Title: Assessing the Implicit Beliefs of Sexual Offenders using the Implicit Relational Assessment Procedure: A First Study

Authors: David L Dawson¹

Dermot Barnes-Holmes ²

D. Mark Gresswell ³

Aidan J P Hart ³

Nick J Gore⁴

Running Title: IRAP: The assessment of implicit sexual beliefs

Key Words: Implicit beliefs; cognitive distortions; sexual offending; Implicit Relational Assessment Procedure

¹ Corresponding author. Department of Clinical Forensic Psychology, Peter Hodgkinson Centre, Lincoln County Hospital, Lincoln. LN2 3UB. Telephone: +44 (0)1522 573243 Email: david.dawson18@gmail.com

² National University of Ireland, Maynooth

³ University of Lincoln, UK.

⁴ University of Kent, UK.
IRAP: The assessment of implicit sexual beliefs

Abstract

Researchers have proposed that the cognitive distortions of sexual offenders are underpinned by a number of implicit cognitive processes termed implicit theories (Ward & Keenan, 1999; Ward, 2000). Although the implicit theory hypothesis has provided a framework for understanding the motivations of sexual offenders, it has until recently received little empirical support due to broader limitations with standard forensic assessment procedures. The current research therefore aimed to determine whether a new assessment methodology, the Implicit Relational Assessment Procedure (IRAP), could provide further evidence for Ward and Keenan’s (1999) children as sexual beings implicit theory. The results indicated that the IRAP was more effective at identifying core differences between sexual offenders against children and non-offenders than a cognitive distortion questionnaire. Furthermore, although both groups demonstrated an overall IRAP response bias towards adults as sexual and children as non-sexual, this was significantly impaired in the sexual offender group, providing support for the implicit theory hypothesis. The findings of the study are discussed in relation to previous research, and recommendations for the further development of implicit measurement methodologies are made.
Introduction

In an influential paper, Ward and Keenan (1999) proposed that cognitive distortions, the beliefs sexual offenders express that function to justify or condone sexually abusive behaviour (Abel, Becker & Cunningham-Rathner, 1984; Calder, 1999; Ward, Hudson, Johnston & Marshall, 1997) are a product of underlying cognitive processes termed implicit theories. Specifically, Ward and Keenan identified five such implicit theories, which they argued account for the majority of cognitive distortions articulated by sexual offenders against children. Briefly, these theories refer to beliefs that (a) children are sexual beings who desire and actively pursue sexual contact; (b) the offender is entitled to behave how he or she wants to towards other people; (c) the world is a dangerous place and one should fight back or else seek out non-threatening sexual partners (e.g. children); (d) sexual behaviour is uncontrollable and therefore individuals that engage in sexually abusive behaviour are not completely to blame; and (e) there are degrees of harm, and sexual behaviour with a child is only harmful if the abuse includes force or threat.

Such implicit theories, it is argued, not only contribute to the generation of cognitive distortions, but also influence attention, social information processing and overt behaviour, and lead an individual to make potentially erroneous assumptions about the motivations and internal states of others. For example, an individual with an implicit theory that children are sexual beings (Ward & Keenan, 1999; Ward, 2000) may be more likely to notice a child sitting with her legs apart, to interpret the child’s behaviour as sexually motivated, and thus target the child for an offence. By contrast, an individual with an implicit theory that children are lacking in self-awareness may be less likely to
notice the child, and more likely to interpret the behaviour as innocuous and innocent. Consequently, implicit theories are hypothesised to play a key role in the offending process (Ward & Keenan, 1999; Ward, 2000) and are thus potentially important targets for Sexual Offender Treatment Programmes (SOTPs). Accurately assessing the cognitive distortions and implicit theories of sexual offenders has, however, proved difficult.

Typically, cognitive distortions and implicit theories are assessed through the use of questionnaires or clinical interviews (e.g. Beech, Fisher, & Beckett, 1998; Brown, 2005; Fisher, Beech, & Browne, 1999; Polaschek & Ward, 2002), but such methods have serious shortcomings when applied to offender populations. In particular, the transparency of explicit measures such as interviews and questionnaires render them vulnerable to social desirability biases and manipulation (e.g. Gannon, 2006; Roche, Ruiz, O’Riordan & Hand, 2005; Tierney & McCabe, 2001; Ward, Hudson, Johnston, & Marshall, 1995). Furthermore, a number of authors have suggested that implicit beliefs are difficult to identify through introspection, even for well-motivated individuals (Greenwald & Banaji, 1995; Mihailides, Devilly & Ward, 2004). Due to these difficulties, research aimed at differentiating sexual offenders from non-offenders using explicit measures of cognitive distortions have provided equivocal results (e.g. Tierney & McCabe, 2001), hindering clinical practice and the empirical development of models such as the implicit theory hypothesis.

