The Use of Implicit Cognitive Measures in the Assessment of Sexual Offenders

By

Caoilte Ó Ciardha

A dissertation submitted for the degree of Doctor of Philosophy of the
University of Dublin, Trinity College, Dublin 2, Ireland

2010

Declaration

I hereby declare that this thesis has not been submitted as an exercise for a degree at this or any other University, that it is entirely my own work and that the contributions of others are duly acknowledged in the text, wherever included. I give permission to the Library to lend or copy the thesis upon request, subject to normal conditions of acknowledgement.

Signed: _____

                 Caoilte Ó Ciardha

Date: _____

Summary

Not all sexual offenders against children have a sexual interest in children. Establishing an individual offenders' degree of sexual interest in children is important as it has implications for risk of re-offence and for treatment. In addition some offenders have distorted cognitions (or implicit theories) surrounding children and their understanding of and willingness to engage in sexual activity. The relationship of such implicit theories and deviant sexual interest has not been adequately explored in the theoretical literature. A reason for this is a lack of available procedures to effectively measure the cognitive structure and processes relating to sexual interests and cognitive distortions relating to sex. This thesis explores the potential utility of two such tasks, a pictorial modified Stroop task using potentially sexually salient stimuli and the Implicit Association Test (IAT). It is hypothesised that the IAT is a measure of schematic associations and could therefore potentially correspond to the cognitive distortions of offenders. The pictorial Stroop task on the other hand was hypothesised to measure an attentional component of the sexual arousal response. Two studies using non-offending control participants demonstrated that both the IAT and the pictorial Stroop task could tap into sexual interest. Both measures were significantly correlated in both studies. The results of the tasks were taken to indicate that while the tasks may measure distinct cognitive processes, both were indirectly measuring sexual orientation.

When versions of both tasks were used with a sample of sexual offenders against children there was no agreement between the tasks. The pictorial Stroop task was again demonstrated to tap into sexual interest towards adults, and showed promise in demonstrating group differences in response times to child stimuli between offenders considered likely to have deviant sexual interest compared with other offenders and control participants. Both IATs used, one exploring gender-sex associations, the other age-sex associations, demonstrated an effect of order that is likely to have masked any clear influence of sexual schema on the results. A second study involving offending participants, conducted in a separate institution, again found the pictorial Stroop task to tap into sexual interest towards adult stimuli (for those cases where a clear indication of sexual orientation was available) but was unable to find group differences between rapists and child molesters in their responses. Additionally the pictorial Stroop task used

in that study correlated with arousal levels as measured by penile plethysmography when a gender preference index was calculated but not using an age preference index. Methodological differences between the pictorial Stroop task used in that study and the pictorial Stroop tasks in the rest of the thesis make a direct comparison between the measures problematic.

To explore methodological questions relating to the pictorial modified Stroop task a study was designed where three versions of the task were compared. The study found that a pictorial Stroop task where images of each trial type were grouped together in a single large block significantly outperformed a version of the task using smaller 'clusters' of matching stimuli and that both outperformed a completely random version of the task. These results suggested that higher order rumination relating to the stimulus types was driving the pictorial Stroop effect. A further study found that a Gender-Sex IAT that used pictures of men and women instead of names to illicit the gender categories and that used a *nonsexual* versus *sexual* contrast rather than a *sex* versus *furniture* (neutral) contrast yielded better discrimination between gay and straight participants. This final study also found that a choice reaction time (CRT) paradigm adopting a blocked design bore no relationship to a blocked pictorial Stroop task (or to the Gender-Sex IAT). These results were taken to indicate that a blocked design is not appropriate for a CRT task and that the task may measure a different component of the arousal process to the blocked pictorial Stroop task.

Taken as a whole the thesis found that while the Implicit Association Test and the pictorial modified Stroop task are both paradigms that offer clear potential for use in the assessment of sexual offenders, both need to be further tested and validated. Future research should test the methodological recommendations made in this thesis. Additionally a clear understanding of the processes measured by each task should be established so that the emerging field involving cognitive or indirect approaches to forensic assessment can establish clear best practices and a framework by which to explore the cognitive processes of sexual offenders.

Acknowledgements

I would like to thank my supervisor, Michael Gormley, whose guidance throughout this project was hugely supportive and his comments always insightful.

