Do You Really Know What You Believe?

Developing the Implicit Relational Assessment Procedure (IRAP) as a Direct Measure of Implicit Beliefs

Dermot Barnes-Holmes, Yvonne Barnes-Holmes, Patricia Power, Eilish Hayden, Rebecca Milne & Ian Stewart

Telling people that they do not know what they believe can be a risky business. Imagine, for example, asking an avid Chelsea football supporter which team they prefer, Chelsea or Liverpool? It would be considered very rude, and perhaps a little dangerous in certain contexts, to challenge the answer provided. Humans appear to have a unique ability to access the contents of their own minds and report with relative accuracy what they find there, and woe betide anyone who seriously challenges this notion, particularly in the context of socially or psychologically sensitive issues. Nevertheless, 20th century psychology has indeed challenged this widely held view of the human mind. For example, Nisbett and Wilson (1977) have shown the frailties of introspection; Wegner (2002) has shown how little control we possess over our own thoughts; and Greenwald and Banaji (1995) have highlighted the hidden or unconscious nature of social attitudes. In response to these and other findings, there has been increasing interest in developing procedures and measures that allow the modern researcher to tap into so-called implicit cognition. The current article is concerned with the theoretical and empirical development of one such methodology, the Implicit Relational Assessment Procedure (IRAP), which was designed to measure implicit beliefs and attitudes.

Implicit Attitudes

Implicit attitudes as defined by Greenwald and Banaji (1995) are “introspectively unidentified (or inaccurately identified) traces of past experience that mediate favourable or unfavourable feeling, thought, or action toward social objects” (p.8). The general argument is that individuals are often not aware of their implicit beliefs or attitudes, or how they can manifest as judgements or actions. There has been a recent explosion of research within psychology using procedures designed to evaluate such implicit attitudes, and the most well-established of these procedures is the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998).

The IAT was developed as a tool for assessing attitudes or beliefs that are easily hidden when explicit measures are employed such as questionnaires and open-ended interviews (De Houwer, 2002). Traditionally, researchers have used questionnaires and other measures of self-report to study and assess beliefs in an objective way, but these methods have a number of limitations (e.g. de Jong, Pasman, Kindt, & van den Hout, 2001; Dovidio & Fazio, 1992; Gemar, Segal, Sagrati, & Kennedy, 2001). One of the key problems is that self-report measures are susceptible to deception and self-presentation strategies, and thus results from such instruments might fail to reflect the implicit or hidden beliefs of a particular group of participants (e.g. de Jong, 2002; de Jong et al., 2001; Gemar et al., 2001; Teachman, Gregg, & Woody, 2001). Some researchers therefore have turned to alternative assessment methods, and the IAT is one such method.

In a seminal study, Greenwald et al. (1998) presented participants with names of flowers (e.g. ‘tulip’), names of insects (e.g. ‘beetle’), pleasant words (e.g. ‘love’), and unpleasant words (e.g. ‘ugly’). The researchers assumed that the concept “flower” and the attribute “pleasant” are likely associated in memory; similarly they also assumed that the concept “insect” and the attribute “unpleasant” will be associated. Insofar as these assumptions are correct, Greenwald et al. reasoned that responses should be faster when one response key is assigned to both “flower” and “pleasant”, and a second response key to “insect” and “unpleasant” (Task 1), than when one key is assigned to “flower” and “unpleasant”, and a second key to “insect” and “pleasant” (Task 2). The results of the experiment supported this prediction; response latencies were shorter in Task 1 than in Task 2 (Greenwald et al., 1998, Experiment 1).

Greenwald and colleagues have now published a range of studies showing that the IAT appears to reveal implicit attitudes, such as racial stereotypes, which participants typically deny when explicit measures such as questionnaires and open-ended interviews are used. For example, IAT research has shown that many participants who claimed not to hold racist attitudes nevertheless responded more quickly and more accurately when asked to categorise names typical of white persons with positive words and names typical of black persons with negative words, than when asked to categorise white with negative and black with positive (see Greenwald, Banaji, Rudman, Farnham, Nosek, & Mellott, 2002, for a review). This basic IAT effect has now been replicated many times (e.g. de Jong, 2002; de Jong et al., 2001; Gemar et al., 2001; Teachman et al., 2001), and it has become an increasingly popular method for assessing implicit attitudes and dysfunctional beliefs. The IAT has not been without its critics, however, and a number of limitations have been identified (e.g. De Houwer, 2002).

Some of the limitations inherent in the IAT have been addressed through the development of alternative methodologies, such as the Go/No-go Association Task (GNAT; Nosek & Banaji, 2001) and the Extrinsic Affective Simon Task (EAST; De Houwer, 2003). Nevertheless, one problem remains; the IAT, and its variants, focus on associations rather than relations among stimuli or events, and as such they provide only an indirect measure of beliefs. When discussing the IAT, De Houwer (2002) put it this way:

Greenwald et al. (1998) designed the IAT to assess the strength of associations between concepts in memory. One can argue that beliefs involve more than just associations between concepts. First, beliefs reflect qualified associations. For instance, the belief “I am a bad person” implies a special type of association between the concept...
“self” and the concept “bad”, namely a directional association which specifies that “bad” is a property or characteristic of “self”. IAT effects do not reflect the nature or directionality of an association between concepts, they can reflect only strength of association. Second, many beliefs involve several associations and several concepts. For instance, conditional beliefs such as “if I do not perform well on a task, then I am an inferior person” involve rather complex structures of qualified associations between several concepts. The IAT cannot be used to directly capture such complex conditional beliefs (also see de Jong et al., 2001, p. 111) . . In sum, the IAT does not provide a measure of beliefs, nor was it designed to do so. It can only provide an index of associations that are assumed to be involved in certain beliefs and thus indirect evidence for the presence of certain beliefs (pp. 117-118, emphases added).