Faced with similar difficulties, researchers within the field of experimental social psychology have developed methodologies to assess and measure implicit cognition. Differences in response latencies, identified using methodologies such as the Implicit Association Test (IAT; Greenwald, McGhee & Schwartz, 1998), have proved to be more
effective than questionnaires at revealing socially sensitive beliefs that an individual may wish to conceal, such as racial stereotyping and prejudice (Dasgupta, McGhee, Greenwald & Banaji, 2000; Greenwald et al., 2002; Rudman, Greenwald, Mellott & Schwartz, 1999). Furthermore, previous research has found that the IAT is difficult to “fake” (Fiedler & Bluemke, 2005), is arguably predictive of overt behaviour (McConnel & Leibold, 2001), and may be able to map clinical change (Teachman & Woody, 2003).

The core assumption underpinning the IAT is that individuals should respond quickly when asked to emit a similar response for two concepts that are closely associated in memory, but should respond more slowly when the two concepts are less closely associated. In an IAT study on racial bias, for example, participants in one task are asked to categorize names typical of white people with positive words and names typical of black people with negative words, but in a second task these categorizations are reversed (white-negative and black-positive). A white in-group racial bias is observed when white participants respond faster on white-positive and black-negative trials than on the reversed counterparts (e.g. Greenwald et al., 1998).

At this stage, it is worth noting that although the IAT is now the most popular test of implicit attitudes, a widely recognized weakness is that it provides only a measure of relative associative strength, which creates a lack of precision in determining the nature of the attitudes under study (see De Houwer, 2003). If, for example, participants respond more quickly on white-positive and black-negative trials than on the reversed counterparts (i.e., white-negative and black-positive), three different interpretations are possible. Participants may (i) like white and dislike black people, or (ii) like both white and black people, but the former is liked more than the latter, or (iii) dislike white and
IRAP: The assessment of implicit sexual beliefs

black people, but the latter is disliked more than the former. In order to measure implicit attitudes to individual stimuli, therefore, a non-relativistic measure is required, and a number of researchers have attempted to develop such tests, including, for instance, the Extrinsic Affective Simon Test (EAST; De Houwer, 2003) and the Go/No-Go IAT (GNAT; Nosek & Banaji, 2001).

A number of authors have advocated the use of implicit methodologies within sexual offending research (e.g. Kalmus & Beech, 2005; Ward, Hudson, Johnston, & Marshall, 1997). In the first IAT study in this area, Mihailides et al. (2004) sought to investigate three of Ward and Keenan’s (1999) implicit theories: children as sexual beings, uncontrollability, and entitlement. The authors found evidence for all three theories amongst child sex offenders, with particularly strong support for the children as sexual beings theory (e.g. child sexual offenders responded faster to word-pairs such as “child” and “lust” than a non-offending control group).

A similar study used an IAT methodology to determine if child sex offenders held stronger implicit associations between sex and child-related words than an offender control group (Gray, Brown, MacCulloch, Smith, & Snowden, 2005). As predicted, the child sex offenders produced significantly shorter mean response latencies than the control group during the trials in which sex-related words (e.g. climax, cock, lust) shared the same response key with child-related words (e.g. innocent, school, kid). Furthermore, the authors reported that the IAT had some predictive validity, correctly identifying 78% of the sexual offenders against children, albeit at the expense of misidentifying 42% of the control participants as sexual offenders. The authors concluded that the IAT can
“identify a core cognitive abnormality that may underpin some paedophilic deviant sexual behaviour” (Gray et al., 2005; p. 304).

A third study employed a lexical decision making task to assess for an implicit association between power and sex amongst child sex offenders (Kamphuis, De Ruiter, Janssen, & Spiering, 2005). The results indicated that child sex offenders responded faster to sex and power-related words when subliminally primed with sex and power-related cues than controls. The authors concluded that the findings support the hypothesis that there is an implicit link between sex and power amongst men who molest children, a finding consistent with Ward and Keenan’s (1999) entitlement and dangerous world implicit theories.

In a fourth and final study, Nunes, Firestone and Baldwin (2007) used an IAT methodology to assess whether child sex offenders held significantly different implicit associations than non-sexual offenders in terms of beliefs about ‘self’ and children (relative to adults) within the domains of evaluation (e.g. pleasant or unpleasant), social power and sexual attractiveness. Similar to the above findings, the authors report that the child sex offenders produced significantly shorter mean response latencies than the control group during the trials in which sex-related words shared the same response key with child-related words (the ‘sexy child’ IAT), although no other significant IAT differences were found between the two groups within the other assessed domains. Interestingly, however, the authors found that the ‘sexy child’ IAT also correlated significantly with increased risk of sexual recidivism as measured by the Static-99 (Hanson & Thornton), an established actuarial measure of sexual offence recidivism.
Although the studies outlined above offer support for Ward and Keenan’s theory and circumvent some of the difficulties associated with explicit assessment methods, a cautious approach is required when interpreting the findings. Specifically, the IAT (and similar methodologies) provide a relatively indirect measure of implicit beliefs. In noting the indirect nature of the IAT, for example, De Houwer (2002) argued that:

. . . [it] 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).

