -2 -1 0 +1 +2 +3

Appendix B continued
The Semantic Differential Scales

White People

Friendly :__:__:__:__:__:__:__: Hostile
-3 -2 -1 0 +1 +2 +3

White People

Honest :__:__:__:__:__:__:__: Deceitful
-3 -2 -1 0 +1 +2 +3

White People

Lazy :__:__:__:__:__:__:__: Hardworking
-3 -2 -1 0 +1 +2 +3

White People

Peaceful :__:__:__:__:__:__:__: Violent
-3 -2 -1 0 +1 +2 +3

White People

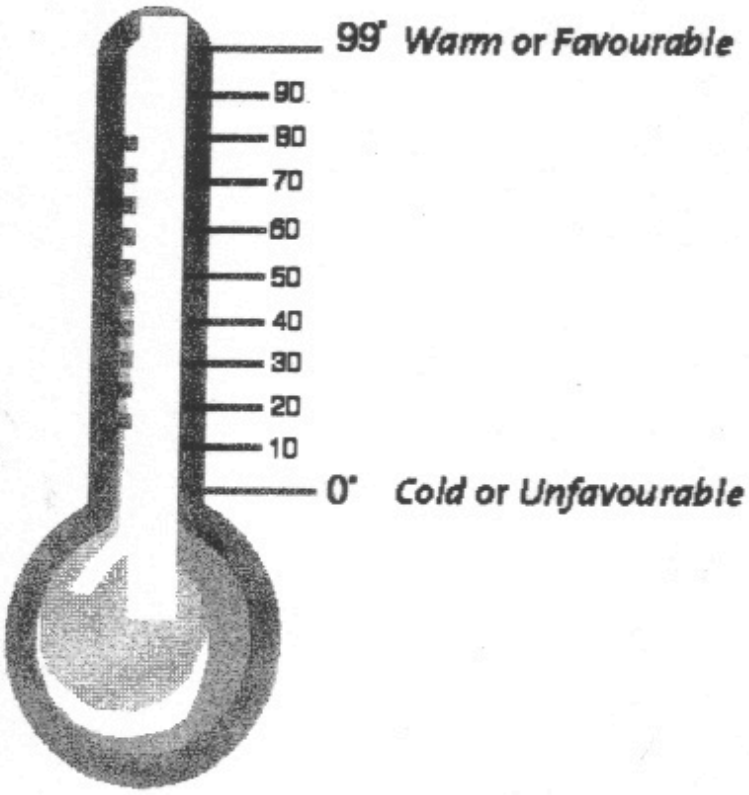
Bad :__:__:__:__:__:__:__: Good
-3 -2 -1 0 +1 +2 +3

White People

Stupid :__:__:__:__:__:__:__: Clever
-3 -2 -1 0 +1 +2 +3

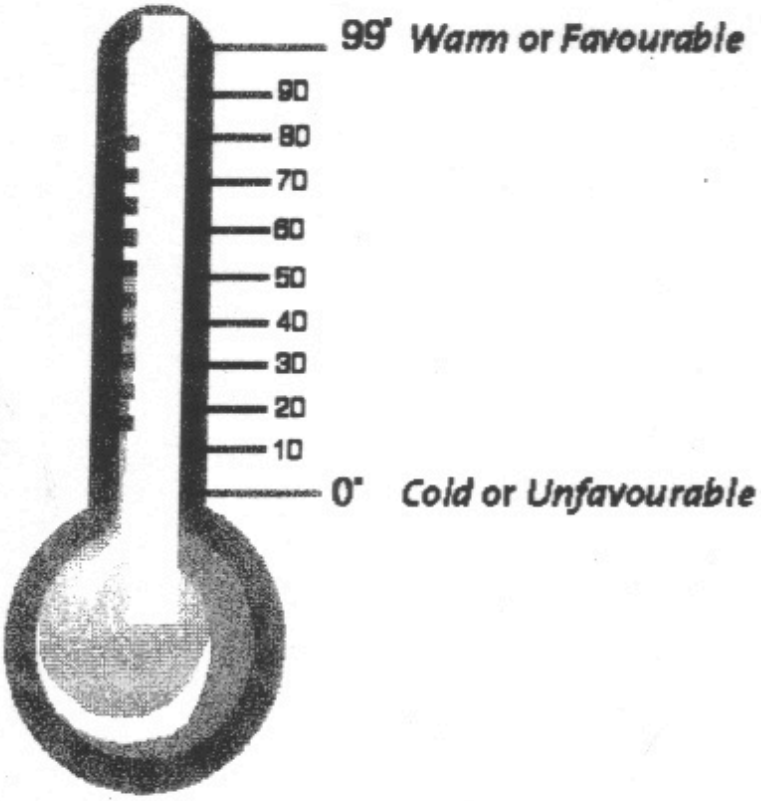
Appendix C
The Feeling Thermometers
Appendix C Continued
The Feeling Thermometers