A key limitation to the IAT, therefore, is that it cannot readily assess the nature or directionality of an association (i.e. a relation), and furthermore it cannot be used to assess a complex structure of directional associations (i.e. a relational network).

In one sense, this limitation to the IAT is not a genuine limitation if the core of human cognition is considered to be associative (see Greenwald, Nosek, Banaji, & Klauer, in press, for a relevant discussion). In other words, if one assumes that associative processes underpin attitudes, beliefs, and virtually all aspects of human higher cognition, then the IAT provides an ideal tool for tapping into these underlying associations. On balance, however, not all psychologists agree that human cognition is inherently associative. In fact, even within behavioural psychology, which historically was the main vanguard of associative theorising, a radically non-associative account of human language and cognition has begun to emerge.

Relational Frame Theory and the Implicit Relational Assessment Procedure

Relational Frame Theory (RFT) is a modern behavioural approach to human language and cognition (http://www.contextualpsychology.org), the central postulate of which is that higher-cognitive functioning is composed of relational acts (Hayes, Barnes-Holmes, & Roche, 2001). Relational Frame Theory assumes that stimulus relations, as opposed to associations per se, are central to understanding human psychological events (Hayes & Barnes-Holmes, 2004). This account has garnered increasing empirical support over the last decade, and thus it seems important to develop methodologies that aim to assess implicit stimulus relations, as opposed to associations.

One of the main methodologies developed from RFT to examine stimulus relations is the Relational Evaluation Procedure (REP; Barnes-Holmes, Healy, & Hayes, 2000; Hayes & Barnes, 1997). The REP allows subjects to evaluate, or report on, the stimulus relation that is presented on a given trial, and a number of studies have now been published that have employed the REP in the analysis of higher cognitive functioning in adult humans (Cullinan, Barnes, & Smeets, 1998; Cullinan, Barnes-Holmes, & Smeets, 2000, 2001; O’Hara, Barnes-Holmes, Roche, & Smeets, 2004; O’Hara, Pelaez, & Barnes-Holmes, 2005; Stewart, Barnes-Holmes, & Roche, 2002, 2004). Critically, the REP provided the methodological basis for the development of the Implicit Relational Assessment Procedure. In fact, the first IRAP was called the IREP, but the former acronym was soon adopted because it can be read as “I rap”, as in “I talk quickly”, which in essence is what the IRAP asks the participant to do (see below).

As De Houwer (2002) pointed out, the IAT assesses associations without providing a means of determining exactly how the elements are associated, or the continua along which the associations occur. In contrast, the IRAP involves presenting specific relational terms (e.g. SIMILAR, OPPOSITE, BETTER, WORSE) so that the properties of the relations among the relevant stimuli can be assessed. In short, the IRAP is a computer-based task that involves asking participants to respond quickly and accurately in ways that are either consistent or inconsistent with their pre-experimentally established verbal relations. The basic hypothesis is that average response latencies should be shorter across blocks of consistent relative to inconsistent trials. Or in other words, participants should respond more rapidly to relational tasks that reflect their current beliefs than to tasks that do not. In the next section of the article we will describe the IRAP in detail and present some preliminary data that have been gathered using this new methodology.

Three IRAP Studies: A Basic Test of the Methodology and the Assessment of Clinically and Socially Sensitive Attitudes

The very first IRAP study. The first IRAP experiment sought to determine if the basic hypothesis would be supported; would average response latencies be shorter across blocks of consistent relative to inconsistent trials? For the first experiment, 16 individuals, aged 18 to 30 years, agreed to participate. All participants were undergraduates at the National University of Ireland, Maynooth.

The experimental condition involved the following procedures. Each participant sat in front of a computer, which presented the instructions and the stimuli, and recorded all responses. In order to counterbalance the experimental sequence, participants were randomly assigned to one of two conditions; consistent-relations-first or inconsistent-relations-first. Before commencing the experiment, each participant was presented with a set of general instructions, which described the task and how to complete it. The instructions also stated that, “In some parts of the experiment the feedback from the computer may make sense to you, but in other parts it may not. This is part of the experiment.”

On each trial of the IRAP, four words appeared simultaneously on the screen. A sample stimulus, either “pleasant” or “unpleasant”, appeared at the top, with a single target word presented in the centre that was deemed either pleasant (e.g. love) or unpleasant (e.g. accident), and two relational terms, “opposite” and “similar”, which appeared at the bottom left- and right-hand corners of the screen (see Figure 1). All of the stimuli remained visible until the participant pressed one of the response keys. The task involved choosing one of the two relational terms (‘similar” or “opposite”); the left-right position of these terms alternated randomly across trials. To choose the term on the left participants pressed the “d” key and to choose the term on the right participants pressed the “k” key.

Choosing the relational term that was deemed correct for that block of trials removed all four stimuli from the screen for 400 ms before the next trial was presented. Choosing the relational term that was deemed incorrect for that block of trials produced a red X in the middle of the screen (immediately below the target word). The participant was not allowed to continue onto the next trial until he or she chose the correct relational term (for that block of trials). In effect, all key presses other than the correct one were non-functioning.
target word and the participant had to make the correct response by choosing “Similar” in order for the computer to progress to the next trial.

The second block of trials in the consistent-relations-first condition reinforced responses that were deemed relationally inconsistent. Given the sample, “Pleasant”, for example, and any of the following target words: abuse, crash, filth, murder, sickness or accident, choosing the relational term “Similar” progressed the computer program to the next trial. If, however, “opposite” was chosen the red ‘X’ appeared and the computer only progressed to the next trial when “Similar” was then chosen.

