Chapter 5

THE IMPLICIT ASSOCIATION TEST AS A MEASURE OF SEXUAL INTEREST[1] [Q1]

NICOLA S. GRAY
Cardiff University, Cardiff, and Caswell Clinic, Bro-Morgannwg NHS Trust, Bridgend, UK

ROBERT J. SNOWDEN[2]
Cardiff University, Cardiff, UK

INTRODUCTION [Q2]

In many domains of psychology, we wish to measure a person's thoughts, attitudes and emotions to a particular situation or event. By far the most popular way to do this is to simply ask the person to report on these. Whilst we do not doubt that such methods are useful and need to continue, there are three reasons to think that other methods are needed to supplement these explicit statements.

Most of our neural lives are not available for introspection in a conscious form. In vision science, we have learnt that activity in the early stages of visual processing, in the retina, thalamus and primary visual cortex, does not produce a conscious experience available for self-report. A person with severe damage to primary visual cortex may report being blind in certain parts of their visual field, but will flinch if an object appears suddenly in this part of the visual field. This phenomenon of 'blindsight' (Cowey, 2004) is just one of many examples of behaviour being guided by non-conscious mental states (Wilson and Dunn, 2004). Perhaps more surprisingly, even mental states such as 'emotions' can be shown to exist without conscious awareness (Winkielman and Berridge, 2004). Thus, if we are effectively

[1]All human images appearing in this chapter are commercially available stock photography licensed to the author for the purposes of software development and publication in print.
[2]The author of this chapter is the developer of the Affinity assessment procedure, which is distributed by Pacific Psychological Assessment Corporation (PPAC). He has received from PPAC royalties on software sales, and also subsidized travel to the United States.

Cognitive Approaches to the Assessment of Sexual Interest in Sexual Offenders Edited by D. Thornton and D. R. Laws

unconscious to many of these influences on our behaviours, we clearly cannot rely on self-report for us to access them. Alternative methods are needed.

Even if I can access my mental state, or indeed reliably report on my overt behaviours, there is still a second problem to overcome. Many self-report questionnaires ask us to comment on how we would react in certain situations – on whether we like going to parties, whether we have a bad temper, etc. To answer such questions, we not only have to speculate and generalize, but also have to 'norm' ourselves to others. So to answer whether I have a bad temper, I have to have some understanding of my temper and to compare it with that of others in order to decide if mine is 'bad' or 'normal'. Of course, *I* believe that I only lose my temper when it is clearly justified; however, others observing me may not agree! This problem is wonderfully illustrated by the finding that most people regard their driving as well above average, indeed over 50% of us think that we are in the top 10%! (Horswill, Waylen and Tofield, 2004; Svenson, 1981). Clearly many individuals regard themselves as good (i.e. above average) as they drive in the way they think is the best way. So someone crawling along the freeway at 30 kmph may regarded this as good safe driving (the people in the cars behind may take a different view), whilst to another executing a handbrake turn on a busy freeway shows their great driving skills (again the cars swerving to avoid this vehicle may beg to differ). There are many other examples of this 'Lake Wobagon' effect (a place where everybody is above average!) (Alicke *et al.*, 1995).

Finally, in order for self-report to be useful the person reporting needs to be honest. For many reasons this may not be the case, and in the domain of sexual offending there is obvious reason for an offender to minimize, deny or distort their responses (Cooper, 2005).

IMPLICIT MEASURES

Given the limitations of self-report just described, we must look to implicit measures to help us index the mental state of the person under scrutiny. Whilst most of this volume is dedicated to 'new' methods, and in particular methods based on measures of cognitions, using implicit measurement is by no means a new enterprise. For many years, a person's reaction to a stimulus has been indexed by psychophysiological measurements such as heart rate, sweating, brain waves, facial expressions, body language, etc. For example, the pupil of the eye reacts to many situations other than just changes in light. Thus a person's excitement, including sexual interest, can be ascertained by measures of changes in pupil size in response to appropriate stimuli (Hess and Polt, 1960). Indeed, in the domain of sexual offenders (and sexual research in general), measurement of sexual arousal via penile tumescence or vaginal blood flow (see chapter by Laws) still remains the 'gold standard' by which the new cognition-based methods will be compared.

[Q3]

More recently, these psychophysiological measurements have been joined by ones based on cognitive methods. Implicit methods have long played a central role in the domain of cognitive psychology. For example, semantic priming experiments have aimed to reveal the nature of how semantic concepts are organized within our minds. In such experiments, a prime is presented (such as the word 'bread')

followed by a target word, such as 'butter'. Decisions about the target word (such as 'was it a real word or a made up one') are much faster for the word 'butter' if primed by 'bread' (but not if primed by 'elbow') showing a semantic association between these two concepts. It was not surprising, therefore, that such techniques have been adopted into the domain of social psychology. For example, the priming task was adapted by Fazio *et al.* (1995) to look at associations to different races by presenting pictures of faces that were either white or black, and then following this prime with a word that had to be categorized as a good or bad word. It was hypothesized that if a person has a negative view of a white face then this would prime the response to 'bad' and make such judgments faster. Using this technique, the authors did indeed show that white people are faster when the white face prime occurs before the good words, and a black face prime occurs before the bad words. This shows that there is a bias in the semantic affective associations to pictures of different races. We shall return to whether this is 'racism' in the next section. In the domain of sexual offending, this technique has yet to be employed, though Snowden, Wichter and Gray (2007) show that the technique has exceptional accuracy at identifying sexual orientation in men.

WHAT ARE 'IMPLICIT ATTITUDES'?

There is often great confusion amongst psychology undergraduates when these techniques are first explained to them. The first is the confusion of implicit attitudes with the implicit measurement of attitudes. For example, I am willing to report my support for the England football team (with therefore appropriate anti-support for the German football team) and I would fully expect this pro-English bias to be revealed by an implicit test. Thus, if I was the participant in a priming experiment with primes of English or German footballers and targets of good versus bad words, I expect myself to be fast when good words follow English players and also fast when bad words follow the German footballer primes. The test would have revealed my bias using an implicit technique. In this instance, the implicit technique has revealed a bias, but this bias was clearly also an explicit attitude and would have also been revealed by explicit techniques.

The more interesting case is that these implicit measurement techniques can often disagree with the explicit measurement. Thus, if it were found that I was fast for the good words following German footballers, and also fast when bad words followed English footballers, then this would show that I have some associations that were not tapped by simply asking me for my views. Clearly, years of watching inept displays by England as they underachieved, in contrast to the overachieving German team, combined with heartbreaking defeats in penalty shootouts at their hands, may have led some part of my mind (some might term this my 'sense') to the unquestionable reality that they are indeed 'good' and we are indeed 'bad' despite my desire for the opposite.

This example shows us that any 'attitude' is made from many influences and that we can have dissociations between those that are held implicitly (without our conscious knowledge) and those that are held explicitly (with our conscious knowledge). Hence, we do not believe in the view that implicit measures are simply

only a 'bona fide pipeline' (Fazio *et al.*, 1995) to measure explicitly held views, rather they offer us a method to probe earlier (and possibly different) automatic associations. These associations may, or may not, be available to consciousness, and may, or may not, be actually endorsed by the person (Olson, Fazio and Hermann, 2007). Nevertheless the presence of such associations influences our behaviours (see also Gawronski, LeBel and Peters, 2007).

So the crucial question becomes 'under what circumstances do these associations have any effect on behaviour'? Indeed, there have been stinging criticisms of the whole endeavour of implicit measurements suggesting that they measure nothing of interest, and that if a person does not endorse the view suggested by the implicit technique then it cannot be called an 'attitude' (Arkes and Tetlock, 2004). To us this is merely a matter of semantics, and we are happy to reserve the term attitude to those that are explicitly endorsed. The most important point, in our opinion, is that the implicit tests can predict future behaviours. The evidence suggests that there are situations where implicit tests can add to that predicted from explicit measurement alone, and in some cases the implicit measurement is by far the greater predictor (Friese, Wänke and Plessner, 2006; McConnell and Leibold, 2001; Phelps *et al.*, 2000; Richeson *et al.*, 2003). Given this evidence, it seems that implicit measures have much to offer in many domains of psychology and to dismiss them as 'mere associations' (Arkes and Tetlock, 2004) is to miss the point that these associations are driving forces behind our behaviour (Banaji, Nosek and Greenwald, 2004).