Please rate how you feel about <u>white people:</u>



Appendix C continued
The Feeling Thermometers

Please rate how you feel about <u>black people:</u>



Appendix D

The Internal and External Motivation to Conceal Prejudice Scale

Instructions: the following questions concern various reasons or motivations people might have for trying to respond in non-prejudiced ways toward Black people. Some of the reasons reflect internal-personal motivations whereas others reflect more external-social motivations. Of course, people may be motivated for both internal and external reasons; we want to emphasize that neither type of motivation is by definition better than the other. In addition, we want to be clear that we are not evaluating you or your individual responses. All your responses will be completely confidential. We are simply trying to get an idea of the types of motivations that students in general have for responding in non-prejudiced ways. If we are to learn anything useful, it is important that you respond to each of the questions openly and honestly. Please give your response according to the scale below.

Please read each of the following statements and rate them as honestly as you can. Answer every question according to the rating scale below.

1--------2--------3--------4--------5--------6--------7--------8--------9

Strongly Disagree **Strongly Agree**

___1 Because of today's PC (Politically Correct) standards I try to appear non-prejudiced toward Black people.

___2 I attempt to act in non-prejudiced ways toward Black people because it is personally important to me.

___3 Being non-prejudiced toward Black people is important to my self-concept.

___4 I try to hide any negative thoughts about Black people in order to avoid negative reactions from others.

___5 If I acted prejudiced toward Black people, I would be concerned that others would be angry with me.

___6 Because of my personal values, I believe that using stereotypes about Black people is wrong.

___7 I attempt to appear non-prejudiced toward Black people in order to avoid disapproval from others.

___8 According to my personal values, using stereotypes about Black people is OK.

___**9** I try to act non-prejudiced toward Black people because of pressure from others.

___**10** I am personally motivated by my beliefs to be non-prejudiced toward Black people.

Appendix E

The Behavioural Intentions Questionnaires employed in Experiment 7

Based on the photograph of the female presented above, how likely is it that you would:

1. want to get to know Her better?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

2. ask Her if you could copy her notes from a class you missed?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

3. want to work on a class project with Her?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

4. invite Her to a study group for an exam?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

5. want to become friends with Her?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

Appendix E continued
The Behavioural Intentions Questionnaires employed in Experiment 7

Based on the photograph of the female presented above, how likely is it that you would:

1. want to get to know Her better?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

2. ask Her if you could copy her notes from a class you missed?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

3. want to work on a class project with Her?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

4. invite Her to a study group for an exam?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

5. want to become friends with Her?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

Appendix E continued
The Behavioural Intentions Questionnaires employed in Experiment 7

Based on the photograph of the female presented above, how likely is it that you would:

1. want to get to know Her better?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

2. ask Her if you could copy her notes from a class you missed?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

3. want to work on a class project with Her?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

4. invite Her to a study group for an exam?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

5. want to become friends with Her?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

Appendix E continued
The Behavioural Intentions Questionnaires employed in Experiment 7

Based on the photograph of the female presented above, how likely is it that you would:

1. want to get to know Her better?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

2. ask Her if you could copy her notes from a class you missed?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

3. want to work on a class project with Her?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

4. invite Her to a study group for an exam?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

5. want to become friends with Her?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

Appendix E continued
The Behavioural Intentions Questionnaires employed in Experiment 7