Thus, if a methodology such as the IAT indicates that ‘Child’ and sexual words are strongly associated, it is then inferred that such implicit associations underlie a belief that children are sexual. Although such an inference seems reasonable, it seems prudent to develop additional methodologies that aim to provide relatively direct measures of implicit cognition. One such methodology has recently been offered: the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes, Barnes-Holmes, Power, Hayden, Milne, & Stewart, 2006, Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, in press; McKenna, Barnes-Holmes, Barnes-Holmes, & Stewart, 2007).

The IRAP was developed from a modern behavioural theory of human language and cognition known as Relational Frame Theory (RFT; Hayes, Barnes-Holmes & Roche, 2001). According to RFT, the core elements of human cognition are relational acts. Unlike the IAT (and other purely associative methods), each trial of the IRAP asks participants to confirm or deny a specific belief by responding, for example, to a simple

---

1 A recent study of child molesters’ beliefs failed to detect implicit theories (Gannon, Wright, Beech, & Williams, 2006), but unlike the studies discussed above the researchers used a non-response-latency based measure involving ambiguous vignettes and free-recall.
statement, such as “Child – Sexual?” with “True” or “False”. The former response would be required on blocks of IRAP trials that are deemed to be inconsistent with prevailing social norms and the latter response would be required for trials that are deemed to be consistent with social norms.

The basic assumption is that most IRAP trials produce an immediate relational response before the participant actually presses a response key (the participant may or may not be consciously aware of this response). It is further assumed that the probability of the initial response will be determined by the verbal and non-verbal history of the participant and current contextual variables. By definition, the most probable immediate response will be emitted first most often, and thus during a consistent IRAP trial that response will tend to match the correct key; during an inconsistent trial, however, the probability of such a match will be reduced. Thus, across multiple trials, the average latency for inconsistent blocks will be longer than for consistent blocks. In short, the IRAP effect is based on immediate and perhaps unconscious relational responding, which is made apparent to the researcher when the behavioral system is put under pressure to respond quickly and accurately. The extent of the observed difference between consistent and inconsistent trials is assumed to provide an index of the strength of the specific belief being assessed.

This basic IRAP effect has now been replicated across a small number of studies, which have shown that the IRAP; (i) compares well with the IAT as a measure of individual differences (Barnes-Holmes, Murtagh, Barnes-Holmes, & Stewart, in press; Barnes-Holmes, Waldron, Barnes-Holmes, & Stewart, in press), (ii) is not easily faked (McKenna, Barnes-Holmes, Barnes-Holmes, & Stewart, 2007), (iii) may be used as a
An important feature of the IRAP is that it consists of four trial-types, which in principle permits the assessment of four specific beliefs (i.e., unlike the IAT, the IRAP is a non-relativistic measure). In the research outlined subsequently, the IRAP involved presenting one of two category labels, “Adult” or “Child”, and two sets of target stimuli; one set was composed of sexual words (e.g., “erotic”) and the other set of non-sexual words (e.g., “innocent”). The IRAP involved presenting each set of target stimuli with both labels, and thus four different trial-types were created (see Figure 1). This four-element structure to the IRAP thus allowed us to determine if the response biases of child sex offenders differed from non-offenders across one or more trial-types.

Although a promising methodology, the IRAP has yet to be used with a forensic population to assess implicit beliefs regarding sexual behaviour. The aim of the present study was to determine if the IRAP would be more effective at revealing sexual offender’s implicit beliefs about children than an explicit (questionnaire-based) methodology. To this end, the IRAP was compared to the Cognitive Distortion Scale (CDS; Gannon, 2006), a questionnaire specifically constructed to measure beliefs.

2 All “in press” articles are available for download from: http://psychology.nuim.ie/IRAP/IRAP_1.shtml
associated with Ward and Keenan’s *children as sexual beings* and *nature of harm* implicit theories. In light of the evidence presented above, it was hypothesised that the IRAP would discriminate offenders from non-offenders more effectively than a traditional questionnaire. Given the exploratory nature of the research, we refrained from making specific predictions pertaining to possible differences that may emerge between the two groups across the four trial-types of the IRAP.

Method

Participants

Two groups (*N* = 32) participated in the present study. The experimental group (hereafter offender group) consisted of 16 male participants who had been convicted for at least one contact sexual offence against a child. The sample was community-based and was recruited on a volunteer basis through the UK National Probation Service. All individuals within the offender group were currently enrolled or had previously completed at least one Probation-run SOTP.