Let us illustrate this in the domain of the implicit measurement of racial bias. Implicit measurements of racial bias (such as priming techniques) have shown that many white Americans appear to have a bias against black people, despite showing no bias on explicit measurements (Nosek, Banaji and Greenwald, 2002). Is an individual with no explicit bias but with an implicitly measured bias a racist? We would argue against this, in that this term should be used for a person that deliberately endorses anti-black (pro-white) sentiments (and vice versa). However, we would argue that such a person is in danger of behaving in a prejudice manner in certain situations (Dovidio *et al.*, 1997). Under situations where they do not have time to consider their decisions, or where decisions are made without the activation of validation processes, then these automatic associations (whether conscious or unconscious) are able to guide behaviour and prejudice may occur (McConnell and Leibold, 2001; Vanman *et al.*, 2004; Strack and Deutsch, 2004; Ziegert and Hanges, 2005; Gawronski and Bodenhausen, 2005; Gawronski, LeBel and Peters, 2007).

The issue of the relationship between implicit and explicit attitudes presented here is only a simplified version of a debate that still attracts much attention (and invective), and the interested reader is guided to more detailed discussion in (Gawronski, LeBel and Peters, 2007; Nosek, 2007; Olson, 2003). We summarize with some simple statements that we think most researchers would agree with:

- Implicit measures capture something different from explicit statements and thus they are not merely another way of measuring the same concept.
- Implicit measures capture automatic associations held by the person (even if they do not explicitly endorse such an association).
- There are conditions under which automatic associations influence behaviour.

- Automatic associations are more likely to guide behaviour when validation processes are not used (either because they cannot be or because the person chooses not to).

Hence, we conclude that implicit measures can bring us information about sex offenders that explicit measures cannot, and that this information might be a guiding influence on the future behaviour of these individuals.

THE IMPLICIT ASSOCIATION TEST

The implicit association test (IAT) is a simultaneous sorting paradigm. We illustrate the technique via an often-used control task (Gray *et al.*, 2003; Greenwald, McGhee and Schwartz, 1998; Snowden *et al.*, 2004) that aims to measure the person's relative attitudes to flowers and insects. The target dimension is represented by words that need to be sorted (via pressing one of two buttons) into either flowers (e.g. tulip, rose) or insects (e.g. ant, locust). The affective or evaluative dimension is then represented by words that have to be sorted into either good (e.g. happy, lucky) or bad (e.g. cancer, frown). This second evaluative sort also used the same buttons as the target dimension, hence we can arrange on one block of trials for the concepts of flowers and good to share the same button (and therefore insects and bad the other button), whilst on another block we can reverse this so that flowers and bad share a button (and therefore insects and good the other button). Now imagine the word 'locust' appears on the screen. This should activate the representation of 'insect' within the mind of the person and in turn the impulse to press the 'insect' button. However, the concept 'insect' might also activate the concept 'bad' via the associations held within the mind of this person, and therefore the impulse to press the 'bad' button. Under conditions where bad and insect share the same button, these response impulses should be reinforcing producing fast and accurate performance, but under conditions where they do not share the button this will create response conflict, and performance will slow and accuracy may fall. By comparing speed and accuracy on the two blocks, we can get some idea of whether the person does have associations between insects and bad (and flowers and good) inside their minds. These two test stages are normally presented as stages 3 and 5 in a 5-stage process (see Figure 5.1).

Figure 5.2 illustrates some data from our own experiments. Note that the 'expected' result is found at a group level – most people did indeed perform faster when insects and bad (and flowers and good) shared response buttons. The effect is very large – producing a difference in reaction times of around 300 milliseconds. This difference is also very large when expressed in effect sizes producing a Cohen's $d = 1.72$. Note that a $d = 0.80$ is normally regarded as a 'large effect', so we need new language ('very large') to describe the effects produced. Such very large effect sizes are one of the reasons the IAT has attracted such attention.

The above results demonstrate that at a group level the IAT can be very powerful – but can it really pick up individual differences? When we first started our research on the IAT in forensic settings, we always used the flower/insect IAT as a control task to examine any generalized effects (e.g. IQ, motivation, institutionalization)

Stage	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Instructions	Insect Flower	Good Bad	Insect or Good Flower or Bad	Flower Insect	Flower or Good Insect or Bad
Example stimuli	Locust Tulip Rose Wasp Ant Daisy . .	Happy Lucky Frown Cancer Vomit Smile . .	Daisy Cancer Rose Smile Ant Frown . .	Rose Daisy Locust Wasp Tulip Ant . .	Ant Rose Cancer Lucky Daisy Smile . .

Figure 5.1 Illustration of a 5-stage IAT. In stage 1, the participant has to press the left key for insect words and the right key for flower words. In stage 2, they have to press the left key for good words and the right key for bad words. In stage 3, they are instructed to press the left key for insect or good words and the right key for flower or bad words. In stage 4, the flowers and insect key are reversed and they are asked to the left key for flower words and the right key for insect words. Finally, in stage 5, they are instructed to press the left key for flower or good words and the right key for insect or bad words. To calculate the IAT score, we compare average reaction time (and errors) in stage 3 with those in stage 5. For someone who likes flowers and doesn't, like insects, we expect them to be faster in stage 5 than in stage 3. Most experiments would also counterbalance whether flowers or insects were presented with the bad words in stage 3 (and therefore stage 5).

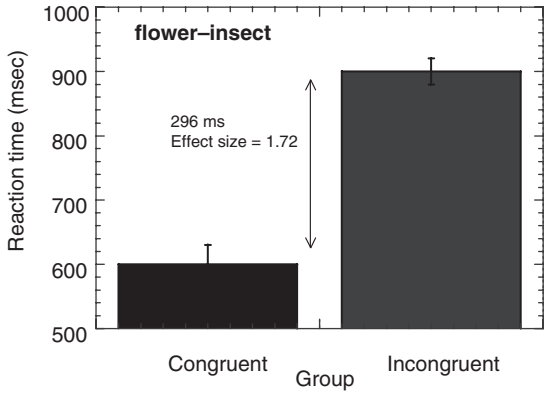

Figure 5.2 Results from a flower–insect IAT in an offender population (from the data of Snowden *et al.*, 2004). Here the term congruent refers to the conditions where flower and good shared the response button (and therefore insect and bad the other button), whilst incongruent is the condition where insect and good shared the response button (and therefore flower and bad the other button).

that might occur. It also served as an introduction to the IAT task and as a form of practice. Our first ever offender participant completed these (and other) IATs, and as we analyzed, the results we were very surprised to find that this person was faster to the insect/good combination. Our ethical protocol was to give feedback on the results to the participants and so we attempted to explain the results to him. We were worried that, perhaps, our task was faulty in some way. However, this offender was not surprised by the result and explained to us that when he was a child he was often locked in the cellar. During this time, his only source of creature comforts was to talk to the spiders and insects and so he had always loved them!

This anecdote does not, of course, provide scientific evidence that the IAT can index individual difference in attitude. So what would? This apparently simple question is actually quite hard to answer. After all how do we really know what the individual differences in attitude are? We could explicitly ask the person, but we have already critiqued the limitations of this self-report. When this has been done, the range of correlations between the explicitly and implicitly measured attitudes ranges markedly, with figures anywhere from near 0 (suggesting no relationships) to up to 0.80 (suggesting a close relationship) (Hofman *et al.*, 2005). It is not clear to us what figure, if any, would provide evidence that implicit measures can access individual's attitudes. A second method is to use a 'known-groups' technique. Here two groups are defined by some externally defined (and hopefully well agreed) criteria, and then the scores from the IAT are used to examine whether they can differentiate these groups. As we describe later, we have shown (Snowden, Wichter and Gray, 2007) that the IAT is capable of near perfect group discrimination under some circumstances.