The six test blocks that followed the two practice blocks alternated between consistent and inconsistent (i.e. test block 1, consistent trials; test block 2, inconsistent trials; and so on until test block 6, inconsistent trials). Participants in the inconsistent-relations-first condition were exposed to the same procedure expect that the practice and test blocks alternated from inconsistent to consistent.

The primary datum was response latency defined as the time in milliseconds (ms) that elapsed between the onset of the trial and a correct response made by the participant. Consistent with common practice in IAT research, latencies over 3000ms were recorded as 3000ms, and latencies less than 300ms were recorded as 300ms (cf. Greenwald, Nosek, & Banaji., 2003). Two mean response latencies were then obtained for each participant; one calculated from across the three consistent test blocks and a second from the three inconsistent test blocks. The overall mean latencies calculated across the 16 participants are presented in Figure 2. As predicted, latencies were shorter for consistent relative to inconsistent trials. Statistical analyses indicated that this difference was significant. Additional analyses showed that this IRAP effect was obtained in both consistent-first and inconsistent-first conditions. Furthermore, separate analyses of the four different combinations of the two samples and two target-types (i.e. Pleasant/Pleasant, Pleasant/Unpleasant, Unpleasant/Pleasant, Unpleasant/Unpleasant) showed the IRAP effect for each combination.

In a second experiment, the response latency data (not shown here) were replicated with 11 naïve participants, and event related potentials (ERPs) were also recorded during exposure to the IRAP (see Staunton & Barnes-Holmes, 2004,

All participants were exposed to eight blocks of 24 trials; two practice blocks followed by six test blocks. Within each block, 12 target words were presented in a random order with the constraint that each word was presented twice, once in the presence of “Pleasant” and once in the presence of “Unpleasant” (see Table 1). The first block of trials in the consistent-relations-first condition reinforced responses that were deemed relationally consistent based on Greenwald et al.’s (1998) consistent and inconsistent categorisation of pleasant and unpleasant terms. Given the sample, “Pleasant”, for example, and any of the following target words, caress, freedom, health, love, peace or cheer; choosing the relational term “Similar” immediately progressed the computer program to the next trial (after 400 ms). If, however, “Opposite” was chosen, a red ‘X’ appeared below the target word and the participant had to make the correct response by choosing “Similar” in order for the computer to progress to the next trial.

The second block of trials in the consistent-relations-first condition reinforced responses that were deemed relationally inconsistent. Given the sample, “Pleasant”, for example, and any of the following target words: abuse, crash, filth, murder, sickness or accident, choosing the relational term “Similar” progressed the computer program to the next trial. If, however, “opposite” was chosen the red ‘X’ appeared and the computer only progressed to the next trial when “Similar” was then chosen.

The six test blocks that followed the two practice blocks alternated between consistent and inconsistent (i.e. test block 1, consistent trials; test block 2, inconsistent trials; and so on until test block 6, inconsistent trials). Participants in the inconsistent-relations-first condition were exposed to the same procedure expect that the practice and test blocks alternated from inconsistent to consistent.

The primary datum was response latency defined as the time in milliseconds (ms) that elapsed between the onset of the trial and a correct response made by the participant. Consistent with common practice in IAT research, latencies over 3000ms were recorded as 3000ms, and latencies less than 300ms were recorded as 300ms (cf. Greenwald, Nosek, & Banaji., 2003). Two mean response latencies were then obtained for each participant; one calculated from across the three consistent test blocks and a second from the three inconsistent test blocks. The overall mean latencies calculated across the 16 participants are presented in Figure 2. As predicted, latencies were shorter for consistent relative to inconsistent trials. Statistical analyses indicated that this difference was significant. Additional analyses showed that this IRAP effect was obtained in both consistent-first and inconsistent-first conditions. Furthermore, separate analyses of the four different combinations of the two samples and two target-types (i.e. Pleasant/Pleasant, Pleasant/Unpleasant, Unpleasant/Pleasant, Unpleasant/Unpleasant) showed the IRAP effect for each combination.

In a second experiment, the response latency data (not shown here) were replicated with 11 naïve participants, and event related potentials (ERPs) were also recorded during exposure to the IRAP (see Staunton & Barnes-Holmes, 2004,

Figure 1. Examples of the four IRAP trial-types. The sample (Pleasant or Unpleasant), target word (love, sickness, peace and filth, etc.), and response options (Similar and Opposite) appeared simultaneously on each trial. Arrows with superimposed text boxes indicate which responses were deemed consistent or inconsistent (boxes and arrows did not appear on screen). Selecting the consistent response option during a consistent block, or the inconsistent option during an inconsistent block, cleared the screen for 400 ms before the next trial was presented; if the inconsistent option was chosen during a consistent block, or the consistent option during an inconsistent block, a red X appeared on screen until the participant emitted the alternative response.

All participants were exposed to eight blocks of 24 trials; two practice blocks followed by six test blocks. Within each block, 12 target words were presented in a random order with the constraint that each word was presented twice, once in the presence of “Pleasant” and once in the presence of “Unpleasant” (see Table 1). The first block of trials in the consistent-relations-first condition reinforced responses that were deemed relationally consistent based on Greenwald et al.’s (1998) consistent and inconsistent categorisation of pleasant and unpleasant terms. Given the sample, “Pleasant”, for example, and any of the following target words, caress, freedom, health, love, peace or cheer; choosing the relational term “Similar” immediately progressed the computer program to the next trial (after 400 ms). If, however, “Opposite” was chosen, a red ‘X’ appeared below the target word and the participant had to make the correct response by choosing “Similar” in order for the computer to progress to the next trial.