The control group consisted of 16 male non-offenders recruited from a university sample, including undergraduates, post-graduates, and non-academic staff members. Participants spoke English as their first language and had normal or corrected-to-normal vision. The participants provided written consent to participate in the research. Table 1 presents the demographic information of both samples.

| Table 1 here |
**Materials**

*Cognitive Distortion Scale (CDS; Gannon, 2006).* Explicit beliefs were measured using the CDS, adapted by Gannon (2006) from the Opinions Questionnaire (Offending Behaviour Programmes Unit, 2000). Gannon reported that the CDS yielded significant differences between sexual offenders against children and a non-offending post-graduate control group. The scale has high internal consistency (Cronbach’s alpha = .94) and test-retest reliability (.89) amongst a small number of child sex offenders.

The scale consists of 13 statements, each relating to cognitive distortions associated with Ward and Keenan’s (1999) *children as sexual beings* and *nature of harm* implicit theories. Participants are required to rate the statements (for example “children that sit in a way that is revealing are suggesting sex”) using a 5-point scale, ranging from 0 (*strongly disagree*) to 4 (*strongly agree*). Total scores range from 0 to 56, with a higher score indicating a stronger endorsement of the cognitive distortion-related statements.

*Implicit Relational Assessment Procedure (Barnes-Holmes et al., 2006).* The IRAP was administered using a portable Intel Pentium 4 laptop (software available from the second author upon request). The IRAP stimulus set was developed to reflect Ward and Keenan’s (1999) *children as sexual beings* implicit theory by the first author and two Clinical Forensic Psychologists working within the field. Two category labels, “Adult” and “Child”, were used with two sets of target stimuli, one set of sexual terms (e.g., “Sexual”) and a set of semantically opposite terms (e.g., “Non-Sexual”). Thus, for example, participants were asked to confirm on some IRAP blocks that children are sexual and to confirm on other blocks that they are not (i.e., directly targeting Ward and Keenan’s *children as sexual beings* implicit theory). The final choice of stimuli (see
Table 2) was based on ethical constraints, pilot testing, and effect sizes obtained in previous research using the IRAP (note that controlling for word length and frequency is unnecessary because each target word acts as its own control).

---

**Table 2 here**

---

**Procedure**

All participants completed the IRAP first. Visual instructions were presented to each participant and were read through with the experimenter (available from the first author upon request). Care was taken to ensure understanding of the task requirements, and specifically that it would sometimes be necessary for the participant to respond to the stimuli in ways that reflected what they believed, and sometimes in ways that appeared to contradict their beliefs. The researcher emphasised the requirement to respond as quickly as possible while also avoiding the “red x” during the computer task, which was used to indicate an incorrect response (see below).

Each participant was administered a number of practice blocks to ensure an average response time of less than 5 seconds and an accuracy rating above 80%. This ensured understanding of the task requirements and minimised the likelihood of random responding. Each participant reached these minimal criteria within four practice blocks.

Each trial involved the presentation of a category label (the word ‘adult’ or ‘child’) at the top of the screen, one of 12 target words in the centre (e.g. ‘sexual’, ‘erotic’, ‘innocent’), and the two response options of ‘true’ and ‘false’ in the bottom corners. All of the stimuli (label, target and response options) appeared simultaneously
(see Figure 1). All of the stimuli remained visible until the participant selected one of the relational terms by pressing either the ‘D’ key for ‘true’ or the ‘K’ key for ‘false’.

Selecting the correct relational term for the trial removed all of the stimuli for 400ms, after which the next trial was presented. Selecting the incorrect relational term produced a red ‘x’ in the centre of the screen, which remained there until the participant selected the correct response option – doing so moved the program to the 400 ms inter-trial interval.

A correct response was defined in terms of whether the participant was completing a consistent or inconsistent block of trials. During a consistent block, participants were required to respond to adults as sexual (e.g. ‘adult’ – ‘sexual’ – true; ‘adult’ – ‘non-sexual’ – false) and to children as non-sexual (e.g. ‘child’ – ‘non-sexual’ – true; ‘child’ – ‘sexual’ – false). During an inconsistent block, the feedback contingencies were reversed, and participants were required to respond to adults as non-sexual (e.g. ‘adult’ – ‘sexual’ – false; ‘adult’ – ‘non-sexual’ – true) and to children as sexual (e.g. ‘child’ – ‘sexual’ – true; ‘child’ – ‘non-sexual’ – false). Table 2 indicates which target stimuli were consistent/inconsistent with each of the two category labels.