WHAT ASSOCIATIONS CAN BE MEASURED BY THE IAT?

Greenwald, McGhee and Schwartz (1998) demonstrated IATs that examined associations between race and valence. Since this time there have been hundreds of different IATs reported in the literature, such that an exhaustive listing would indeed be exhausting. We mention a few here to give an idea of the huge range of areas where the IAT has been used. Attitudes to homosexuals (Banse, Seise and Zerbes, 2001) and the obese (Craeynest *et al.*, 2005; Teachman and Brownell, 2001), naughty versus nice foods (Richetin *et al.*, 2007), consumer choices (Coke vs. Pepsi?) (Karpinski, Steinman and Hilton, 2005), job-hiring decisions (Ziegert and Hanges, 2005) and voting preferences (Karpinski, Steinman and Hilton, 2005) have all been measured. Of particular interest to our laboratory are the studies that have aimed to measure concepts related to clinical constructs, such as IATs to measure phobias (Teachman, Gregg and Woody, 2001), anxiety (Egloff and Schmukle, 2002; de Jong, 2002), concepts related to eating disorders and body image (Vartanian, Polivy and Herman, 2004), attitudes to alcohol (De Houwer *et al.*, 2004; Wiers *et al.*, 2002) and smoking (De Houwer, Custers and De Clercq, 2006; Huijding *et al.*, 2005), attitudes to violence (Gray *et al.*, 2003; Luo *et al.*, 2006; Snowden *et al.*, 2004), self-harm (Nock and Banaji, 2007) and self-esteem (Greenwald and Farnham, 2000). It would appear that the ability to measure any association using the IAT is limited only by the experimenter's ingenuity.

METHODOLOGICAL CONSIDERATIONS

Since the first report on the IAT (Greenwald, McGhee and Schwartz, 1998), there have been many attempts to understand if and when variations in the technique produce different results. For instance, is the effect driven by the associations to the exemplars (e.g. locust) or by the category to which they have to be assigned (e.g. insect)? To test these ideas, we can attempt to manipulate the expected valence of the exemplars to be non-congruent with the category. So, for example, what would happen if English participants had to perform a German–English IAT in which the concept of German was represented by much loved Germans (e.g. Beethoven), but the concept of English represented by much hated English (e.g. Thatcher). Surprisingly the results of this type of test (De Houwer, 2001) show that the English participants were still faster under the English-good conditions, and the results differed little from when the concept of English was represented by much-loved English people (e.g. Churchill) and the concept of German by much-hated Germans (e.g. Hitler). Thus these results (and others) (Olson and Fazio, 2003) suggest that the IAT is driven at the categorical level, rather than by the exemplars of this category. This finding is theoretically important in that it tells us about the level of mental representations at which this test operates (note that other implicit tests, such as the priming task may operate at the level of exemplars; see Olson and Fazio, 2003), and also of practical importance as we only need to have sensible representatives of the category and it should therefore be robust against small changes in just how the category is represented. In line with these findings, the IAT appears to be also robust against the number of exemplars used to define the category (Brown, 2006; Nosek, 2005).

Our research in forensic settings has a target population containing many individuals with low cognitive abilities, and/or problems with memory and/or attention. We therefore have performed some experiments to adapt the basic technique to make it more acceptable and easier to administer in such forensic and clinical settings (Brown, 2006). To control for the effects of poor literacy in offender populations, we explored whether substituting pictures for words has an effect. We found that whilst picture versions of the task produced faster reaction times, they did not alter the basic effect, nor the effect sized produced compared to word versions (Brown, 2006). Therefore the use of pictures, or spoken words, may be most useful in applying the IAT to less literate populations such as children (Baron and Banaji, 2006) or those with intellectual disabilities. We now tend to use a hybrid picture–word version in which the target concept (e.g. flowers/insects) is represented by words and the evaluative concept (e.g. good/bad) is represented by words. We have found that the participants regard this task as 'easier' as it appears to demand less memory for just how to classify each stimulus. Whilst the standard 5-stage IAT is quite short (approximately 15 minutes), we found that it was sometimes too long in these applied settings, especially when we required several IATs to be completed. We therefore compared a shortened version in which only the two crucial stages are presented, to the standard 5-stage version using the same stimuli and the same participants. Again, we found no differences in the efficacy of this shortened version (Brown, 2006) and thus can recommend this shortened version for use in such settings (see also Teachman, Gregg and Woody, 2001)

Finally, the associations in the mind that produces the IAT effects could be manifest in slower reaction times, or in a greater number of errors, or in some combination of these effects. It would therefore be of great value to somehow combine the two effects into a single score. Greenwald, Nosek and Banaji (2003) have tested several scoring algorithms that attempt to do this and recommend one in which trials on which an error occurs are substituted by a 'punishment' reaction time (for details see their paper), and that the overall IAT score (the difference in the reaction times between the two crucial blocks of trials) is then expressed in terms of the variance in reaction times. This *D*-score therefore gets round some individual difference problems in overall reaction times (e.g. offenders tend to have slower overall reaction times than non-offenders). We have also found that this method of scoring eliminates a small, but problematic, influence of IQ on the IAT (Brown, 2006; see also Cai *et al.*, 2004) and produces greater effect sizes.

In conclusion, the results of many tests seem to show that the IAT is very robust to the exact methodology employed. The scoring method of Greenwald, Nosek and Banaji (2003) also eliminates 'unwanted' influences of individual differences that are not related to implicit associations *per se* (e.g. IQ). Clearly, these properties are highly desirable in any test that aims to have an applied use.

PSYCHOMETRIC PROPERTIES OF THE IAT

The IAT is still in its infancy and, as mentioned above, there are many methodological differences between studies that make definitive statements upon the psychometric properties of the IAT impossible. It will also be the case that the subject matter under consideration will play a strong role in determining the psychometric properties of the task. Therefore, the properties illustrated in this section are nothing more than illustrative. (For a more comprehensive look at these issues, see Nosek, Greenwald and Banaji, 2007.)

Psychologists normally look for three types of reliability when examining a test – internal reliability (do different parts of the test purporting to measure the same thing give similar results?), inter-rater reliability (do two different raters scoring the same case give similar results?) and test–retest reliability (does the same individual tested at different times give similar results?).

The internal reliability of reaction time tests can be calculated using Cohen and Cohen's (1982) formula which treats each response latency as an item. Using this technique, we found very good internal consistency (Cronbach's alphas = 0.81−0.93) for our flower/insect task (Brown, 2006), a result echoed by others (e.g. Nosek, Greenwald and Banaji, 2007). We also showed the same internal consistency for an IAT designed to look at child and sex associations (Cronbach's alphas = 0.80–0.87; Brown, 2006).

The inter-rater reliability is not an issue for the IAT as it is scored via a computer with no human judgment involved.

The test–retest reliability of the IAT is a more problematic issue. In order to get a sensible figure, we need a mental concept that is not changing between the test sessions. There is no guarantee that this will be the case for the IAT, and what little research there is on this matter certainly does not rule out 'state' influences

on the task (Gawronski, LeBel and Peters, 2007; Nosek, 2005). Nosek, Greenwald and Banaji (2007) examined the results from many IATs designed to measure a wide range of concepts and report an average test–retest reliability of $r = 0.60$. This represents a robust and reliable instrument. Of particular relevance to this chapter is that we (Brown, Gray and Snowden, 2008) have measured the test–retest reliability for our child/sex IAT (see below) and found a value of $r = 0.63$.

In conclusion, all measures of the psychometric properties of the IAT, accrued from a diverse range of IATs, methods and population samples, point to an instrument with robust and reliable properties.

FAKING THE IAT

Whilst we have argued that the IAT is more than just a method for measuring the same concepts as those tapped by explicit measures, it is also clear that the IAT hopes to measure attitudes that the person may wish to hide or dissimulate. To this end, there have been several reports of attempts to see if a person can fake the results on an IAT.