The second block of trials in the consistent-relations-first condition reinforced responses that were deemed relationally inconsistent. Given the sample, “Pleasant”, for example, and any of the following target words: abuse, crash, filth, murder, sickness or accident, choosing the relational term “Similar” progressed the computer program to the next trial. If, however, “opposite” was chosen the red ‘X’ appeared and the computer only progressed to the next trial when “Similar” was then chosen.

The six test blocks that followed the two practice blocks alternated between consistent and inconsistent (i.e. test block 1, consistent trials; test block 2, inconsistent trials; and so on until test block 6, inconsistent trials). Participants in the inconsistent-relations-first condition were exposed to the same procedure expect that the practice and test blocks alternated from inconsistent to consistent.

The primary datum was response latency defined as the time in milliseconds (ms) that elapsed between the onset of the trial and a correct response made by the participant. Consistent with common practice in IAT research, latencies over 3000ms were recorded as 3000ms, and latencies less than 300ms were recorded as 300ms (cf. Greenwald, Nosek, & Banaji., 2003). Two mean response latencies were then obtained for each participant; one calculated from across the three consistent test blocks and a second from the three inconsistent test blocks. The overall mean latencies calculated across the 16 participants are presented in Figure 2. As predicted, latencies were shorter for consistent relative to inconsistent trials. Statistical analyses indicated that this difference was significant. Additional analyses showed that this IRAP effect was obtained in both consistent-first and inconsistent-first conditions. Furthermore, separate analyses of the four different combinations of the two samples and two target-types (i.e. Pleasant/Pleasant, Pleasant/Unpleasant, Unpleasant/Pleasant, Unpleasant/Unpleasant) showed the IRAP effect for each combination.

In a second experiment, the response latency data (not shown here) were replicated with 11 naïve participants, and event related potentials (ERPs) were also recorded during exposure to the IRAP (see Staunton & Barnes-Holmes, 2004,

Figure 1. Examples of the four IRAP trial-types. The sample (Pleasant or Unpleasant), target word (love, sickness, peace and filth, etc.), and response options (Similar and Opposite) appeared simultaneously on each trial. Arrows with superimposed text boxes indicate which responses were deemed consistent or inconsistent (boxes and arrows did not appear on screen). Selecting the consistent response option during a consistent block, or the inconsistent option during an inconsistent block, cleared the screen for 400 ms before the next trial was presented; if the inconsistent option was chosen during a consistent block, or the consistent option during an inconsistent block, a red X appeared on screen until the participant emitted the alternative response.

All participants were exposed to eight blocks of 24 trials; two practice blocks followed by six test blocks. Within each block, 12 target words were presented in a random order with the constraint that each word was presented twice, once in the presence of “Pleasant” and once in the presence of “Unpleasant” (see Table 1). The first block of trials in the consistent-relations-first condition reinforced responses that were deemed relationally consistent based on Greenwald et al.’s (1998) consistent and inconsistent categorisation of pleasant and unpleasant terms. Given the sample, “Pleasant”, for example, and any of the following target words, caress, freedom, health, love, peace or cheer; choosing the relational term “Similar” immediately progressed the computer program to the next trial (after 400 ms). If, however, “Opposite” was chosen, a red ‘X’ appeared below the target word and the participant had to make the correct response by choosing “Similar” in order for the computer to progress to the next trial.

The second block of trials in the consistent-relations-first condition reinforced responses that were deemed relationally inconsistent. Given the sample, “Pleasant”, for example, and any of the following target words: abuse, crash, filth, murder, sickness or accident, choosing the relational term “Similar” progressed the computer program to the next trial. If, however, “opposite” was chosen the red ‘X’ appeared and the computer only progressed to the next trial when “Similar” was then chosen.

The six test blocks that followed the two practice blocks alternated between consistent and inconsistent (i.e. test block 1, consistent trials; test block 2, inconsistent trials; and so on until test block 6, inconsistent trials). Participants in the inconsistent-relations-first condition were exposed to the same procedure expect that the practice and test blocks alternated from inconsistent to consistent.

The primary datum was response latency defined as the time in milliseconds (ms) that elapsed between the onset of the trial and a correct response made by the participant. Consistent with common practice in IAT research, latencies over 3000ms were recorded as 3000ms, and latencies less than 300ms were recorded as 300ms (cf. Greenwald, Nosek, & Banaji., 2003). Two mean response latencies were then obtained for each participant; one calculated from across the three consistent test blocks and a second from the three inconsistent test blocks. The overall mean latencies calculated across the 16 participants are presented in Figure 2. As predicted, latencies were shorter for consistent relative to inconsistent trials. Statistical analyses indicated that this difference was significant. Additional analyses showed that this IRAP effect was obtained in both consistent-first and inconsistent-first conditions. Furthermore, separate analyses of the four different combinations of the two samples and two target-types (i.e. Pleasant/Pleasant, Pleasant/Unpleasant, Unpleasant/Pleasant, Unpleasant/Unpleasant) showed the IRAP effect for each combination.