Richard Laws, David Thornton, Carmen Gress and Kevin Nunes have all, whether they realise it or not, been instrumental in shaping my understanding of what exactly I was exploring with this thesis. Likewise, Dave Gogan, Leslie Helmus, Kelly Babchishin, Katriona O'Sullivan, Ashling Bourke, Rosaleen McElvaney and Rick McKee have offered me invaluable help and advice. I'd like to thank my academic appraisers Mac MacLachlan and Hugh Garavan not just for their helpful comments but for their implicit reassurance that I was on the right track. I thank David Hevey for his constant willingness to answer my progressively stranger statistics enquiries. Anyone who doesn't judge you when you sneak into their office to steal books on statistical power is a true friend.

I owe a huge debt of gratitude to Jeanine De Volder, not only for helping me access participants when I wasn't sure if I would be able to get the project off the ground but for also taking a genuine interest in my research and my wellbeing while I conducted it. The rest of the staff, both past and present, at the Granada Institute were always welcoming and supportive. I would especially like to thank Pat Walsh and Niall Muldoon for their help. I would also like to thank the staff of the Irish Prison Service who helped me so efficiently to get ethical approval, access, and ultimately, to recruit participants. I am grateful to Paul Murphy, Governor Liam Dowling and especially Esther Lonergan for their help.

I would like to thank those people in the School of Psychology who helped make everything run smoothly and who generally made my life as a postgrad a lot easier; in particular I'm grateful to Luisa, Michelle, Pat, Eddie, Siobhán, Rose, Lisa and June for their help and friendship over the years.

I would like to thank all those postgrads, lecturers and researchers who have made my time in Trinity so enjoyable. I thank you all wholeheartedly. You know who you are;

you could usually be found in the staffroom from about half past eleven, in O'Neill's after work on a Friday or on the Botany Bay soccer court long after the ball should have been returned to the sports hall. In particular I'd like to thank my numerous officemates, especially Justin Gregg and James Dixon. Perhaps no-one had to suffer through more of my half-cocked theories over my time here than Justin and James and for that I thank them both. I also thank Justin for showing the rest of us how to do a PhD: quickly, competently and in a wetsuit. I thank James for being a close friend throughout and for constantly bringing out my competitive side.

To those who added months onto the length of this PhD I owe my sanity at the end of the process. Their distractions came in many welcome guises. Late nights, road trips, fishing and golf came courtesy of Feargal Chambers and Ciarán Ó Brian, usually disguised as 'working from home'. For funny emails, gossip and hours wasted in bizarre corners of the internet I thank Darragh, Ciarán, Andy and Dessie. To Áine, Eimear, Eamonn, Becky, Johanna, Caroline, Sinéad, Jane, Sarah, Aoife, Aoileann, Doreen; you all know how much you have helped me at different points of this process.

I would like to thank the participants who took part in my study.

Ach go háirithe, ba mhaith liom mo bhuíochas ó chroí a ghabháil le mo chlann, a thug tacaíocht, cabhair agus comhairle fhial dom agus mé ag dul i ngleic leis an ábhar seo; Sally, Cairbre, Darach, Osgur, Seán agus Mamó, is daoibh-se a thiomnaím an tráchtas seo.

Table of Contents

List of Figures

6

List of Tables

Chapter 1: Introduction

This thesis explores the utility of several indirect or implicit cognitive tasks in the assessment of sexual interest and sexual associations. Specifically the thesis attempts to investigate the utility of such tasks in the assessment of sexual offenders. Chapter 1 outlines previous relevant research and theory and hypothesises what cognitive processes or structures are implicated in participants' responses to these tasks. Chapter 2 further examines some methodological issues pertinent to the research reported in the thesis.

The empirical chapters in the thesis are presented chronologically. In chapter 3, the Pictorial modified Stroop task and the Implicit Association Test are piloted to establish their potential utility and to determine what methodology and analysis might best suit the testing of offenders. Chapter 4 reports the findings of the administration of the tasks to offenders. Chapters 5 and 6 report the results of two studies carried out with non-offenders that further explore the questions of how best to carry out these types of tasks and also what exactly is being measured by the tasks. Chapter 7 brings together the findings of all the empirical chapters and discusses those findings in light of the hypotheses made in the first chapter.