Banse, Seise and Zerbes (2001) asked participant to fake a positive attitude towards homosexual males. It was found that they were unable to do so, and this lack of ability to fake responses to the IAT when not told how to do so is supported by several studies (Asendorpf, Banse and Mucke, 2002; Egloff and Schmukle, 2002; Kim, 2003), with other studies showing only small effects of attempted faking (Steffens, 2004). However, there is some evidence that if given clear instructions on how to successfully fake an IAT participant can do this (Fiedler and Bluemke, 2005). It should be noted that these deliberate strategies should produce a series of responses that are not the same as someone with the genuine mental association that they are trying to emulate, and efforts are being made to devise algorithms that could detect such deliberate faking strategies (Cvencek and Greenwald, submitted).

In conclusion, it appears that the IAT is hard to fake, but not impossible. Given the possible high stakes for a sex offender, and the possibility that implicit tests could be used to monitor distorted cognitions (thus being tested on multiple occasions), further research is needed so that faking can be minimized, or at least detected when it occurs.

INVESTIGATING SEXUAL ORIENTATION USING THE IAT

As a precursor to testing for distorted cognitions in those with a history of sexual offences, we decided to test the ability of the IAT to successfully detect sexual orientation. We devised an IAT where the target dimension was male versus female (represented by pictures) and the evaluative dimension were words that were sexually attractive versus unattractive. To test the efficacy of this IAT to identify sexual orientation, we tested men who described themselves as either heterosexual or homosexual (Snowden, Wichter and Gray, 2007). The results were unremarkable – the IAT for the heterosexual men were consistent with an association between sex and

females, whilst those for the homosexual men were consistent with an association between sex and males. What was remarkable was the magnitude of this effect. In terms of effect sizes, this IAT produced an effect size of 2.73 (remember an effect size of 0.80 is regarded as large!).

We then used signal detection theory to examine just how well the IAT could detect a person's sexual orientation. We constructed the receiver operating characteristic (ROC) for the IAT by plotting the percentage of hits (correctly predicting that someone is more attracted to males on the basis of their IAT score) against the false alarms (incorrectly predicting that someone is more attracted to males on the basis of their IAT score) for many possible IAT scores (see Figure 5.2). For a test that cannot distinguish between groups, the probability of the false alarms will equal to that of the hits and the points of the ROC will follow the diagonal line through the middle of the ROC (see dashed line), whilst if we could perfectly detect one group from the other, the data points would fall along the *X*-axis and then the top of the *Y*-axis. To quantify the ROC, we can calculate the resulting area under the curve (AUC). If the points fall along the diagonal, the AUC would be half the graph or 0.5. If the points call on the axis, the AUC should be the entire graph or 1.0. The ROC produced by our IAT was 0.97 (which was significantly different from chance levels, $p < 0.0001$).

Armed with this sex IAT, we were able to investigate theories of human sexuality. It has been suggested that women's sexual attractions are far more ambiguous and malleable that those of males (Baumeister, 2000; Bem, 1996; Diamond, 2003). For instance, as male sex drive increases, this is associated with an increased sexual attraction to females, but a decreased sexual attraction to males. For females, increased sex drive is associated with increases in sexual attraction to both males and females (Lippa, 2006; Lippa and Arad, 1997). We (Snowden, Brown and Gray, submitted) therefore predicted that the sex IAT would not show a strong preference for males or females in heterosexual women, despite these women explicitly stating that their sexual preference was for males (and, if we can assume their honesty, that they were involved only in heterosexual relationships). Our results (see Figure 5.3) showed that these heterosexual females showed no strong sex association to either gender, in fact what little bias was found was towards an association between sex and female! Thus, our results suggest that heterosexual women (as a group) have implicit automatic cognitions relating to sex that are quite unlike those of males (be these hetero- or homosexual males), and are quite unlike the ones they explicitly express. The results from the sex IAT mirror those that have used physiological measures of sexual arousal and women (Chivers and Bailey, 2005; Chivers *et al.*, 2004), which also show similar sexual responses to stimuli depicting male erotic and female erotic stimuli in heterosexual women. Again, this differs from the category-specific sexual arousal responses of both heterosexual and homosexual males.

We also tested females who stated that their preferred sexual partners were female (Snowden, Brown and Gray, submitted). These lesbians clearly showed implicit cognitions consistent with an association between sex and female (see Figure 5.3). Thus, the responses of the lesbians on the sex IAT suggest that their **[Q5]** sexuality is more like that of a male, with category-specific associations, than those of heterosexual females. Whilst this is clearly a finding in need of replication and

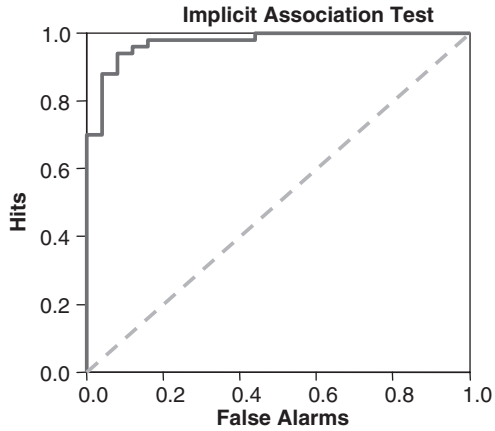

Figure 5.3 Receiver operating characteristic (ROC) analysis of the data from Snowden, Wichter and Gray (2007) of the ability of a sex–gender IAT to predict a male's stated sexual preference. The proportion of 'hits' (the correct prediction that someone is a homosexual) is plotted against the proportion of 'false positives' (the incorrect prediction that someone is a homosexual). The dotted line represents the expected result if the test cannot make this prediction. Our line falls a long way from this 'chance' performance revealing the test has exceptional ability to identify sexual orientation in this population.

exploration using other techniques, it illustrates the power of implicit techniques to explore theories of human sexuality.

CHILD/SEX IAT

Gray *et al.* (2005)

We began our work on the child/sex IAT in 2001 as part of a larger study looking at many different offence-specific IATs (including those related to attitudes to violence, arson, rape as well as control IATs; see Gray *et al.*, 2003; Snowden *et al.*, 2004) in offenders at HMP Grendon (Oxfordshire, UK). One of the advantages of this prison is that it is run as a therapeutic community and all the inmates must not be on psychotropic medication, or take any illegal substances, in order to be admitted and continue to be residence within this community. We designed an IAT that we believed might be able to index associations between sex and children. At this time, only IATs using the affective dimension (good vs. bad) against a target dimension had been published and so our first thoughts were to devise a target dimension of child–sex versus adult–sex and an affective dimension of good versus bad. However, this was not the way forward. Firstly, representing the concept of child–sex is fraught with ethical and moral problems, and the South Wales police service gave us quite unambiguous advice that this was not an area we should explore! Further reflection also suggested that this was not what we wanted to measure. What we would have got from this is a person's view of

child–sex (i.e. whether it is a good or bad thing), not whether the person has an association between sex and children. This idea of whether child–sex is bad is not without its own interest in view of evidence that cognitive distortions are prevalent in paedophiles (e.g. Ward, 2000; Ward *et al.*, 1997; Ward and Keenan, 1999). However, this was not our main area of interest, so we devised an IAT in which the usual affective dimension of good versus bad was replaced by one in which the dimension was sex versus not sex – a radical departure (or so we thought at the time) from the usual IAT. We represented this dimension with words that the offenders at HMP Grendon associated with sex (e. g. cock, pussy) and used words of similar length that we believed were not related to sex to represent the not sex category (e.g. elbow, eye – it actually proved quite hard to find any words related to body parts that no one reported as being related to sex. Our investigations suggest a surprisingly large predilection for toes!). The adult–child dimension was represented by words such as 'beard' and 'mature', and the child dimension by words such as 'infant' and 'school'.