Based on the photograph of the female presented above, how likely is it that you would:

1. want to get to know Her better?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

2. ask Her if you could copy her notes from a class you missed?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

3. want to work on a class project with Her?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

4. invite Her to a study group for an exam?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

5. want to become friends with Her?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

Appendix E continued
The Behavioural Intentions Questionnaires employed in Experiment 7

Based on the photograph of the female presented above, how likely is it that you would:

1. want to get to know Her better?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

2. ask Her if you could copy her notes from a class you missed?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

3. want to work on a class project with Her?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

4. invite Her to a study group for an exam?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

5. want to become friends with Her?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

Appendix E continued
The Behavioural Intentions Questionnaires employed in Experiment 7

Based on the photograph of the male presented above, how likely is it that you would:

1. want to get to know Him better?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

2. ask Him if you could copy his notes from a class you missed?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

3. want to work on a class project with Him?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

4. invite Him to a study group for an exam?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

5. want to become friends with Him?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

Appendix E continued
The Behavioural Intentions Questionnaires employed in Experiment 7

Based on the photograph of the male presented above, how likely is it that you would:

1. want to get to know Him better?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

2. ask Him if you could copy his notes from a class you missed?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

3. want to work on a class project with Him?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

4. invite Him to a study group for an exam?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

5. want to become friends with Him?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

Appendix E continued
The Behavioural Intentions Questionnaires employed in Experiment 7

Based on the photograph of the male presented above, how likely is it that you would:

1. want to get to know Him better?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

2. ask Him if you could copy his notes from a class you missed?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

3. want to work on a class project with Him?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

4. invite Him to a study group for an exam?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

5. want to become friends with Him?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

Appendix E continued
The Behavioural Intentions Questionnaires employed in Experiment 7

Based on the photograph of the male presented above, how likely is it that you would:

1. want to get to know Him better?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

2. ask Him if you could copy his notes from a class you missed?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

3. want to work on a class project with Him?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

4. invite Him to a study group for an exam?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

5. want to become friends with Him?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

Appendix E continued
The Behavioural Intentions Questionnaires employed in Experiment 7

Based on the photograph of the male presented above, how likely is it that you would:

1. want to get to know Him better?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

2. ask Him if you could copy his notes from a class you missed?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

3. want to work on a class project with Him?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

4. invite Him to a study group for an exam?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

5. want to become friends with Him?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

Appendix E continued
The Behavioural Intentions Questionnaires employed in Experiment 7

Based on the photograph of the male presented above, how likely is it that you would:

1. want to get to know Him better?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

2. ask Him if you could copy his notes from a class you missed?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

3. want to work on a class project with Him?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

4. invite Him to a study group for an exam?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

5. want to become friends with Him?

1 2 3 4 5 6 7

Very Unlikely Neutral Very Likely

Appendix F

The Acceptance and Commitment Therapy (ACT)-based and Education-based protocols employed in Experiment 7

ACT-based protocol

The following information and exercises are going to deal with prejudicial thoughts and feelings.

Many people believe that prejudicial thoughts and feelings are the biggest barrier to a society that is accepting, inclusive, and benevolent.

Ask yourself if you would agree with that.

What I want you to do now is to take a look at your own prejudicial thoughts and feelings, in a different way than you may have in the past, and see what your experience tells you about what role that these thoughts and feelings play in your life.

The first thing I want you to do is turn your attention to your mind and your thoughts. During this exercise, I want you to focus on NOTICING and RECORDING the thoughts that your mind gives you, without trying to fight or change them. Do NOT censor yourself. Whatever comes up is fine, just notice…

I want you to write down what comes to mind when you read the following…

Most white people are… _____

Most black people tend to…_____

Some racial slurs I know are…

_____

People who live in this country and don't speak the language are…

_____

If I was the only person of my race in a public place, I would feel the most uncomfortable around people who were…

_____

Many blacks and don't do well in school because…

_____

From time to time, white people can be…

_____

Appendix F continued
The Acceptance and Commitment Therapy (ACT)-based and Education-based protocols employed in Experiment 7

ACT-based protocol

Some racially insensitive thoughts I have are…

_____

I wish black people wouldn't…

_____

What did you notice? Reactions? What do you think about that stuff? Were you able to notice what your mind was giving you, the fact that thoughts came up? Were you able to let stuff come up without censoring, fighting, changing?