1.1   Implicit Cognition: an emerging field within applied psychology

Implicit cognition refers to cognitive processes that are not consciously accessible to the individual through self report or introspection. Such processes interact with more conscious, deliberate cognitions to produce behaviour. Therefore, how people behave is a product, not only of how they know themselves to be but also of biases/beliefs/schemas that they are not aware of. The identification of implicit cognitive processes can be carried out through the comparison of self-reported attitudes etc. and related behaviour or by the use of what are known as 'implicit measures'. The use of self-report is problematic. When the results of a self-report measure is at odds with perceived behaviour it is not clear whether the individual was presenting themselves in a certain light but aware of their underlying attitude etc. or whether there truly was a conflict between that person's implicit and explicit attitudes. Similarly the

use of implicit measures to 'prove' the existence of implicit cognition is problematic. An implicit measure that yields a result that is predictive of behaviour but at odds with self report may simply be identifying the same presentation bias in self report. Therefore it remains unclear in this way whether the biases/associations etc. identified by implicit methods were available to introspection but 'censored' at self report. The argument against this interpretation arises out of the ability of some implicit methods to identify biases that are at odds with self report in situations where there would be no stigma attached to admitting such a bias, i.e. the individual is not strongly motivated to disguise their true cognitions.

De Houwer and others veer away from the use of the word implicit when talking about these measures or tasks in favour of the word indirect. This avoids the implied assumption when using the word implicit that the tasks are measuring cognitions that are truly beyond the individual's conscious awareness. To an extent, it is sufficient for this thesis to assume simply that these measures may have the potential to tap into cognitions that an individual cannot *or* will not access through introspective methods. The thesis uses methods that are argued to access both participants' sexual interests and certain schemas in an indirect manner. The degree to which the cognitions measured are available to introspection, and therefore whether they are part of a true implicit cognition, is beyond the scope of this dissertation. The dissertation will instead focus on the use of implicit or indirect measures of assessment.

1.2   Theories of sexual offending

Ward, Polaschek and Beech (2006) in *Theories of Sexual Offending* outline and organise most of the major theories that have been posited on sexual offending. They divide these theories into levels constituting different levels of explanatory sophistication. Level I theories are multifactorial theories that seek to provide a unified account of sexual offending. These theories attempt to identify the core features of offenders[1] and to construct a theory that looks at causes of those features and attempts to explain how they result in abusive behaviour. The theories that are placed in this

---

[1] The term offence and its derivatives are used throughout the text to refer to sexual offences against children. Offences such as rape against adults and non-sexual offences/offenders etc. are explicitly identified.

level are Finkelhor's precondition model, Marshall and Barbaree's integrated model, Hall and Hirschman's quadripartite model, Ward and Siegert's pathways model, Malamuth's confluence model of sexual aggression along with theories that look at sexual offending from an evolutionary perspective (e.g. Thornhill & Palmer, 2000). Level II theories represent single-factors that typically could be individual components of the more extensive level I theories. For example theories of cognitive distortions would expand upon the cognitive distortion component identified by most if not all of the level I theories. Other examples of level II theories, according to Ward, Polaschek and Beech (2006), include theories of deficient victim empathy, theories of deviant sexual preferences, feminist theories of child sexual abuse, theories of intimacy deficits and theories of arising out of the assessment of risk. Level III theories are micro-models that seek to explain, using multiple factors, specific offence related behaviours. These include, according to Ward, Polaschek and Beech (2006), relapse prevention and self-regulation models along with offence chain, offence cycle and offence process models. In addition to these three theory levels, the authors look at theories based on treatment and classification. Their review sets down the groundwork for creating what they consider is a coherent, adequate, consistent and 'fertile' unified theory of sexual offending. They consider a fertile theory to be one that opens up new avenues of enquiry. Ward and Beech have gone on to attempt to develop such a theory, through a process of theory knitting, where useful and complimentary aspects of many different theories are 'knitted' together; they refer to the outcome of this process as an Integrated Theory of Sexual Offending (Ward & Beech, 2006, 2008). In terms of level I theories Ward and Beech (2006, 2008) incorporate much of the strengths of other theories, in particular those of Finkelhor, Hall and Hirschman; Marshall and Barbaree; and Ward and Siegert, in their Integrated Theory. This introduction will outline the framework of the theory, without much reference to older theories, and will then focus on the areas of cognitive distortion and sexual interest in particular.