Our results (Gray *et al.*, 2005) showed that the control offenders (no conviction for offending against children under 16) showed faster reaction times when adult and sex concepts shared the same response button, but the sex offender group (any conviction for sexual offending against a child) had faster reaction times when the child and sex concepts shared the response button. The group differences were highly significant ($p < 0.001$) despite the quite small numbers ($N = 18$ for the sex offender group), which reflected a large effect size (Cohen's $d = 0.83$). Construction of the ROC also showed that the child/sex IAT has a moderate ability to distinguish those with a child–sex offence from the other offenders (AUC = 0.73). It should be noted that this AUC is likely to be a lower limit of accuracy as we have no means of proving that our control group did not include some people with unconvicted child–sex offences within it. We concluded the paper by suggesting that implicit methods might one day be a useful additions to the armoury of the clinician, as well as a tool for research into offence-related cognitions. Since then there have been four further reports using IATs to explore cognitions in child–sex offenders.

Brown (2006)

Following our initial results, we were keen to replicate and extend the child/sex IAT towards a possible clinical tool. Along with our PhD student, Anthony Brown, we identified a number of factors that were problematic in the original IAT we devised. Amongst these were the long duration of the test(s), memory confusions over what classification was needed (e.g. should they be classifying the word 'beard' according to whether it was adult vs. child or whether it was sex vs. not sex?), and the possible confounds of IQ. In order to get round these problems, we initially tested the effects of changes in the IAT using the control task (flowers and insects) in a sample drawn from the local community, and found that (1) a shortened version of the task (using only the crucial two stages) produced an equivalent IAT effect to the original 5-stage version, (2) using pictures for one dimension (e.g. flowers vs. insects) and words for the other (e.g. good vs. bad) gave the same results as using all words (and indeed all pictures) and (3) there were small effects of IQ on

[Q6] the reaction time scores (such that those with higher IQs had faster reaction times) which in turn should have affected IAT scores (such that those with the higher IQs should have smaller IAT scores). However, as predicted, the use of the *D*-scoring algorithm (Greenwald, Nosek and Banaji, 2003) eliminated this IQ confound (Cai *et al.*, 2004).

Given these results, we developed a 2-stage child–sex IAT that used pictures for the dimension of child versus adult, and words for the concepts or sex versus not sex. We then tested this version, and our original 5-stage version on participants drawn from offenders serving community rehabilitation orders, offenders on licence from prison sentences or on remand in bail hostels. The experimental group consisted of 11 men convicted of sexual offences against children. The control group were 15 other offenders without convictions for sexual offences against children. Both versions of the task showed that the experimental group showed an IAT *D*-score consistent with an association with child and sex, whilst the control group showed an IAT *D*-score consistent with an association with adult and sex. Thus, the results replicated those of Gray *et al.* (2005), extended them to this to sexual offenders living in the community, and showed that the shortened version of the IAT was also effective.

Brown, Gray and Snowden (2008)

It is clear that offenders who have sexually offended against children are not a homogenous group. The motivations to offend are likely to differ between offenders (Beech, 1998) (Marziano *et al.*, 2006). For instance, distinctions have been drawn [Q7] between offenders that are 'pedophilic' – defined by being sexually aroused by

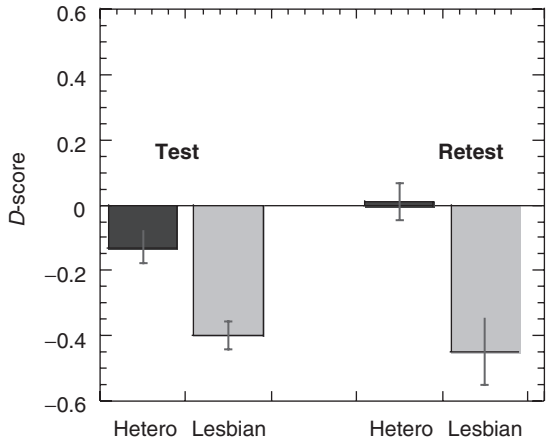

Figure 5.4 Results of the sex–gender IAT in females (from Snowden, Brown and Gray, submitted). *D*-scores below 0 indicate faster performance when female and sex are paired (whilst scores above 0 indicate faster performance when male and sex are paired). Participants were tested twice. Error bars represent ±1 SEM.

stimuli of pre-pubertal children, and offenders that are 'hebephilic' – defined as being aroused by stimuli of adolescents (Blanchard *et al.*, 2001; Freund and Blanchard, 1989). We therefore hypothesized that it would only be the paedophilic offenders that would show the association between children and sex on the IAT. However, there is the question of how one can identify 'pedophiles' and 'hebephiles' so as to test their implicit cognitions using the IAT. One method might be to use PPG measures of sexual arousal, but this was not available to us, and if PPG and IAT were to be just measuring the same thing using different instruments then the results would only be of interest in the sense that either method could be used. We therefore operationalized the 'pedophile' versus 'hebephile' groups by the nature of the victims of their previous offences. Any sex offender with a victim younger than 12 years was termed a 'pedophile', and those with victims only between 12 and 16 years of age were termed 'hebephiles'. We note that this is a loose approximation of the true meaning of these terms, but any errors in classification caused by this operationalization would work against our hypothesis.

We managed to obtain a much larger sample of sex offenders than our previous studies and our groups contained 54 paedophiles, 21 hebephiles and 49 offender controls that had no convictions for child–sex offences. We used our shortened 2-stage IAT. Our IAT results clearly showed that the paedophile group had implicit cognitive associations between sex and children, whilst the offender controls had implicit cognitive associations between sex and adults – replicating our previous results. However, the results from the child–sex offenders that we termed hebephiles did not show associations between children and sex, instead their implicitly measured associations sexual cognitions appeared similar to the offender controls. This shows that the child/sex IAT appears to have the ability to distinguish between the different sexual cognitions that may characterize individual sex offenders, and therefore strengthens the possible use of implicit measurements in the understanding of sexual offenders at both the group and individual level.

We commenced this chapter by outlining the problems of self-report measures, and that in sex offenders this may be a substantial problem due to deliberate dissimulation. So it was a natural question to consider whether our child/sex IAT would be as effective in a sample of sex offenders who denied or minimized their offences and/or sexual attraction to children, compared to those that admitted their offences. Again 'denial' is a multifaceted term that we cannot hope to cover in such a chapter, so we again used a pragmatic definition that was commensurate with practice within the UK prison service. 'Deniers' and 'admitters' were classed according to the assessments of HM Prison Service sex offender treatment program (SOTP). The theory manual for the SOTP in the Unites Kingdom lists certain 'not ready criteria'. These are used to classify prisoners who are 'blocked from benefiting from the programme at the present time, but who could potentially overcome the blockage at some point in the future'. Chief among these is 'in total denial of his sexual offending'. This is assessed in two ways. First, clinicians use a structured interview to assess denial and minimization of the offence. Among other questions, the offender is asked what he found attractive about the child and what sexual acts were engaged in. Second, by using scores from the Sex Offence Attitudes Questionnaire (Hogue, 1994), a 50-item self-report questionnaire designed to assess the degree to which the offence is denied, distorted and minimized. In other words,

the assessment is made as to the degree the offender denies sexual interest in children in general, and his own offences in particular. On the basis of these criteria, the offender is classed as either a non-denier or a denier and is included or excluded from SOTP by the clinician. Given the results from the hebephiles outlined above, the analysis of denial was confined to those with an offence against a child younger than 12 years.

We found that both groups produce scores that were indicative of an association between sex and children, and that we would not distinguish the scores of the deniers from the non-deniers. Hence, we suggest that the child–sex IAT is able to identify distorted implicit cognitions about sex and children even in offenders who are in denial of their paedophilic tendencies.

Nunes, Firestone and Baldwin (2007)

Nunes, Firestone and Baldwin (2007) tested several IATs relating to possible cognitive distortions in a sample of 27 child molesters (defined by having a victim younger than 14 years) in comparison to offenders without convictions for sexual offences (and who denied ever committing such an offence). In total, they tested six IATs. Each of three concepts (pleasant vs. unpleasant, powerful vs. weak, sexy vs. not sexy) was paired with the attributes of me versus not me, and child versus adult. Only one of these IATs revealed a significant between-group difference – the IAT that examined the concepts of child versus adult against the attribute of sexy versus not sexy. They found that the child molesters were faster when the concepts of child and sexy were on the same response key, and the control offenders were faster when the concepts of adult and sexy shared the same key. The results are therefore in accord with those reported by Gray *et al.* (2005).