Now I want you to write down what comes to mind when you read the following…

Mary had a little…_____

Blondes have more… _____

There's no place like…_____

Wow. How did that happen? Do you think that most people would have come up with the same answers?

What's the deal with thoughts? It seems like we go through life looking to our own thoughts as the authority on everything. We take our thoughts VERY literally, use them as evidence that something is good or bad, or that someone is right or wrong.

Only, as we just saw, it seems like a lot of our thought content is programmed in from past experiences. Not only that, but once something is programmed in, it is unlikely to disappear altogether, so it is subject to the whim of your life. You may have a whole host of experiences that could bring up this thought or that. But because we don't pay attention to the actual process of having thoughts, we don't really see how random this is. Next I want you to do a little exercise and see if this doesn't fit for you…

I want you to say and to remember the numbers, 1, 2, 3. What are the numbers?
- what are the chances tomorrow, next week, next year, even on your death bed that you will forget?
- all because some crazy psychology student put it in your head

–isn't that silly, all that valuable brain space wasted?

Appendix F continued
The Acceptance and Commitment Therapy (ACT)-based and Education-based protocols employed in Experiment 7

ACT-based protocol

- but, what if we can take perspective, just notice our prejudices, become aware of them, and recognize that in some way they may be the same as 1, 2, 3?

What if it were the case that we were never going to be able to get rid of own prejudicial thoughts?
What if we could simply notice them, be aware that they are historical, and not buy into them? Would we need to get rid of them to behave in ways that are open, sensitive, and positive, towards others?

I want to make a distinction between having a thought and *buying* a thought. When you BUY into a thought, you look at the world from that thought, like looking through yellow colored glasses. When you simply HAVE a thought, you notice that your mind produced a thought, the process, and that you are not the thought, the thought need not necessarily be true, but rather just something that came up that **may or may not** be useful to pay attention to, like a pop-up add on a computer…

Imagine that you are applying for a desirable job. You've been out of work. Money is tight. You really need to get back to work. Get present with what that would feel like for a moment… You've met the other candidates and you're sure that you are the best. Your resume is better, you interviewed well. You're confident that you're getting this job. Just get present with this for a moment, what that would feel like… Then you find out that you're not getting the job, that the job went to another candidate, and that that candidate was a different race than you. Notice what happens in your body and try to sit with it, without changing it…
Check in with yourself, what did you feel? _____
What race did you picture the other person to be? _____

Now, imagine you are on vacation, having a fun time in a city you've visited in the past. Take a moment to get present with that…After a fun day, it's time to go home, and you're walking back to your hotel room. It's late, though, so it's dark. You have some familiarity with the city, but you're not entirely sure you're going the right way. You may be lost. As you walk, many people are outside store fronts and buildings talking to each other and watching you as you go by, because you stick out, you're obviously not from the city. Now imagine all those people are the same race as you. Take a moment to get present with this situation, and notice how your body feels… Now imagine that you're walking, possibly lost, people outside are talking and watching you, and NONE of them are the same race as you, you're the only one. Now check in with yourself, take a moment, what's happening in your body…what has changed?

Appendix F continued
The Acceptance and Commitment Therapy (ACT)-based and Education-based protocols employed in Experiment 7

ACT-based protocol

Reactions? What did you notice?

_____

Think about some of your own stereotypes or prejudices and write down what shows up for you below:

It's hard sometimes to look at the fact that we have emotional and cognitive reactions that we don't want to have. We may think that that means something bad about us. Or we may use that to judge how open we are to other people, or how accepting we are. But, what if we could just notice and acknowledge these reactions, without attaching anything to them? If we could do this, maybe we wouldn't need to get rid of them or unlearn them to BEHAVE the way we want with other people…

Let's take a look at the issue of trying to control or change how we think or feel…

Don't think of chocolate cake. . .

Really, try your very hardest not to think of deep, soft, brown freshly baked chocolate cake. . .
Try your best not to think of the smell and the taste of that cake. . .