1.2.1   Ward and Beech's Integrated Theory of Sexual Offending

Figure 1.1 outlines schematically the Integrated Theory of Sexual Offending (ITSO, Ward & Beech, 2006, 2008). The theory still needs to be fleshed out but it is one that shows potential since it draws on disparate fields of research to inform our understanding of sexual offending. The framework proposed suggests that the clinical

symptoms (or state factors) i.e., emotional problems, social difficulties, cognitive distortions and deviant arousal, that may be seen in offenders and that directly result in offending behaviour (Ward & Beech, 2008) arise out of underlying problems in neuropsychological functioning. These underlying problems are envisaged as originating in three neurological systems (the motivational/emotional; action selection and control; and perception and memory systems). Ward and Beech (2006) draw on Pennington (2002) for the proposed functioning and neuroanatomical structure of these systems. Problems in underlying neuropsychological functioning are posited to be a result of abnormal brain development and/or what Ward and Beech (2006) refer to as an ecological niche. Brain development is seen as a result of evolution and genetics, and the neurobiological infrastructure that results is implicated in the neuropsychological functioning already mentioned. 'Ecological niche' encompasses the social, cultural, and personal circumstances of the individual along with the physical environment in which they find themselves and impacts on neuropsychological functioning primarily through a process of social learning (Ward & Beech, 2006, 2008). The theory also states that these ecological variables can function as a trigger for offending behaviour and that once triggered offending results in a positive feedback loop to entrench the individual's vulnerabilities and thus maintains and/or escalates the offending behaviour (Ward & Beech, 2008). While, as stated, the framework still needs to be fleshed out, the theory currently demonstrates, using Ward and Beech's (2008) criteria, strong unifying power in accommodating many existing theories of offending and drawing on varied disciplines. It also has value in being a fertile source of future research. It is perhaps unfair to be critical of a theory that is in its infancy and that has tried to draw together many of the existing theories into a coherent whole. However, as with previous theories, the theory lacks an explanation of how deviant interest is represented in the brain and also how that interacts with other state factors such as distorted thinking. At this early stage of development the ITSO may be more of a framework for investigation rather than a functioning theory.
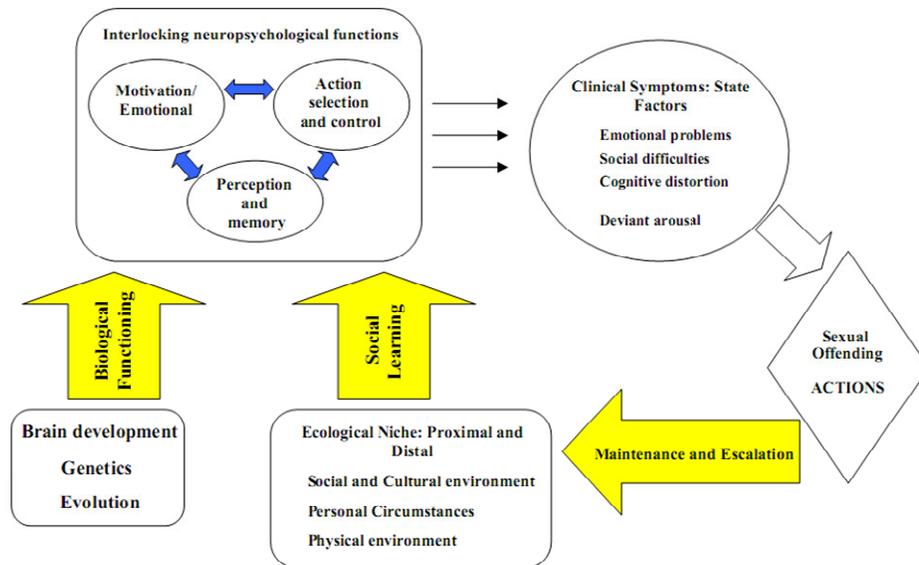
Figure 1.1: Integrated Theory of Sexual Offending. Reproduced from Ward and Beech (2006) with permission.