In a second experiment, the response latency data (not shown here) were replicated with 11 naïve participants, and event related potentials (ERPs) were also recorded during exposure to the IRAP (see Staunton & Barnes-Holmes, 2004,
The Irish Psychologist

previously, but different stimulus words were used (see Table 2). The participants was exposed to the same IRAP procedure as described who had never worked with autistic individuals. Each of the participants with under 6 months experience; and 16 participants the study: 16 participants with 6 months to 6 years experience; 12 working with a population diagnosed with autism participated in an implicit measure were used (cf. Hayes, et al., 2004). Although the first two experiments demonstrated that the IRAP produces significant differences in response latencies and ERP waveforms between consistent and inconsistent trials, these data were gathered with beliefs that could hardly be defined as socially sensitive or indeed implicit. The belief that sickness is unpleasant, for example, is unlikely to be an unconscious attitude or something that most of us would wish to conceal. A more adequate test of the IRAP, as a measure of implicit beliefs, would involve using it to examine socially sensitive attitudes that individuals are either unaware of or do not wish to express openly. Measuring attitudes to autism using the IRAP. The second IRAP study focused on attitudes to autism among three groups that differed in levels of experience of working with individuals with Autistic Spectrum Disorder. We speculated that professionals working with individuals with autism would tend to express more positive attitudes to the disorder, using explicit measures (i.e. questionnaires), than would professionals with no direct experience of autism, but this difference could well be absent if an implicit measure were used (cf. Hayes, et al., 2004).

Three groups of professionals varying in experience of working with a population diagnosed with autism participated in the study: 16 participants with 6 months to 6 years experience; 12 participants with under 6 months experience; and 16 participants who had never worked with autistic individuals. Each of the participants was exposed to the same IRAP procedure as described previously, but different stimulus words were used (see Table 2).

Figure 3. Grand average ERP waveforms from 11 participants for consistent (light lines) and inconsistent (dark lines) test blocks at electrode sites C3, P3 (left column), Cz, Pz (center column), C4, and P4 (right column). Stimuli were presented at 0ms.

for further information on ERPs as a neural measure of cognitive activity. Interestingly, the ERP waveforms differentiated between performance on the consistent and inconsistent trials. Specifically, greater negative deflections emerged for inconsistent relative to consistent trials between approximately 600 and 1000ms, and these differences proved to be statistically significant for most of the sites from which ERP signals were recorded (see Figure 3). Although the ERP data are preliminary, the fact that a reliable difference emerged between consistent and inconsistent trials with a relatively low n (compared to most ERP studies) indicates that additional research with the IRAP and neural measures may be worthwhile. It is also important to note that the same stimulus configurations are presented on consistent versus inconsistent trials on the IRAP (only the feedback differs), and thus the differential waveforms, which occurred before the feedback, appear to reflect the different psychological processes that occur just prior to emitting a well-established relational response versus a relatively weak one. Or, in other words, the intention to report a belief that one does not believe produces a more negative ERP waveform than the intention to report a belief that one actually believes. If these findings hold up in future research the implications could be wide-ranging indeed (see Knight, 2004).

Following exposure to the IRAP, each participant was asked to complete specific sections of the Challenging Behaviours Attribution Scale (CHABA; Hastings, 1997). Participants who had no direct experience of autism were asked to imagine that they did in fact work with this population, and to indicate how they thought they would feel or deal with challenging behaviour. Sections 3 and 4 of the CHABA were directly relevant to the current study. Section 3 is related to perceived self-efficacy in dealing with challenging behaviour. There are five questions with a score of 1 to 7 with 1 signifying low self-efficacy and 7 signifying high efficacy. Section 4 provides a list of emotions that may have been experienced when dealing with individuals who display challenging behaviours. Answers are rated on a scale of 0 = no, never; 1 = yes, but infrequently; 2 = yes, frequently; and 3 = yes, very frequently. The emotional terms are separated into two categories, depression and anger (10 items), or fear and anxiety (5 items). The maximum score, therefore, for depression and anger is 30 and for fear and anxiety it is 15.

In a second questionnaire, the Attitudes to Autism Scale (AAS), participants were given a list of statements in relation to children with autism or normally developing children. The AAS was developed specifically for the study based on contributions from a range of autism and non-autism professionals who were asked to provide statements that they considered to be commonly applied to autism. Participants were instructed to circle the number they thought they would feel or deal with challenging behaviour. There are five questions (e.g. “Children with autism are more difficult than normally developing children”), the responses to which were scored on a scale ranging from 1 to 5 (from “Very True” to “Very Untrue”). A possible score ranging from 10 to 50 was thus obtained for

Each block of 24 trials involved presenting each of the two sample phrases, “Autistic Spectrum Disorder” and “Normally Developing”, in a random order with the constraint that each sample was presented 12 times. The 12 target words were also presented in a random order with the constraint that each one appeared once with the two sample phrases. These target words were pairs of generic positive and negative terms, not specific to either Autistic Spectrum Disorder or normally developing individuals. For illustrative purposes, consider the following trial: “Autistic Spectrum Disorder” is presented as a sample with the target word “Difficult”. During consistent blocks choosing “Similar” would progress the participant to the next trial, but choosing “Opposite” would not (the red X would appear); the reverse applied during inconsistent blocks.
each participant, with 10 to 25 signifying a negative view of autism, 25 to 35 signifying a neutral view of autism in relation to normally developing children, and 35 to 50 signifying a positive view of autism.

Results from the study showed that the mean adjusted response latencies from the IRAP were significantly longer on inconsistent tasks relative to consistent tasks for each of the three groups (see Figure 4). Although the Years-Experience Group responded more rapidly overall on both tasks than the other two groups, statistical analyses found no significant difference among the groups in terms of the relative difference in latencies between consistent and inconsistent tasks. Additional analyses showed that the IRAP effect was obtained in both consistent-first and inconsistent-first conditions for all three groups. Furthermore, separate analyses of the four different combinations of the two samples and two target-types (i.e. Autism/Negative, Autism/Positive, Normal/Negative, Normal/Positive) showed the same response pattern for each group. Specifically, they each responded more rapidly to ‘Autism-Negative-Similar’ than to ‘Autism-Negative-Opposite’, and more rapidly to ‘Autism-Positive-Opposite’ than to ‘Autism-Positive-Similar’; each group also responded more rapidly to ‘Normal-Positive-Similar’ than to ‘Normal-Positive-Opposite’, and more rapidly to ‘Normal-Negative-Opposite’ than to ‘Normal-Negative-Similar’.