Nunes, Firestone and Baldwin (2007) also administered an actuarial measure of future risk of sexual offending to each of the child molester group – the Static-99 (Hanson and Thornton, 2000). They found that the child/sex IAT was correlated ($r = 0.43$) with the Static-99 risk, such that those with IAT scores indicating stronger implicit associations between sex and children have greater risk scores for sexual offending. This is a suggestive finding that the IAT score may have some ability to predict future offending, though direct tests of such a hypothesis are needed.

Mihailides, Devilly and Ward (2004)

Mihailides, Devilly and Ward (2004) attempted to test models of implicit cognitions (Ward and Keenan, 1999; Ward, 2000) in sexual offenders using three IATs that related to children as sexual beings, the uncontrollability of sexuality and a sexual-entitlement bias. The first of these IATs is the closest to those so far described. The technique, however, differed to those of our group (and those of Nunes, Firestone and Baldwin, 2007) in that instead of having a child–adult dimension they employed a child–not child dimension (examples of not child words were 'lids', 'rifle' and 'underline'). They found that their experimental group (25 incarcerated male child–sex offenders) showed faster responses to the condition

where the child words shared the same button with the sex words, compared to the combination of child and not sex. The difference between conditions was very large (895 milliseconds). However, this difference was also true for the controls (25 incarcerated other offenders) who showed a difference of 513 milliseconds. The groups were distinguishable in that the difference was larger for the sex offenders than the controls. Thus this was interpreted as a greater association of child and sex in the child–sex offender group.

Interpretation of this paradigm is less clear than the previous studies. The results suggest that all groups tested have a greater association with 'child and sex' than between 'not child and sex', though it is hard to really understand what the association between the 'not child' and 'sex' might have meant to the participant given the exemplars under its umbrella include 'lids', 'rifle' and 'underline'.

Discussion and Conclusions

The above review contains five independent data sets of child–sex offenders versus non-sex offender controls in which a child–sex IAT has been tested. All five show the pattern of results that child–sex offenders have a greater association between child and sex than did the other offender controls. This result was consistent across differences in methodologies across the groups (and within the Gray/Snowden group). The finding therefore appears robust. The child/sex IAT appears to be able to identify those offenders who have a sexual offence against children with an AUC on the order of 0.73–0.81 – figures that are comparable with those produced by PPG measurements.

Limitations

Whilst we have done our best to highlight the possible importance of the IAT in the domain of sexual offending, we would be remiss if we did not address its limitations.

First, the IAT is a relational measure. It requires the comparison of two concepts (e.g. flowers vs. insects or child vs. adult) against an evaluative dimension (e.g. good vs. bad or sex vs. not sex). Thus, the current data tell us that child–sex offenders have greater sexual associations with children than with adults. This could, in theory, be because they are attracted to both (but more so to children), or not attracted to either (but less so to adults – if this makes sense). It would be of value to get more data that specifically examines the absolute associations between children and sex, rather than the association relative to that of sex and adult that our IAT produces. There have been attempts to 'rescore' the results from IAT experiments to try and obtain these 'absolute' associations, but the evidence shows that the reaction times from the IAT cannot be rescored to allow this (Nosek, Greenwald and Banaji, 2005). However, there are now techniques with conceptual similarities to the IAT that aim to measure attitudes to a single concept using implicit measurement (e.g. Go–No-Go task – Nosek and Banaji, 2001; Extrinsic Affective Simon

Task – De Houwer, 2003). We (Snowden, Gray, Brown and Power, unpublished report) have collected some preliminary data using the Go–No-Go task (Nosek and Banaji, 2001) that shows that those with a conviction for child–sex offences show a greater association between children and sex than did other offender controls, and than did a group of rapists.

Second, whilst the IAT (and other implicit measurements) may be able to index an association between the concepts of child and sex, it does not tell us why this has arisen. It is feasible that someone undergoing extensive therapy for their sexual offending might well have such associations as they have to discuss their offending behaviour and listen to others discuss such matters. Whilst this is feasible, we are encouraged that this is not the likely explanation of our results as we have also found similar results in those not undergoing treatment programmes (Brown, Gray and Snowden, 2008), and in those that are now in the community (Brown, 2006). Nevertheless, any individual might have an association between child and sex for any number of reasons (e.g. an expectant mother!).

Third, we do not yet have any data on the predictive validity of the IAT. The methods used so far are only 'predictive' of past behaviour. Studies are needed to establish the IAT's ability in this domain.

CLINICAL USE OF THE IAT

We have suggested that the IAT may become a valuable addition to the clinician's tool bag in the future. It appears to be able to capture cognitions that are not necessarily accessible to the offender themselves (or, indeed, the assessing clinician), and appears to have a greater robustness against dissimulation than self-report measures. As such, we would like to suggest how it might be used in clinical settings. We fully acknowledge that the IAT is still in its infancy and issues of reliability and dissimulation need to be addressed further before it is actually employed in such a manner.

As a Facilitator to Understanding a Person's Cognitions

As outlined in the introduction, people do not always have conscious access to all their mental processes. IATs could be devised to measure cognitive distortions underpinning their offending behaviour (e.g. Marziano *et al.*, 2006; Ward *et al.*, 1997) in order to understand the cognitions held by each individual offender. The flexibility of the IAT gives it advantages over other physiological techniques that require sexual arousal (e.g. PPG – see chapter by Laws) or other specific autonomic responses that are more invasive. Armed with such information, the clinician can make the offender aware of these cognitions, and thus the need for treatment in this area. Of course, we can also rule out the need for specific treatments in some cases and therefore save time, effort and resources providing treatment where none is needed.

[Q8]

As an Indicator of Change

One of the problems faced by those working with sex offenders is how to index the effectiveness of therapy. It is well known that treatment effects are modest in this group (Brooks-Gordon and Bilby, 2006) and currently there is strong reliance on self-report measures to index therapeutic change. What is commonly found is that treatment programmes show large effects on self-report measures, but limited, if any, effects on reoffending rates (leaning some to question the validity of treatment effectiveness as indexed via self-report).

Of interest here is whether the IAT could ever index clinical change. The answer appears to be positive in that Teachman and Woody (2003) show that an IAT designed to measure specific phobias (e.g. spider phobia) showed change as treatment for this phobia progressed, and that this change in the IAT was mirrored by behavioural change in approach to the phobic object. Thus, in at least one domain, the IAT has been shown to index treatment effects. Studies are now required that monitor these cognitions via implicit (and explicit) measures during the course of treatment, and consider the long-term outcome (including sexual recidivism) in such offenders.

CONCLUSIONS

The IAT appears to hold great promise as a tool for investigating cognitions related to sexual offending. It is flexible and can be adapted to measure a variety of concepts. It returns large effect sizes, has good psychometric properties and is relatively hard to dissimulate. It can measure associations that the offender may not be aware of, or ones that they are not willing to explicitly report. We believe, therefore, that the IAT will continue to be a most valuable instrument for research into sexual offending, and we can envisage that it may eventually provide the clinician with a instrument that will help them understand the motivations and sexual interest profile of the offender, inform treatment and management regimes, and monitor the effectiveness of such treatment.

REFERENCES

Alicke, M.D., Klotz, M.L., Breitenbecher, D.L. *et al.* (1995) Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology*, **68**, 804–25.

Arkes, H.R. and Tetlock, P.E. (2004) Attributions of implicit prejudice, or 'would Jesse Jackson "fail" the implicit association test?' *Psychological Inquiry*, **15**, 257–78.

Asendorpf, J.B., Banse, R. and Mucke, D. (2002) Double dissociation between implicit and explicit personality self-concept: the case of shy behavior. *Journal of Personality and Social Psychology*, **83**, 380–93.