Just try to do this for about 30 seconds or so. . .

What did you notice?

Perhaps you deliberately thought about something other than cake (e.g., a tractor). . .
"I Must think about a tractor…because it's not cake."

But you had to think about cake, if only momentarily, to check that you weren't thinking about cake!

In fact, research says this type of thought suppression is impossible for more than a short while, and often causes the suppressed thought to show up more, and be more intense.

Appendix F continued
The Acceptance and Commitment Therapy (ACT)-based and Education-based protocols employed in Experiment 7

ACT-based protocol

So, what if, instead of fighting with these things, instead of using prejudices, stereotypes, and racist thoughts as evidence of problems with our selves and others, we could simply notice the process by which they come about, acknowledge that that is part of us, accept this, and **choose to behave in ways that do no reflect those thoughts or feelings whatsoever**?

It's like your mind is a computer. It constantly feeds you output, it never stops. "I'm hungry, what do I need to do today, I hope I see Mary today, I'm really ugly, When will I graduate, Should I work out later?" When you're right up close to the monitor, you can't see anything but the output. But if you stand back from the computer screen, you can notice that there is a difference between you and the output. You can watch it, see the process, call someone over and say, "Hey, look at that, that's interesting." Now imagine also that at this computer, various people sit down on it and program stuff in. Sometimes your parents are there, typing away, sometimes you friends, teachers, romantic partners, people in the media, songs you listen to, and so on and so on…The output comes from so many sources, we couldn't possibly track how or why we think what we do. Maybe if we stand back from the computer screen, we don't have to.

What are the numbers???

Think of your own prejudices, evaluations, and negative thoughts and emotions. Is it possible to make a little room for them by stepping back from the computer screen?

On the following page please summarise in your own words the main points contained in the foregoing information and exercises. Please feel free to add anything else that you feel is important or you want to say.

Appendix F continued

The Acceptance and Commitment Therapy (ACT)-based and Education-based protocols employed in Experiment 7

Education-based protocol

The gap in Black-White views appears to be growing. While 85% of Whites believe that Black children have the same educational opportunities as White children, only 52% of Blacks agree with that statement (Tilove, 2001).

A human rights commission successfully brought a civil lawsuit against two Klan groups for harassing and threatening African Americans who had moved into an all-White housing project (Baldauf & Johnson, 1998).

The Supreme Court agreed to decide whether a suit involving "environmental racism" could be brought in federal courts. Chester, Pennsylvania, is a town of 42,000 (65% of which is African American) and has five major waste facilities. The rest of the country, which is 91% white, has 500,000 people but has only two waste facilities (Watson, 1998).

The poverty rate for African Americans remains nearly three times higher than that of White Americans (33.1% versus 12.2%), and the unemployment rate twice as high (11% versus 5%; U.S. Bureau of the Census, 1995). Their disadvantaged status, as well as racism and poverty, contribute to the following statistics. About one third of African American men in their 20s are in jail, on probation, or on parole. This rate has increased by over one third during the past five years (Freeberg, 1995). Over 20% of black males are temporarily or permanently banned from voting in Texas, Florida, and Virginia because of felony convictions (Cose et al., 2000). The lifespan of African Americans is five to seven years shorter than that of White Americans (N.B. Anderson, 1995; Felton, Parson, Misener, & Oldaker, 1997).

Other health statistics are equally dismal. Twenty percent of African Americans have no health insurance (Giachello & Belgrave, 1997). About 40% of new AIDS cases in 1995 were African Americans (Talvi, 1997). Rates of hypertension (National Center for Health Statistics, 1996) and obesity (Kumanyika, 1993) are higher than those of the White population. Although hypertension has been thought to be primarily biological in African Americans, psychological factors may also be involved. African Americans exposed to videotaped or imaginal depictions of racism showed increases in heart rate and digital blood flow (D. R. Jones, Harrell, Morris-Prather, Thomas, Omowale, 1996). Systolic blood pressure also appears to be influenced by response to discrimination. African Americans who responded by accepting discrimination showed higher blood pressure than did those who challenged the situation (Krieger & Sidney, 1996). Medical researchers (Ayanian, Udvarhelyi, Gatsonis, Pashos, & Epstein, 1993; Harris, Andrews, & Elixhauser, 1997) have found that compared to White patients, African American patients were less

The Acceptance and Commitment Therapy (ACT)-based and Education-based protocols employed in Experiment 7

Education-based protocol

likely to undergo corrective surgeries or major therapeutic procedures. Since all had insurance coverage, the reason for the difference in care is unclear, although race-based decisions remain one possibility.