Participants were exposed to six test blocks (3 consistent and 3 inconsistent). The blocks alternated between consistent and inconsistent (i.e. test block 1 = consistent trials; test block 2 = inconsistent trials; test block 3 = consistent trials etc.). Within each block, the category label was presented with a target word in a pseudo-random order with the constraint that each target was presented twice with each sample. After each test block participants were given visual instructions indicating that the previous correct and incorrect answers would be reversed in the next block. Following the sixth and final block, an instruction appeared on screen asking the participant to report to the researcher.
Following the IRAP, participants completed the cognitive distortion scale and a brief demographic questionnaire.

Results

Scoring the IRAP

In line with contemporary implicit measurement research, latency data from the IRAP were transformed into a $D$ measure (hereafter $D$-IRAP; Barnes-Holmes, Murtagh, et al., in press; Barnes-Holmes, Waldron, et al., in press; Cullen & Barnes-Holmes, in press; Vahey, et al., in press) using an adapted version of Greenwald, Nosek and Banaji’s (2003) IAT $D$-algorithm. The $D$ transformation functions to minimise the impact of factors such as age, motor skills, and/or cognitive ability on latency data, allowing researchers to measure differences between groups using a response-latency paradigm with reduced contamination by individual differences associated with extraneous factors (Greenwald et al., 2003). A larger $D$ score indicates a greater difference in response latencies between consistent and inconsistent trials. Positive scores indicate responding in accordance with pre-experimentally defined biases (i.e., within the current study, adults as sexual and children as non-sexual) and negative scores indicate the opposite (i.e., adults as non-sexual and children as sexual). Scores that approach zero indicate no discrimination between adults or children as sexual or non-sexual.

The raw latency scores were converted into five $D$-IRAP scores: a score for each of the four different trial-types (adult-sexual; adult-non-sexual; child-non-sexual; and child-sexual) and an overall $D$-IRAP effect score (the mean of the four trial-type scores). The procedure for the transformation of the raw latency data is provided in Table 3.
IRAP Results

Data from all participants were retained following the transformation of raw latencies into D-IRAP scores. The five mean D-IRAP scores for both groups are presented in Figure 2. The overall D-IRAP scores for both groups were in a positive direction, but the non-offender score was almost twice that of the offender group. In effect, both groups demonstrated a bias towards adults as sexual and children as non-sexual, but the bias for the non-offenders was double that of the offenders. The results for the individual trial-types also indicated lower D-scores in each case for the offenders, but the group difference was most pronounced for the child-sexual trial-type. In fact, this trial-type produced the largest D-IRAP score for the non-offenders and the smallest (approaching zero) for the offenders.

A 2x4 mixed analysis of variance (ANOVA) was conducted with group (offenders vs non-offenders) as the between-participant variable and the four IRAP trial-types as the repeated measure. The ANOVA revealed a significant effect for group \( (F_{(1,30)} = 4.381, p = .045, \eta_p^2 = .127) \), with no main effect for trial-type \( (p > .14) \). The interaction between group and trial-type approached significance \( (F_{(3,90)} = 2.273, p = .085) \). Four one-way between-participants ANOVAs were used to conduct planned comparisons for each trial-type. Only the child-sexual trial-type produced a significant
group difference \( F(1,30) = 6.136, p = .019, \eta^2_p = .170 \) (remaining \( p \) values > .12). In effect, the offender group produced a significantly smaller bias against children as sexual relative to the non-offender group.

Eight planned one-sample \( t \)-tests were used to determine if the \( D \)-IRAP scores for each of the trial-types for both groups were significantly different from zero. All but one of the \( t \)-tests proved to be significant (all \( ps < .03 \), with the non-significant effect obtained for the child-sexual trial-type for the offender group (\( p > .96 \)). The non-offender group thus responded in accordance with the pre-experimentally defined social norms with faster responses to; (i) \textit{adults-sexual-true} relative to \textit{adults-sexual-false}, (ii) \textit{adults-non-sexual-false} relative to \textit{adults-non-sexual-true}, (iii) \textit{children-non-sexual-true} relative to \textit{children-non-sexual-false}, and (iv) \textit{children-sexual-false} relative to \textit{children-sexual-true}. Although the offender group reproduced this pattern for three of the trial-types, critically, the child-sexual trial-type failed to produce a significant IRAP effect. In other words, the offenders appeared to be unable to discriminate, in terms of response latency, between children as sexual versus non-sexual.

\textit{Cognitive Distortion Scale}

Unexpectedly, the non-offender group achieved a higher mean score (\( M = 8.19, SD = 7.26 \)) than the offender group (\( M = 4.62, SD = 4.73 \)) on the CDS, indicating that the non-offenders actually endorsed more cognitive distortion-related statements than the offender group, though this result did not reach statistical significance (\( t(30) = 1.645, p = .110 \)).
Previous Offences

Within the offender group, no significant correlation was identified between number of previous offences and the CDS or any of the five D-IRAP scores (all $p$ values $>.16$).