Banaji, M.R., Nosek, B.A. and Greenwald, A.G. (2004) No place for nostalgia in science: a response to Arkes and Tetlock. *Psychological Inquiry*, **15**, 279–89.

Banse, R., Seise, J. and Zerbes, N. (2001) Implicit attitudes towards homosexuality: reliability, validity, and controllability of the IAT. *Zeitschrift fur Experimentelle Psychologie*, **48**, 145–60.

Baron, A.S. and Banaji, M.R. (2006) The development of implicit attitudes: evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science*, **17**, 53–8.

Baumeister, R.F. (2000) Gender differences in erotic plasticity: the female sex drive as socially flexible and responsive. *Psychological Bulletin*, **126**, 347–74.

Beech, A.R. (1998) A psychometric typology of child abusers. *International Journal of Sex Offender Therapy and Comparative Criminology*, **42**, 319–39.

Bem, D.J. (1996) Exotic becomes erotic: a developmental theory of sexual orientation. *Psychological Review*, **103**, 320–35.

Blanchard, R., Klassen, P., Dickey, R. *et al.* (2001) Sensitivity and specificity of the phallometric test for pedophilia in nonadmitting sex offenders. *Psychological Assessment*, **13**, 118–26.

Brooks-Gordon, B. and Bilby, C. (2006) Psychological interventions for treatment of adult sex offenders. *British Medical Journal*, **333**, 5–6.

Brown, A., Gray, N.S. and Snowden, R.J. (2008) Implicit measurement of sexual preferences in child sex abusers: role of victim type and denial. *Journal of Abnormal Psychology*, Submitted.

[Q9]

Brown, A.S. (2006) *Developing an Implicit Association Test for Forensic Use: Discriminating Paedophiles from Other Offenders*. Cardiff University, Wales, UK.

Cai, H.J., Sriram, N., Greenwald, A.G. and McFarland, S.G. (2004) The implicit association test's *D* measure can minimize a cognitive skill confound: comment on McFarland and Crouch (2002). *Social Cognition*, **22**, 673–84.

Chivers, M.L. and Bailey, J.M. (2005) A sex difference in features that elicit genital response. *Biological Psychology*, **70**, 115–20.

Chivers, M.L., Rieger, G., Latty, E. and Bailey, J.M. (2004) A sex difference in the specificity of sexual arousal. *Psychological Science*, **15**, 736–44.

Cohen, J. and Cohen, P. (1982) *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 2nd edn, Erlbaum, Hillsdale, NJ.

Cooper, S. (2005) Understanding treating and managing sex offenders who deny their offence. *Journal of Sexual Aggression*, **11**, 85–94.

Cowey, A. (2004) The 30th Sir Frederick Bartlett lecture: fact, artefact, and myth about blindsight. *Quarterly Journal of Experimental Psychology. Section A – Human Experimental Psychology*, **57**, 577–609.

Craeynest, M., Crombez, G., De Houwer, J. *et al.* (2005) Explicit and implicit attitudes towards food and physical activity in childhood obesity. *Behaviour Research and Therapy*, **43**, 1111–20.

Cvencek, D. and Greenwald, A.G. (submitted) Faking of the Implicit Association Test is statistically detectable.

[Q10]

De Houwer, J. (2001) A structural and process analysis of the implicit association test. *Journal of Experimental Social Psychology*, **37**, 443–51.

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

De Houwer, J., Crombez, G., Koster, E.H.W. and De Beul, N. (2004) Implicit alcohol-related cognitions in a clinical sample of heavy drinkers. *Journal of Behavior Therapy and Experimental Psychiatry*, **35**, 275–86.

De Houwer, J., Custers, R. and De Clercq, A. (2006) Do smokers have a negative implicit attitude toward smoking? *Cognition & Emotion*, **20**, 1274–84.

de Jong, P.J. (2002) Implicit self-esteem and social anxiety: differential self-favouring effects in high and low anxious individuals. *Behaviour Research and Therapy*, **40**, 501–8.

Diamond, L.M. (2003) What does sexual orientation orient? A biobehavioral model distinguishing romantic love and sexual desire. *Psychological Review*, **110**, 173–92.

Dovidio, J.F., Kawakami, K., Johnson, C. *et al.* (1997) On the nature of prejudice: automatic and controlled processes. *Journal of Experimental Social Psychology*, **33**, 510–40.

Egloff, B. and Schmukle, S.C. (2002) Predictive validity of an implicit association test for assessing anxiety. *Journal of Personality and Social Psychology*, **83**, 1441–55.

Fazio, R.H., Jackson, J.R., Dunton, B.C. and Williams, C.J. (1995) Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *Journal of Personality and Social Psychology*, **69**, 1013–27.

Fiedler, K. and Bluemke, M. (2005) Faking the IAT: aided and unaided response control on the Implicit Association Tests. *Basic and Applied Social Psychology*, **27**, 307–16.

Freund, K. and Blanchard, R. (1989) Phallometric diagnosis of pedeophilia. *Journal of Consulting and Clinical Psychology*, **57**, 100–105.

Friese, M., Wänke, M. and Plessner, H. (2006) Implicit consumer preferences and their influence on product choice. *Psychology & Marketing*, **23**, 727–40.

Gawronski, B. and Bodenhausen, G.V. (2005) Accessibility effects on implicit social cognitions: the role of knowledge activation versus retrieval processes. *Journal of Personality and Social Psychology*, **89**, 672–85.

Gawronski, B., LeBel, E.P. and Peters, K.R. (2007) What do implicit measures tell us? *Perspectives on Psychological Science*, **2**, 181–93.

Gray, N.S., Brown, A.S., MacCulloch, M.J. *et al.* (2005) An implicit test of the associations between children and sex in pedophiles. *Journal of Abnormal Psychology*, **114**, 304–8.

Gray, N.S., MacCulloch, M.J., Smith, J. *et al.* (2003) Violence viewed by psychopathic murderers. *Nature*, **423**, 497–8.

Greenwald, A.G. and Farnham, S.D. (2000) Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, **79**, 1022–38.

Greenwald, A.G., McGhee, J.L. and Schwartz, J.L. (1998) Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, **74**, 1464–80.

Greenwald, A.G., Nosek, B.A. and Banaji, M.R. (2003) Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, **85**, 197–216.

Hanson, R.K. and Thornton, D. (2000) Improving risk assessments for sex offenders: a comparison of three actuarial scales. *Law and Human Behavior*, **24**, 119–36.

Hess, E.H. and Polt, J.M. (1960) Pupil size related to interest value of visual stimuli. *Science*, **132**, 349–50.

Hofman, W., Gawronski, B., Gschwendner, T. *et al.* (2005) A meta-analysis on the correlation between the Implicit Association Test and the explicit self-report measures. *Personality and Social Psychology Bulletin*, **31**, 1369–85.

Hogue, T.E. (1994) Sex offence information questionnaire: assessment of sexual offenders' perceptions of responsibility, empathy and control. *Issues in Criminology and Legal Psychology*, **21**, 68–75.

Horswill, M.S., Waylen, A.E. and Tofield, M.I. (2004) Driver's rating of different components for their own driving skill: a greater illusion of superiority for skills that relate to accident involvement. *Journal of Applied Social Psychology*, **34**, 177–95.

Huijding, J., de Jong, P.J., Wiers, R.W. and Verkppijen, K. (2005) Implicit and explicit attitudes toward smoking in a smoking and non-smoking setting. *Addictive Behaviors*, **30**, 949–61.

Karpinski, A., Steinman, R.B. and Hilton, J.L. (2005) Attitude importance as a moderator of the relationship between implicit and explicit attitude measures. *Personality and Social Psychology Bulletin*, **31**, 949–62.

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

Lippa, R.A. (2006) Is high sex drive associated with increased sexual attraction to both sexes? *Psychological Science*, **17**, 46–52.

Lippa, R.A. and Arad, S. (1997) The structure of sexual orientation and its relation to masculinity, femininity, and gender diagnosticity: different for men and women. *Sex Roles*, **37**, 187–208.