In contrast to the IRAP, however, the questionnaires did yield significant differences across the three groups (see Figure 5). Specifically, the Years- and Months-Experience groups showed significantly more positive attitudes to autism (Figure 5, top-left panel), and significantly lower levels of emotional reactions of depression/anger (bottom-left panel) and fear/anxiety (bottom-right panel) than the No-Experience group. Finally, the Years-Experience Group reported significantly greater self-efficacy than the other two groups (top-right panel), with no difference between the latter two. Overall, therefore, the current data appear to show the classic separation between implicit and explicit measures of attitudes, in that the IRAP showed no significant effect across groups, but the questionnaires did.

This separation between the implicit and explicit measures could indicate that the IRAP simply failed to capture the different attitudes that were expressed explicitly by the experienced and non-experienced participants. Certainly, many of the questionnaire items, particularly on the CHABA, were not directly relevant to the content of the IRAP tasks, and thus the observed separation between the measures may be expected. On balance, the content of the AAS was designed to overlap to some degree with the content of the IRAP and yet the two measures also showed a different effect. Perhaps, therefore, the current results indicate that autism professionals are unwilling or unable to express (perhaps even to themselves) negative views of children with a diagnosis of autism.

This latter conclusion is consistent with recent research on stigmatisation and burn-out among caring professionals. For example, data reported by Hayes et al. (2004) suggest that caring professionals may suppress or avoid negative thoughts and feelings pertaining to a client population, thus producing various forms of psychological stress, which in turn may lead to burnout and high staff turnover rates. The current IRAP data, although preliminary, could be seen therefore as highlighting the need for professional training programmes aimed at dealing with stigmatisation and burn-out among those working with individuals with autism in an open and accepting way (see Hayes, et al., 2004, for an example of such a programme for substance-abuse counsellors). In any case, the current results do suggest, if only tentatively, that the IRAP provides a measure of implicit beliefs. The final study outlined subsequently offers additional evidence to support this conclusion. Furthermore, the study illustrates how the IRAP may be used to assess a network of interconnected beliefs or relative preferences pertaining to different social groups.

Measuring relative likeability for different social groups using the IRAP. In this third study we assumed that individuals would generally show a preference for their own group over that of others (Tajfel, 1982). Furthermore, we assumed that relative preferences for other groups would be determined, in part, by the extent to which the other groups are perceived to be more or less similar to one’s own (see Cota, Evans, Dion, Kilik, & Longman, 1995; Goethals & Darley, 1977). Pilot work with Irish students indicated that when given the choice among Scottish, American and African, Scottish is typically ranked as most similar to Irish, followed by American and then African. Based on this ranking one might assume that Irish participants will show a preference for Irish over Scottish, Scottish over American, and American over African.

To determine if this pattern of preferences would be obtained using an explicit measure, 12 Irish participants were asked to rate on four separate likert scales the extent to which they found Irish, Scottish, American and African people likeable or unlikeable. The scales ranged from –6 indicating extremely unlikeable to +6 indicating extremely likeable. The results obtained from the four scales are presented in Figure 6. Statistical analyses indicated no difference in ratings between Irish and Scottish, but these two groups were rated as significantly more likeable than both American and African, with African being rated as significantly more likeable than American. The explicit measure thus produced responses that did not accord with our predictions – Irish was not ranked as more likeable than Scottish, and African was ranked as more likeable than American, rather than vice versa.

The IRAP was then employed to determine if the participants’ implicit relative preferences for the four groups would differ from their explicit responses to the likert scales. An IRAP procedure similar to that described previously was used, but the stimuli were different (see Table 3). The two sample stimuli were “More Likeable” and “Less Likeable”; the target stimuli were the word-pairs “Irish Scottish”, “Scottish American” and “American African”, and their reversed counterparts; and the
response options were “True” and “False”. Consistent with the previous IRAP studies, each block of 24 trials involved presenting the two samples in a random order with the constraint that each sample was presented 12 times. The three target word-pairs, and their reversed counterparts, were also presented in a random order with the constraint that each one appeared twice with the two samples. On one trial, for example, “More Likeable” was presented as a sample with the target word-pair “Irish Scottish” (participants were instructed to read this configuration as “Irish more likeable than Scottish?”). Choosing “True” during consistent blocks progressed the participant to the next trial, but choosing “False” did not (the red X appeared); the reverse applied during inconsistent blocks. Trials were deemed consistent or inconsistent in accordance with our prediction that Irish participants will show a greater liking for Irish over Scottish, Scottish over American and American over African. Thus, for example, response latencies should be shorter when Irish participants have to respond “True” to “Irish more likeable than Scottish” and “Scottish less likeable than Irish” than when they have to respond “False” on these trials; similarly, response latencies should be shorter when participants have to respond “False”, rather than “True”, to “Scottish more likeable than Irish” and “Irish less likeable than Scottish”.