1.3 Cognitive distortions, schemas and implicit theories

Cognitive distortions are a key component of virtually all modern theories of sexual offence aetiology (Gannon, Ward, & Collie, 2007). Ward and Beech (2006) see cognitive distortions as clinical symptoms or state factors, related to the committing of sexual offences in many cases. Cognitive distortions were first proposed as a mechanism underpinning child abuse by Gene Abel and his colleagues in the 1980s (e.g. Abel, Becker, & Cunningham-Rathner, 1984; Abel, Gore, Holland, & Camp, 1989). Definitions of cognitive distortions in the context of sexual offending have been, and, in some cases, continue to be vague. For example Bumby states: "Cognitive distortions related to sexual offending are learned assumptions, sets of beliefs, and self-statements about deviant sexual behaviors such as child molestation…which serve to deny, justify, minimize, and rationalize an offender's actions (Bumby, 1996, p. 38)". As a definition, this leaves the term of cognitive distortion so broad that it can apply, and is applied, to attitudes and thought processes that are offence-supportive, and play a causal role in offending, or processes that are a post-hoc rationalisation or neutralisation of the offence (Blake & Gannon, 2008; Gannon & Polaschek, 2006; Maruna & Mann, 2006). This results in a situation where the term's intended meaning often remains ambiguous (Blake & Gannon, 2008). A more useful way to consider the cognitions that are

12

involved in offending may be to talk in terms of cognitive structures (e.g. schemas, implicit theories), cognitive processes or operations (e.g. information processing) and cognitive products (e.g. self-statements, beliefs and attributes)(Ward, Hudson, Johnston, & Marshall, 1997). It is these cognitive products that are the state factor cognitive distortions referred to by Ward and Beech (2006).

There has been quite a large amount of research to date attempting to measure and identify cognitive products held by offenders. In addition, one of the most effective approaches to treatment to date has come from the school of Cognitive Behavioural Therapy, which assumes a key role of cognition in behaviour. However, there has, until recently, been a lack of theory-driven research looking at offenders' cognitions, especially when it comes to differentiating products from processes and structures (Mann & Beech, 2003; Ward et al. 1997). Polaschek and Gannon (2004) state that: "Until recently, researchers focused on measurement of surface cognitions at the expense of developing an understanding of the underlying architecture responsible for generating and organizing them (p. 300)". Ward hypothesises that cognitive structures emerge from an offender's early developmental experiences (i.e., trait factors), while the cognitive products (state factors) are produced by an interaction of those cognitive structures and environmental factors (2000). As already stated, Ward and Beech later flesh out this interaction in terms of neurobiological and neuropsychological functioning along with ecological factors (Ward & Beech, 2006). Much of the recent literature has focused on the fact that cognitive distortions as a term or concept may be too all-encompassing and need to be elucidated. The literature has especially tried to address the difficulty of trying to understand to what extent offence-supportive statements are indicative of deep-rooted faulty belief structures that have a causal role in offending and to what extent they are as a result of post-hoc minimisations, justifications, excuses etc. (Blake & Gannon, 2008; Gannon, Ward & Collie, 2007; Mann & Beech, 2003; Mann, Webster, Wakeling, & Marshall, 2007; Maruna & Mann, 2006; Ward, Keown, & Gannon, 2007). Given that many current treatment programs emphasise the identification and challenging of cognitive distortions, Maruna and Mann (2006) question whether all of these so called cognitive distortions are problematic (and therefore should be a focus for treatment). They ask whether some of what are sometimes considered distorted cognition (especially justifications and excuse making) is not an attempt to reconcile in a normative way the fact that the offender considers

themselves an inherently good person albeit one who has done bad things. In addition they make the point that some of the excuses offered by individuals for criminal behaviour may have in fact contributed to their offences and are thus not necessarily the product of sloppy or distorted thinking (e.g. drug and alcohol use, family upbringing etc.). They also suggest that some degree of excuse-making may indicate less risk of recidivism (Maruna & Mann, 2006) since to make excuses may indicate an acceptance of society's norms regarding the inappropriateness of the offending behaviour. To compel someone, as seems to be suggested by some approaches to treatment, e.g. Salter (1988) and other examples cited in Maruna and Mann (2006), to say of themselves "I am someone who carried out criminal acts because I wanted to" as opposed to "I am someone who carried out criminal acts for a number of reasons, some of which are to do with how I see the world and some of which are to do with situational factors" may not afford the offender with the best opportunity to avoid recidivism.