Please write down your immediate reactions to this information. . .

Now, please write down what you think it would be like to be a black person living in **Ireland**. . .

Although these statistics are grim, Ford (1997) pointed out that much of the literature is based on individuals of the lower social class who are on welfare or unemployed, and not enough is based on other segments of the African American population. This focus on one segment of African Americans masks the great diversity that exists among African Americans, who may vary greatly from one another on factors such as socioeconomic status, educational level, cultural identity, family structure, and reaction to racism. More than one third of African Americans are now middle-class or higher. They tend to be well-educated, married homeowners. In 1989, one out of seven African American families had an income of $50,000 or higher (Hildebrand, Phenice, Gray, & Hines, 1997). These are important distinctions. Many middle- and upper-class African Americans are receptive to the values of the dominant society, believe that advances can be made through hard work, feel that race has a relative rather than a pervasive influence in their lives, and embrace their heritage. However, they may feel bicultural stress. As Leanita McClain, the first

Appendix F continued

The Acceptance and Commitment Therapy (ACT)-based and Education-based protocols employed in Experiment 7

Education-based protocol

African American elected to the Board of Directors of the Chicago Tribune, reported,

I run a gauntlet between two worlds, and I am cursed and blessed by both. I travel, observe, and take part in both; I can also be used by both. I am a rope in a tug of war… Whites won't believe that I remain culturally different; Blacks won't believe that I remain culturally the same (Ford, 1997, p. 93).

However, middle-class African Americans are also exposed to feelings of guilt for having "made it," frustrations by the limitations imposed by the "glass ceiling," and feelings of isolation. Often, upward mobility can produce unintentional effects, as shown in the following case study.

A 14-year old African American boy, Joseph, came into counselling because of feelings of depression and anger. His parents are professionals and moved to a predominantly White suburb. Prior to the move, Joseph attended a mainly Black school, where he received many awards for academic achievement. Since his enrolment in a primarily White school, Joseph's performance has fallen. His teachers report him to be disruptive, off-task, and argumentative- particularly on issues of justice and minority groups. Joseph complains that they are insensitive and resents being the "expert" on Blacks. He has been asked why Blacks commit so many crimes and why they are so good in sports. He is also teased when he visits friends at his first school for speaking "proper English." Joseph has stolen money from his parents in an attempt to "buy" friendship with his white peers. (Ford, 1997)

The move from his predominantly Black school to one that is primarily White has exposed Joseph to issues of racism and the feelings of being different from both White Americans and African Americans. Issues of racial identity are also evident. It is also apparent that Joseph's parents are not aware of the racial issues that have surfaced with the change in schools. These factors need to be addressed with both the parents and Joseph.

Ford (1997) believes that middle- and upper-class African Americans may suffer a negative impact on mental health from issues such as believing a double standard exists (having to work twice as hard to succeed); feelings of isolation (being the only African American in the organisation); powerlessness (given responsibility only on tasks pertaining to minorities); being an "expert" or a "representative" on minority issues (e.g., African American professors might be asked to teach multicultural classes even if it is not their area of expertise); and "survival guilt" in moving to a higher class and

The Acceptance and Commitment Therapy (ACT)-based and Education-based protocols employed in Experiment 7

Education-based protocol

neighbourhood. Because of this, middle- and upper-class African Americans may occupy a marginal status in which they are not fully accepted by White Americans and are rejected by African Americans. Please write down any ideas you might have for overcoming current or potential race-related difficulties in the Irish context. . .

Please summarise in your own words the main points contained in the foregoing information and exercises. Please feel free to add anything else that you feel is important or you want to say. . .