Results from the IRAP showed that the mean adjusted response latencies were significantly longer on inconsistent tasks relative to consistent tasks for each of the three target-pairs (see Figure 7). In general terms, therefore, the participants responded more rapidly across tasks that confirmed rather than denied that; (i) Irish is more likeable than Scottish, (ii) Scottish is more likeable than American, and (iii) American is more likeable than African (additional analyses showed that these IRAP effects were obtained in both consistent-first and inconsistent-first conditions). These implicit measures of likeability therefore diverged from the pattern of preferences obtained using the explicit likert scales. First, the large IRAP effect for the Irish-Scottish pairing contrasts rather starkly with the explicit measure, which produced virtually no difference in likeability for these two groups. Second, the clear IRAP effect for the American-African pairing is in the opposite direction to the explicit measure; the former indicates a preference for American whereas the latter indicates a preference for African. Overall, therefore, the IRAP data were consistent with our original prediction that Irish participants will show a preference for Irish over Scottish, Scottish over American, and American over African.

At the present time we can only speculate as to why these clear differences emerged between the explicit and implicit measures. It does seem likely, however, that responses to the likert scales were influenced to some degree by variables that extended beyond the relational or contextual control of likeability per se. For example, some participants may have rated “African” more likeable than “American” because doing so was perceived to be “politically correct” (i.e. one does not want to be seen to favour a super-power over poor and oppressed nations). In fact, one of the participants during debriefing actually offered this explanation for his divergent performances across the two measures. Furthermore, because Irish and Scottish are perceived to be so similar it is possible that some participants did not rate the former as more likeable than the latter because doing so might be seen as excessively parochial or patriotic. When exposed to the IRAP, however, these and other sources of additional contextual control over responding would be much reduced because participants do not have sufficient time, on a trial-by-trial basis, to engage in the additional and sometimes complex relational activity that serves to generate a “carefully considered” answer. In short, therefore, the IRAP effect may diverge from an explicit measure because a rapid-response requirement significantly reduces the sources of contextual control that may “contaminate” the targeted relational responding (cf. Wilson, Lindsey, & Schooler, 2000).

If the foregoing interpretation of the IRAP effect turns out to be correct, or at least partially so, then it certainly seems to provide a direct measure of implicit beliefs. In drawing this conclusion, however, the terms “direct” and “implicit” require some scrutiny. In the closing section of the current article, therefore, we will examine these concepts and attempt to provide a brief theoretical analysis of the IRAP effect.
The IRAP as a Direct Measure of Implicit Beliefs

*Is the IRAP a direct measure?* Most procedures that were designed to measure implicit cognitions do not ask participants to report the extent to which they hold a particular belief or attitude. Rather, the belief is determined indirectly by assessing its effect on a particular performance. When exposed to an IAT for assessing attitudes to old age, for example, participants are not asked to report their beliefs about old people directly. Instead, the IAT might involve categorising stimuli associated with old and young people with names of insects and flowers. If response latencies are shorter on tasks that involve categorising old with insects and young with flowers, relative to old with flowers and young with insects, a negative attitude towards old age may be inferred from the test (cf. Dasgupta & Greenwald, 2001). In contrast, the traditional approach to measuring attitudes simply involves asking participants to express the attitudes or thoughts that are measured, and most other implicit tests, appears to be a direct measure. That is, each of the IRAPs described in the current article involved contextual cues and little else provide a rough RFT analogue of a *behavioural* explanation rooted firmly in RFT concepts. At this point, therefore, it seems more appropriate for us to provide a brief RFT explanation for the IRAP effect, and in so doing highlight possible points of contact between behavioural and cognitive approaches to implicit beliefs or attitudes. Let us now turn to this issue.

On each IRAP trial, contextual cues are presented that specify particular relational (e.g. same, opposite, more/less) and functional dimensions (e.g. pleasant/unpleasant, likeability). We would argue that relational responses that are controlled primarily by these cues and little else provide a rough RFT analogue of so-called implicit cognition. From this perspective, the IRAP may give us a measure of implicit beliefs because, as explained previously, contaminating sources of contextual control will be reduced by the procedure’s rapid response requirement. If this interpretation is correct, a possible explanation for the IRAP effect might be as follows.

One of the concerns about implicit cognition is that are assumed to be *indirectly* related to the belief under study such an IRAP would seem to provide a relatively indirect measure. Imagine an IRAP, for example, that asked participants to confirm and deny, across consistent and inconsistent blocks, respectively, that old people are similar to insects and opposite to flowers, and young people are similar to flowers and opposite to insects. Although this would be a rather bizarre task, more rapid responding across consistent versus inconsistent blocks could be used to infer, as one would from the IAT using the same categories, a negative attitude towards old relative to young people.

In summary, the IRAP may be considered a relatively direct measure when participants are asked to respond to questions concerning the actual attitude or belief under study; if the questions are not directly related to the targeted cognition (as is typical with the IAT), then the measure seems more indirect. In either case, however, it could be argued that the IRAP constitutes a less direct measure than a traditional questionnaire because differential response latency, rather than a self-expressed verbal response or rating, is taken as the primary datum.

*Is the IRAP effect a measure of implicit beliefs?* Providing an adequate answer to this question is problematic for at least two reasons. First, the concept of implicit cognitions remains a topic of hot debate within the literature, and widespread agreement concerning basic definitions has not yet emerged (De Houwer, in press). For example, some researchers have emphasised the “unconscious” aspect of implicit attitudes (Greenwald & Banaji, 1995), whereas others have focused on their unknown origin, habit-like qualities and their impact on uncontrolled responses (Wilson, et al., 2000). And still others have focused on the automatic (versus deliberative) properties of implicit cognitions (De Houwer; Fazio & Olson, 2003).