In order to attempt to understand the phenomenon of faulty cognition among offenders researchers have tried to place 'cognitive distortions' in the broader framework of social cognition and the theory underlying cognitive therapy. Specifically they have looked at certain cognitions held by offenders as schemas. These schemas are cognitive heuristics employed by individuals to interpret their social environment. They "are the structures within memory that guide our attention, inform our perceptions, prompt our inferences, and save us energy by providing shortcuts to interpreting social situations. Thus, schemas are structures, but they affect our cognitive processes" (Mann & Beech, 2003, p. 139). Three theories in particular have used schemas to attempt to explain the presence of cognitive distortions among offenders: Ward's *implicit theories* (Ward, 2000; Ward & Keenan, 1999); Mann and Beech's *schema-based model* (Mann & Beech, 2003); and Ward, Gannon and Keown's update of the implicit theories model, the *judgement model of cognitive distortions* (Ward, Gannon, & Keown, 2006; Ward et al., 2007).

1.3.1   Schema-based theories of cognitive distortion

Schemas represent 'deep' cognition and may only be able to be assessed indirectly through cognitive products (Kwon & Oei, 1994; cited in Ward, Polaschek & Beech, 2006). Ward (2000) argued that if some of the cognitive distortions identified by

14

previous research are in fact cognitive products, then they must be products of underlying schema. However since the definitions of schema can be quite broad, Ward instead regards these schema as causal (or implicit) theories to emphasise the fact that he believes they behave similarly "to scientific theories, that are used to explain, predict, and interpret interpersonal phenomena (Ward, 2000, p. 494)." Ward and Keenan (1999) introduced five of these implicit theories that they see as schemas that may be present among offenders against children. These theories were developed by condensing distortions from various measures (Abel et al., 1984; Bumby, 1996; Hanson, Gizzarelli, & Scott, 1994), interview studies (Neidigh & Krop, 1992; Ward, Fon, Hudson, & McCormack, 1998) and a review paper (Ward et al., 1997)

1. *Children as sexual beings*: the individual hypothesises children to be inherently sexual creatures who can enjoy, and even seek out, sex with adults. Innocent childhood behaviours may, thus, be misinterpreted as sexual by someone who holds this distortion.

2. *Nature of harm*: sexual contact between children and adults is seen as harmless unless significant physical violence and/or injury is involved.

3. *Uncontrollability*: individuals who hold this implicit theory believe that events just happen, and that behaviour is dictated mainly by powerful urges and emotions. Offenders, as a consequence do not believe they have control over their own actions.

4. *Entitlement*: the offender's needs are more important than anyone else's and therefore he[2] is entitled to use inferior individuals such as children to satisfy those needs, including sexual needs.

5. *Dangerous world*: the world is a menacing, excessively hostile place, teeming with abusive and exploitative people. There are two variants of this implicit theory: (1) offenders perceive all individuals, including children, as hostile and rejecting, and attempt to control them through sexual abuse, (2) they perceive adults but not children to be dangerous. By comparison, children represent a safe haven and therefore are a preferred choice for a sexual partner (Ward, Polaschek & Beech, 2006).

---

[2] For the most part offenders are referred to in this thesis as male. The vast majority of sexual offenders are male (Craig, Browne, & Beech, 2008) therefore the vast majority of research in the area has focused on males and the majority of research drawn upon in this text relates to male sex offenders. While the potential utility of using the measures outlined in this text with females is an important question, it is one that is beyond the scope of this study.