Treatment Effect

Each participant within the offender group was currently enrolled or had previously completed an SOTP. The number of treatment sessions received was not significantly correlated with any of the D-IRAP scores or the CDS (all $p$ values $>.3$).

Education Effects

The offender and non-offender groups did not differ significantly in terms of age, but did differ significantly in relation to years of education ($t_{(30)} = 2.450, p = .02$). As outlined above, the $D$-score transformation is used to control for individual differences in response speed and cognitive ability more generally, and thus educational attainment should not impact significantly on the transformed IRAP measure. Nevertheless, a series of correlations were conducted to ensure that the $D$-IRAP scores did not in fact correlate with the education variable. All correlations proved to be non-significant (all $p$ values $>.19$)

Prediction of Group Membership

The IRAP data indicated that offenders produced significantly smaller $D$-IRAP scores than non-offenders at the group level for the child-sexual trial-type. In order to determine the predictive validity of this $D$-score, a discriminant analysis was performed. The value of this function was significantly different for offenders and non-offenders ($\chi^2_{(1, 32)} = 5.489, p = .019$), and overall the discriminant function successfully predicted
IRAP: The assessment of implicit sexual beliefs 19

outcome for 62.5% of cases, with accurate predictions being made for 68.8% of the offender group, and 56.3% of the non-offender group. This indicated a 31.2% \((n = 5)\) false negative misclassification of the offender group, and a 43.7% \((n = 7)\) false positive classification of the non-offender group.

Discussion

The current research aimed to determine if the IRAP could prove a useful methodology for identifying key differences between the implicit beliefs of sexual offenders and non-offenders, and provide further empirical support for Ward and Keenan’s (1999) *children as sexual beings* implicit theory. The sexual offender group within the current research obtained a significantly smaller overall \(D\)-IRAP score than the non-offender group. Although both groups responded in a positive direction, indicating an overall response bias towards adults as sexual and children as non-sexual, this response bias was significantly reduced in the sexual offender group. In other words, the overall \(D\)-score suggested that although both groups were able to discriminate between adults as sexual and children as non-sexual, this ability was significantly impaired in the offender group.

Critically, the difference between the two groups was most pronounced for the child-sexual trial-type of the IRAP. Although the non-offender group demonstrated a significant IRAP effect for this trial-type, suggesting a strong response bias towards children as *not* sexual, the sexual offender group failed to make this distinction. In fact, there was an almost complete absence of an IRAP effect, indicating that the sexual offenders confirmed and denied with almost equal speed that children were sexual. This response profile offers support for Ward and Keenan’s (1999) *children as sexual beings*
implicit theory, in that the offender group failed to discriminate children as non-sexual, whereas the non-offender group demonstrated a strong response bias to children as non sexual.

Interestingly, the offender group showed a significant IRAP effect for the child-non-sexual trial-type. Although this result might appear to contradict the absence of an effect for the child-sexual trial-type, it could be seen as broadly consistent with Ward and Keenan’s (1999) dangerous world implicit theory. Central to this theory is that children can be both sexual and innocent, and thus make attractive, non-threatening sexual partners to individuals that feel intimidated by age-appropriate sexual relationships. Although this is offered as an entirely post-hoc and speculative interpretation, it serves to highlight the potential of the IRAP to access multiple aspects of the implicit beliefs of child sex offenders.

Overall, therefore, although the IRAP methodology targets belief-statements as opposed to the “raw” associations targeted by the IAT and related measures (Barnes-Holmes et al., 2006), the current results support previous findings that have identified stronger implicit child-sex associations amongst child sex offenders (e.g. Gray et al., 2005; Kamphuis et al., 2005). Previously expressed concerns regarding the use of explicit assessment methods to ascertain socially sensitive attitudes and beliefs were also upheld (Schwartz & Baer, 1996) in that, contrary to the findings of Gannon (2006), no significant differences were found between the two groups based on their endorsement of cognitive distortion-related statements. In fact, the mean CDS scores were more elevated in the non-offender group, though not reaching statistical significance. This finding is consistent with other research that suggests sexual offenders are often indistinguishable
from non-offenders when using cognitive distortion questionnaires (e.g. Tierney & McCabe, 2001), and that explicit attitudes often differ from implicit attitudes when the latter are measured by procedures such as the IAT or IRAP (Barnes-Holmes, 2006; Greenwald & Banaji, 1995).

The current results indicate that the IRAP may have some, albeit moderate, discriminative validity in predicting group membership. In the present study, 68.8% of the offenders were correctly identified based on their $D$-IRAP scores for the child-sexual trial-type. Unfortunately, this indicated that 31.2% of the offenders were incorrectly categorised as non-offenders, and 43.7% of the non-offenders as sexual offenders. This compares slightly less favourably to the IAT data presented by Grey et al. (2005), which correctly categorised 78% of paedophiles at the expense of incorrectly categorising 42% of the control participants. Although number of sexual convictions was not found to be significantly related to the IRAP results or the CDS in the current study, this may be due to the very low base rate within the offender group. It is therefore possible that the predictive validity of the IRAP may increase amongst individuals with more severe offending histories, although this is clearly a matter for further research. Nevertheless, it must be acknowledged that although the IRAP may have some discriminative validity, the results presented here suggest that at present, similar to the IAT, it is an imprecise tool.