Luo, Q., Nakic, M., Wheatley, T. *et al.* (2006) The neural basis of implicit moral attitude – an IAT study using event-related fMRI. *NeuroImage*, **30**, 1449–57.

Marziano, V., Ward, T., Beech, A.R. and Pattison, P. (2006) Identification of five fundamental implicit theories underlying cognitive distortions in child abusers: a preliminary study. *Psychology Crime & Law*, **12**, 97–105.

McConnell, A.R. and Leibold, J.M. (2001) Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attributes. *Journal of Experimental Social Psychology*, **37**, 435–42.

Mihailides, S., Devilly, G.J. and Ward, T. (2004) Implicit cognitive distortions and sexual offending. *Sexual Abuse: A Journal of Research and Treatment*, **16**, 333–50.

Nock, M.K. and Banaji, M.R. (2007) Assessment of self-injurious thoughts using a behavioral test. *American Journal of Psychiatry*, **164**, 820–23.

Nosek, B.A. (2005) Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, **134**, 565–84.

Nosek, B.A. (2007) Implicit–explicit relations. *Current Directions in Psychological Science*, **16**, 65–9.

Nosek, B.A. and Banaji, M.R. (2001) The go/no-go association task. *Social Cognition*, **19**, 625–64.

Nosek, B.A., Banaji, M.R. and Greenwald, A.G. (2002) Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, **6**, 101–15.

Nosek, B.A., Greenwald, A.G. and Banaji, M.R. (2005) Understanding and using the implicit association test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, **31**, 166–80.

Nosek, B.A., Greenwald, A.G. and Banaji, M.R. (2007) The implicit association test at age 7: a methodological and conceptual review, in *Automatic Processes in Social Thinking and Behavior* (ed. J.A. Bargh), Psychology Press, New York, pp. 265–92.

[Q11]

Nunes, K.L., Firestone, P. and Baldwin, M.W. (2007) Indirect assessment of cognitions of child sexual abusers with the implicit association test. *Criminal Justice and Behavior*, **34**, 454–75.

Olson, M.A. and Fazio, R.H. (2003) Relations between implicit measures of prejudice: what are we measuring? *Psychological Science*, **14**, 636–9.

Olson, M.A., Fazio, R.H. and Hermann, A.D. (2007) Reporting tendencies underlie discrepancies between implicit and explicit measures of self-esteem. *Psychological Science*, **18**, 287–291.

Phelps, E.A., O'Conner, K.J., Cunningham, W.A. *et al.* (2000) Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, **12**, 729–38.

Richeson, J.A., Baird, A.A., Gordon, H.L. *et al.* (2003) An fMRI investigation of the impact of interracial contact on executive function. *Nature Neuroscience*, **6**, 1323–8.

Richetin, J., Perugini, M., Prestwich, A. and O'Gorman, R. (2007) The IAT as a predictor of food choice: the case of fruits versus snacks. *International Journal of Psychology*, **42**, 166–73.

Snowden, R.J., Brown, A.S. and Gray, N.S. (submitted) Implicit sexual associations in heterosexual women and lesbians.

[Q12]

Snowden, R.J., MacCulloch, M.J., Smith, J. *et al.* (2004) Implicit affective associations to violence in psychopathic murderers. *Journal of Forensic Psychiatry and Psychology*, **15**, 620–41.

Snowden, R.J., Wichter, J. and Gray, N.S. (2007) Implicit and explicit measurements of sexual preference in gay and heterosexual men: a comparison of priming techniques and the implicit association task. *Archives of Sexual Behavior*, **37** (4), 558–65.

[Q13]

Steffens, M.C. (2004) Is the implicit association test immune to faking? *Experimental Psychology*, **51**, 165–79.

Strack, F. and Deutsch, R. (2004) Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, **8**, 220–47.

Svenson, O. (1981) Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, **47**, 143–8.

Teachman, B.A. and Brownell, K.D. (2001) Implicit anti-fat bias among health professionals: is anyone immune? *International Journal of Obesity*, **25**, 1525–31.

Teachman, B.A., Gregg, A.P. and Woody, S.R. (2001) Implicit associations for fear-relevant stimuli among individuals with snake and spider fears. *Journal of Abnormal Psychology*, **110**, 226–35.

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

Vanman, E.J., Saltz, J.L., Nathan, L.R., and Warren, J.A. (2004) Racial discrimination by low-prejudiced whites – facial movements as implicit measures of attitudes related to behavior. *Psychological Science*, **15**, 711–4.

Vartanian, L.R., Polivy, J. and Herman, C.P. (2004) Implicit cognitions and eating disorders: their application in research and treatment. *Cognitive and Behavioral Practice*, **11**, 160–67.

Ward, T. (2000) Sexual offenders' cognitive distortions as implicit theories. *Aggression and Violent Behavior*, **5**, 491–507.

Ward, T., Hudson, S.M., Johnston, L. and Marshall, W.L. (1997) Cognitive distortions in sex offenders: an integrative review. *Clinical Psychology Review*, **17**, 479–507.

Ward, T. and Keenan, T. (1999) Child molesters' implicit theories. *Journal of Interpersonal Violence*, **14**, 821–38.

Wiers, R.W., Van Woerden, N., Smulders, F.T.Y. and De Jong, P.J. (2002) Implicit and explicit alcohol-related cognitions in heavy and light drinkers. *Journal of Abnormal Psychology*, **111**, 648–58.

Wilson, T.D. and Dunn, E.W. (2004) Self-knowledge: its limits, value, and potential for improvement. *Annual Review of Psychology*, **55**, 493–518.

Winkielman, P. and Berridge, K.C. (2004) Unconscious emotion. *Current Directions in Psychological Science*, **13**, 120–23.

Ziegert, J.C. and Hanges, P.J. (2005) Employment discrimination: the role of implicit attitudes, motivation, and a climate for racial bias. *Journal of Applied Psychology*, **90**, 553–62.

Queries

Q1: We have set the text ‘All human images … publication in print’ as chapter title footnote and ‘The author … the United States’ as author footnote for stylistic purposes. Could you please confirm that it is OK? Also, the second footnote, ‘The author … the United States’, is talking about only one author, whereas the chapter has two authors. Could you please specify for which author this note has been given?

Q2: Kindly check the hierarchy of headings in this chapter (<H1>: first-order head; <H2>: second-order head).

Q3: Would it not be better to mention here the specific chapter, as the mentioned author (Laws) has more than one chapters in this book? Kindly suggest.

Q4: You have cited reference ‘Olson (2003)’ here but you have not given this reference in the reference list. Could you please provide the details of this reference to be included in the reference list?

Q5: Does Figure 5.4 mean here in place of Figure 5.3 in this paragraph? If no, kindly provide the cross-citation of Figure 5.4, as this figure has not been cited anywhere in the text.

Q6: We have spelt out the term ‘RT’ as ‘reaction time’. Could you please confirm that it is OK?

Q7: British spellings have been used in this chapter, whereas the terms ‘pedophilic’ and ‘pedophile’, which are American spellings, have been used within the quotes. Would it not be better to change these terms in ‘paedophilic’ and ‘paedophile’ to make the spellings consistent throughout the chapter. Kindly suggest as appropriate.

Q8: Would it not be better to mention here the specific chapter, as the mentioned author (Laws) has more than one chapters in this book? Kindly suggest.

Q9: Kindly update the reference ‘Brown *et al.* (2008)’, if possible.

Q10: Please update the reference ‘Cvencek and Greenwald (submitted)’, if possible.

Q11: We have replace ‘****’ appearing in the reference ‘Nosek, Greenwald and Banaji (2007)’ with the author’s name (J.A. Bargh). Would you please confirm that it is OK?

Q12: Kindly update the reference ‘Snowden *et al.* (submitted)’, if possible.

Q13: We have located volume number (37), issue number (4) and page range (558--65) on PubMed search and added those in the reference ‘Snowden *et al.* (2007)’. Could you please check that it is OK. Also, please ensure that the year in the same reference is correct as given, or should it be ‘2008’.