These five child molester implicit theories accounts for many of the cognitive products identified by other researchers. For example they account for the majority of 38 'cognitive distortions' identified by Neidigh and Krop (1992) through the use of an open-ended questionnaire. It should be noted that Keenan and Ward (2003) state that any given sex offender is only likely to hold a subset of implicit theories. In a qualitative analysis of interviews, Marziano, Ward, Beech and Pattison (2006) found evidence for all five implicit theories in the majority (18 participants were judged to hold the five distortions with the remaining 4 participants holding four) of their heterogeneous but small sample of offenders. They also found that offenders, who themselves had been abused, held the *dangerous world* theory to a significantly higher degree than those who had not. This finding makes intuitive sense: victims of abuse may well have an entrenched view of the world as a hostile and dangerous place as a result of that victimisation. Offenders against males also showed a higher level of endorsement of the *children as sexual beings* implicit theory which would be consistent with the fact that offenders with male victims are more likely to show paedophilic sexual interest as measured by penile plethysmography than offenders with female victims (Seto & Lalumière, 2001). As mentioned, the sample in the Marziano et al. study was small but did show support for Ward's five implicit theories. It was also promising to see that the strength of the implicit theories held seemed to suggest that different distortions may be indicative of different offending pathways and offending behaviour. It is important to note that it is not entirely surprising that this study lent support to Ward's theory since Ward devised his implicit theories to fit the cognitive products that were typically seen among offenders. Therefore, unless Marziano found atypical cognitive products among their sample, they were always likely to confirm the five theories. Thus, it could be argued that the measurement of behaviour and of cognitive products is not sufficient to explore the cognitive structure underpinning sexual offending.

Information processing methods or methods which are influenced by implicit cognition may well provide an avenue for exploring cognitive structures that avoids some of the problems associated with looking at cognitive products only (Langton, 2007; Mann & Beech, 2003; Segal & Stermac, 1990; Ward et al., 1997). Mann and Beech (2003) state: "It is our recommendation that future research move from measurement of attitudes to

16

identification and measurement of schemas. It is particularly important that a variety of assessment methodologies are employed for identification of schemas. Theories of schema such as social cognition theory and cognitive therapy theory indicate that schemas are not usually consciously accessible and therefore may not be uncovered by self-report inventories. Instead, methods such as think-aloud tests, Stroop tests (Stroop, 1938), repertory grids (Kelly, 1955), or even projective tests may be more suitable ways of identifying schemas in sexual offenders (Mann & Beech, 2003, p. 150)". To this end several studies have used implicit measures that have potentially tapped into the some of the proposed implicit theories (Banse, Schmidt, & Clarbour, 2009; Brown, Gray, & Snowden, 2009; Gray, Brown, MacCulloch, Smith, & Snowden, 2005; Mihailides, Devilly, & Ward, 2004; Nunes, Firestone, & Baldwin, 2007). Of these studies, only Mihailides et al. (2004) set out explicitly to test Ward et al.'s implicit theories. They tested three of the implicit theories and found support for all three. Some methodological concerns remain with the Mihailides et al. (2004) study and these will be considered later during a general discussion on whether such studies are measuring cognitive distortions or sexual interest.

Four studies by Gannon and colleagues also explored the presence of cognitive distortions among sexual offenders against children (Gannon, 2006; Gannon, Keown, & Polaschek, 2007; Gannon & Polaschek, 2005; Gannon, Wright, Beech, & Williams, 2006). The results are somewhat mixed and indicate that the distribution of cognitive distortions among offenders may not be as clear-cut as suggested by the study by Marziano et al. (2006). In one of the four studies Gannon and colleagues (Gannon et al., 2006) found their sample of (intrafamilial) offenders were no more likely than incarcerated controls to use distortions indicative of Ward and Keenan's (1999) implicit theories when recalling vignettes of offence scenarios specifically designed to tap into those theories. Another one of the studies found that treated and untreated child molesters along with nonsexual offending and non-offending controls all tended to perform similarly on a computerised cognitive distortion questionnaire (Gannon & Polaschek, 2005). When response times for items were measured, only the treated offenders showed faster reaction times potentially indicating that they were faking good, though a number of alternative explanations may explain the faster responding. A lack of support for widespread cognitive distortions was also found when offenders were administered a cognitive distortion questionnaire on two occasions, the second

time while connected to a pseudo lie detector (Gannon, 2006). The second administration did not yield more distorted belief disclosure relative to the first administration or relative to controls. However, those offenders who participated in this study were intrafamilial offenders. When the study was replicated using extrafamilial offending participants it was found that the second administration did increase cognitive distortion endorsements (Gannon, Keown & Polaschek, 2007). Taken together these studies seem to suggest that cognitive distortions can be somewhat elusive to measure and that methodology and sample makeup play an important role in their detection.