Whether the discriminative accuracy of the IRAP, or indeed the IAT, could be enhanced through further research is dependent on additional factors. Although implicit theories have been hypothesised to play an important role in the offending process, and the results presented here suggest differences between the implicit beliefs of sexual
offenders and non-offenders, little empirical support demonstrating the causal nature of these beliefs has been demonstrated within the broader sex offender literature. It is therefore plausible that implicit theories and beliefs may be demonstrated to be poor predictors of overt sexual behaviour, and implicit assessment methodologies will therefore never attain substantial predictive validity: not all offenders may hold implicit theories about children, and conversely, some non-offenders may hold sexual beliefs about children but never act upon such beliefs. Indeed, the CDS results presented here indicate that non-offenders report at least some of the factors associated with blurred boundaries regarding children and sex.

The predictive or discriminative validity of the IRAP as a technology is also inextricably linked to the saliency of the stimulus set being used. The stimulus set within the present study was chosen to reflect Ward and Keenan's (1999) children as sexual beings implicit theory, and therefore contained words pertaining to sex and attributes traditionally related to children (e.g., innocent). However, the saliency and meaning of particular words will be dependent on an individual's personal history, and it is possible that the stimulus set used within this study did not adequately access those beliefs related to sexual offending. More simply, sexual offenders are known to be a heterogeneous population (e.g. Gudjonsson & Sirgurdsdson, 2002) and it is possible that the words chosen within the stimulus set had no functional relationship to offending behaviour amongst some individuals within the offending sample, and therefore did not produce an IRAP effect (indeed, it could be argued that both the offending and non-offending samples employed in the current study were quite heterogeneous). Future research with sexual offenders may consider testing different stimulus sets to assess whether particular
target stimuli are associated with different sexual offence typologies, or whether stimulus sets can be ‘tailored’ to individual offenders, taking into account factors such as victim specificity and so on.

The data obtained from the current study have a number of implications for those involved in the assessment, treatment and management of sexual offenders. While technologies such as the IRAP are in their infancy, the data presented here indicates that there may be merit in developing these methodologies for use within clinical research. If the IRAP is a valid measure of implicit beliefs, future research may attempt to determine whether a functional relationship exists between these beliefs and overt, recidivistic behaviour, through longer-term follow-up studies. Assuming that a functional relationship is identified, this research would allow for the relationship between implicit beliefs and re-offending to be explored, with the aim of developing more effective treatment programmes that attempt to target and modify the relevant processes.

As indicated above, the relationship between implicit beliefs and overt sexual behaviour is far from clear, and it is likely that these beliefs play only a part in what is undoubtedly a multifactorial, social, and contextually determined offending process (e.g. Ward & Siegert, 2002). The future utility of these methods appears best suited to identifying processes that may contribute to offending behaviour, with the focus on developing treatment programmes that target those processes. Within this context, it is possible that implicit methodologies could be used to supplement pre- and post-treatment assessment measures, although it remains to be determined if the IRAP can map clinical change, and indeed if such utility would be beneficial.
Finally, despite a burgeoning international research programme in the domain of implicit cognition, an agreed operational definition of implicit beliefs as a psychological construct has yet to emerge from the broader empirical literature (Barnes-Holmes et al., 2006; De Houwer, 2006). Consequently, assessing whether the IRAP is a valid measure of implicit beliefs is problematic, in that the term is yet to be strictly operationalised. Whether the IRAP accesses beliefs that an individual may wish to conceal (Barnes-Holmes et al., 2006), beliefs that are activated automatically (Wilson, Lindsey & Schooler, 2000), or beliefs that an individual may not be able to identify introspectively (Greenwald & Banaji, 1995) is yet to be determined. Nevertheless, it is important to remember that the IRAP discriminated between the offender and non-offender groups, which the CDS did not. Consequently, it seems important to undertake further research to assess by what definition the IRAP is an implicit measure, and whether through further development it can make a significant contribution to the clinical assessment and treatment of sexual offenders.
IRAP: The assessment of implicit sexual beliefs 25

References


measure attitudes towards meat and vegetables in vegetarians and meat-eaters.

_The Psychological Record._


32nd Annual Convention of the Association for Behaviour Analysis. Atlanta, USA.


*Social Psychology Quarterly, 66*, 83-96.