The second problem is that the IRAP emerged directly from a behavioural theory of human language and cognition (RFT), and the concept “implicit” is not a technical or explanatory term within this theoretical framework. From an RFT perspective, therefore, the IRAP effect, and implicit cognition itself, requires a *behavioural* explanation rooted firmly in RFT concepts. At this point, therefore, it seems more appropriate for us to provide a brief RFT explanation for the IRAP effect, and in so doing highlight possible points of contact between behavioural and cognitive approaches to implicit beliefs or attitudes. Let us now turn to this issue.

In fact, it is worth noting that the IRAP may fall at different points along the continuum depending upon the stimuli that are used for a given experiment. If the tasks require participants to confirm or deny beliefs that the experimenter is aiming to assess, then the IRAP seems more direct than indirect. If, however, participants were asked to confirm or deny beliefs that are assumed to be *indirectly* related to the belief under study such an IRAP would seem to provide a relatively indirect measure. Imagine an IRAP, for example, that asked participants to confirm and deny, across consistent and inconsistent blocks, respectively, that old people are similar to insects and opposite to flowers, and young people are similar to flowers and opposite to insects. Although this would be a rather bizarre task, more rapid responding across consistent versus inconsistent blocks could be used to infer, as one would from the IAT using the same categories, a negative attitude towards old relative to young people.

In summary, the IRAP may be considered a relatively direct measure when participants are asked to respond to questions concerning the actual attitude or belief under study; if the questions are not directly related to the targeted cognition (as is typical with the IAT), then the measure seems more indirect. In either case, however, it could be argued that the IRAP constitutes a less direct measure than a traditional questionnaire because differential response latency, rather than a self-expressed verbal response or rating, is taken as the primary datum.
inconsistent block will tend to possess the wrong function. In the latter case, therefore, additional private responding will be needed to establish the correct function (the ERP data presented earlier provide some evidence to support this interpretation). More informally, during consistent trials participants can simply do the first thing that comes to mind, but during inconsistent trials they must suppress this tendency before choosing a response key. The need to “monitor” and “self-correct” before pressing the correct key during inconsistent trials requires time. Consequently, across multiple trials, which control for extraneous variables, such as fatigue and random distraction effects, the average latency for inconsistent blocks will be longer than for consistent blocks.

If the foregoing interpretation of the IRAP effect is correct it could explain why implicit beliefs or attitudes are difficult to control. Specifically, implicit cognitions reflect the most immediate and often private or incipient relational response that occurs for an individual. Subsequent and more complex relational responding may fail to cohere with the first private response, but when the behavioural system is put under pressure, as with the IRAP, the impact of the first or most probable response is difficult to hide (cf. Wilson, et al., 2000). This type of analysis could also be relevant to the so-called unconscious nature of implicit attitudes. If we accept that an implicit attitude constitutes a probabilistic and often private and fleeting relational response it may indeed be difficult for an individual to discriminate this response from the subsequent pattern of extended and public relational responding that also occurs. When Irish participants were asked in our study, for example, to rate their likeability of Americans, any immediate and fleeting positive response may have been “drowned out”, within a matter of milliseconds, by the production of a more complex and extended relational network in which Americans were co-ordinated with negatively valenced descriptors. If this type of process typically occurs, participants may well be surprised when they produce a performance on the IRAP that does not co-ordinate with their more complex relational responses. In short, relational responding is so continuous, dynamic and fast-flowing that discriminating each individual response, and its control over the statement of a particular belief, is far from a simple matter for most of us, and thus some responses might be described as relatively “unconscious”.

As an aside, preliminary findings from our research suggest that exposure to the IRAP may in fact serve to increase the salience of previously “unconscious” beliefs. Specifically, when participants are asked to complete an explicit measure before and after an IRAP performance, the second explicit measure sometimes moves in the direction of the IRAP effect. Of course, this raises the thorny question as to whether or not the original attitude or belief was genuinely unconscious, but addressing this issue would take us well beyond the scope of the current article (but see De Houwer, in press).

**Conclusion**

We fully recognise that the data we have presented are quite preliminary and the foregoing interpretation of the IRAP effect is rather speculative. Indeed, a considerable body of empirical research will be needed to assess the reliability and validity of the IRAP, and also the RFT interpretation of its effect. Nonetheless, current theory and data are consistent with the general thrust of our arguments, and a growing number of researchers across Ireland, North America, England and Sweden are currently engaged in IRAP research. In closing, therefore, we invite the interested reader to visit the IRAP website (http://www.nuim.ie/academic/psychology/IRAP/IRAP%20page1.shtml), download the free software and material, and see if you really do know what you believe.

**References**


Contributions Sought for The Irish Psychologist

Contributions from members continue to be sought, particularly in the form of review-type articles. Suggested special issues are also welcomed. Without your contributions, the development of the IP as a forum for review, debate and discussion of key issues in psychology cannot be sustained.

The Irish Psychologist is produced by The Psychological Society of Ireland and is intended to serve as a means of disseminating information as well as a forum for discussion of issues relevant to psychologists in Ireland. It is aimed to appeal to academic psychologists and practitioners and to keep members informed of events and developments concerning psychology in Ireland.

The editor welcomes
• articles
• reports of events
• letters
• news of members
• book reviews
• research reports
• diary of events entries (now free of charge)
and any other material relevant to psychologists in Ireland.

In contributing articles, authors should provide a broad overview of a particular issue or debate matters of professional relevance. A non-specialist audience should be assumed. Prospective authors may contact the editor in advance of submission at the email address below.

News of members, or of general interest to psychologists in Ireland, is welcome, as are reviews of events.

Material should be sent (preferably as electronic copy) to: The Editor, The Irish Psychologist, The Psychological Society of Ireland, CX House, 2A Corn Exchange Place, Dublin 2.

email irishpsychologist@psihq.ie

Vol.32 No.7 p 177