Table I. The demographic characteristics of both groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean Age</th>
<th>Mean Years of Education</th>
<th>Mean Number of Sexual Offence Convictions</th>
<th>Mean Number of Weeks within SOTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-offender (N = 16)</td>
<td>37.2 (SD 9.3)</td>
<td>14.9 (SD 4.1)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Offender (N = 16)</td>
<td>42.8 (SD 15.3)</td>
<td>12 (SD 2.4)</td>
<td>1.63 (SD 1.3)</td>
<td>28 (SD 37.4)</td>
</tr>
</tbody>
</table>
Table II. The complete stimulus set for the IRAP

<table>
<thead>
<tr>
<th>Sample 1: Adult</th>
<th>Sample 2: Child</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relational Term 1: <em>True</em></td>
<td>Relational Term 2: <em>False</em></td>
</tr>
<tr>
<td>Target words consistent with Adult:</td>
<td>Target words consistent with Child:</td>
</tr>
<tr>
<td>Sexual</td>
<td>Non-sexual</td>
</tr>
<tr>
<td>Flirty</td>
<td>Pure</td>
</tr>
<tr>
<td>Seductive</td>
<td>Innocent</td>
</tr>
<tr>
<td>Arousing</td>
<td>Unarousing*</td>
</tr>
<tr>
<td>Erotic</td>
<td>Non-erotic</td>
</tr>
<tr>
<td>Sexually Aware</td>
<td>Sexually Unaware</td>
</tr>
</tbody>
</table>

* Although not found in an English dictionary, pilot testing indicated that this word was semantically equivalent to “Not arousing”.

Table III. The method for transforming raw latency scores to $D$-IRAP scores

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Use only test-block data</td>
</tr>
<tr>
<td>2</td>
<td>Eliminate latencies above 10,000 ms from the dataset</td>
</tr>
<tr>
<td>3</td>
<td>Remove all data for a participant if 10% of the test-block response latencies are less than 300 ms</td>
</tr>
<tr>
<td>4</td>
<td>Calculate 12 standard deviations for the four trial-types: four for the response-latencies from test-blocks 1 and 2, four from the latencies from test-blocks 3 and 4, and a further four from test-blocks 5 and 6</td>
</tr>
<tr>
<td>5</td>
<td>Calculate 24 mean latencies for the four trial types in each test-block</td>
</tr>
<tr>
<td>6</td>
<td>Calculate difference scores for each of the four trial types, for each pair of test blocks, by subtracting the mean latency of the consistent test-block from the mean latency of the corresponding inconsistent test block</td>
</tr>
<tr>
<td>7</td>
<td>Divide each difference score 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</td>
</tr>
<tr>
<td>8</td>
<td>Calculate the four overall trial-type $D$-IRAP scores by averaging the three scores for each trial-type across the three pairs of test blocks</td>
</tr>
<tr>
<td>9</td>
<td>Calculate an overall relative $D$-IRAP score by averaging all 12 trial-type $D$-IRAP scores from step 8</td>
</tr>
</tbody>
</table>
Fig. 1

Adult - Sexual

- Adult
  - Sexual
    - Consistent
    - Inconsistent
    - Select 'd' for
    - Select 'k' for
    - True
    - False

Child - Sexual

- Child
  - Sexual
    - Inconsistent
    - Consistent
    - Select 'd' for
    - Select 'k' for
    - True
    - False

Adult - Non-sexual

- Adult
  - Non-Sexual
    - Inconsistent
    - Consistent
    - Select 'd' for
    - Select 'k' for
    - True
    - False

Child - Non-sexual

- Child
  - Non-Sexual
    - Consistent
    - Inconsistent
    - Select 'd' for
    - Select 'k' for
    - True
    - False
Fig. 2

The figure shows the D-IRAP scores for different trial types and groups. The bars represent the means, with error bars indicating the standard error. The trial types include Adult Sexual, Adult Non-sexual, Child Sexual, Child Non-sexual, and Overall D-IRAP. The groups are Non-offender and Offender. The graph illustrates the implicit sexual beliefs across these categories.
IRAP: The assessment of implicit sexual beliefs 37

Figure Captions

*Figure 1.* Examples of the four IRAP trial-types. The category label (“Adult” or “Child”), target word (“Sexual,” “Non-sexual,” etc.) and response options (“True” and “False”) appeared simultaneously on each trial. Arrows with superimposed text boxes indicate which responses were deemed consistent or inconsistent with social norms (boxes and arrows did not appear on screen). Selecting the consistent response option during a consistent block of trials, or the inconsistent option during an inconsistent block, cleared the screen for 400 ms before the next trial was presented; if the consistent option was chosen during an inconsistent block, or the inconsistent option chosen during a consistent block, a red X appeared on screen until the participant emitted the alternative response.

*Figure 2.* A graph depicting the mean $D$-IRAP scores for each of the individual trial-types, and the overall mean $D$-IRAP effect score.