In addition to the five implicit theories posited to be held by offenders against children, Polaschek and Ward (2002) also proposed five implicit theories for rapists. Three of these overlap considerably with those of child molesters (i.e. m*ale sex drive is uncontrollable*, e*ntitlement* and d*angerous world*). In addition rapists may also hold a w*omen are unknowable* theory, which proposes that women and men are fundamentally different and that men cannot understand them. This may lead a rapist to consider sexual encounters as adversarial and that as part of that, women will be deceitful in order to get what they want from men. The fifth implicit theory states that *women are sex objects* and are therefore constantly receptive to male sexual needs, whether they are consciously aware of that or not. Similarly to the *nature of harm* theory for child molesters, rapists who hold the *women are sex objects* theory may not consider a woman to have been harmed by an assault unless she has been physically injured. Polaschek and Gannon (2004) assessed the five rape implicit theories proposed by Polaschek and Ward (2002) much in the same way as Marziano et al. (2006) did. Similarly to that study's results, support was found for the existence of all five implicit theories and there was no evidence to suggest that further implicit theories were required to explain additional cognitive products. Unlike Marziano et al. they didn't find evidence that their entire sample held all or even most of the implicit theories. Instead they found that 65% of rapists held the *women are unknowable* theory (which they changed to *women are dangerous* in line with their findings that statements in support of that theory were overwhelmingly negative), 70% held the *women are sex objects* theory while 68% held the *entitlement* theory. Both *male sex drive is uncontrollable* and *dangerous world* were less common at 16% and 19% respectively.

The difference in the prevalence of implicit theories among rapists and child molesters may represent differences in the true frequency of these cognitive structures among the different type of offenders or may be due to differences in methodology: Marziano et al (2006) collected their data using a specially designed interview schedule, whereas Polaschek and Gannon (2004) used extant interview data and used offence-process descriptions from those interviews to collect their data. However, like Marziano et al. (2006) group differences (between deniers and admitters) in the presence of implicit theories were in line with hypotheses suggesting that the presence or absence of distortions may be indicative of offence pathways or behaviours. It should be noted that Polaschek and Gannon (2004) stipulated that their research did not answer the question of whether the implicit theories supported post-hoc explanations, were ameliorations of 'self-referent negative affect' or whether they actually played a causal role in offending. It follows then that the same can be said of Marziano et al's work also. Polaschek and Gannon (2004) go on to stress that structures such as implicit theories need indirect measures to explore them since these structures, according to information-processing theory, are not accessible through conscious scrutiny. Qualitative investigations that indicate the presence of implicit theories need to be supported by the use of other methodologies, including experimental paradigms (Polaschek & Gannon, 2004).

Mann and Beech (2003) outlined a second schema-based model for the role of cognition in sexual offending. The model is not in opposition to Ward and Keenan's (1999) theory; in fact they include the five child sex offender implicit theories as likely schemas in the model. The model is more concerned with reconciliation of the existing sex offender cognitive distortion literature with social cognition and cognitive therapy literature. Drawing on social cognition theory and cognitive therapy theory, and using the example of cognitions in depression, Mann and Beech (2003) outline a framework by which developmental experiences result in dysfunctional category and belief schemas which interact with ambiguous or negative life events to influence (and bias) information processing. It is this process that results in the surface cognitions or cognitive products that have been previously referred to by many authors as cognitive distortions. Mann and Beech (2003) emphasise that they do not see schemas as the 'driving force' in offending, but something that interacts with other factors to result in sexual assault. These factors include deviant sexual arousal, poor conflict management skills and emotional loneliness. The strength and influence of this theory lies in the fact

that the vague term of cognitive distortion is grounded by Mann and Beech (2003) in the broader cognition literature and is teased apart into cognitive structures and cognitive products. In this way the authors have moved away from simply measuring these products. In addition the authors have proposed a framework by which dysfunctional schemas may come about and interact with other factors to contribute to distorted cognitive products.

The schema-based model of sexual assault proposed by Mann and Beech (2003) is a step forward but suffers from some limitations. Ward, Polaschek and Beech (2006) highlight several difficulties (while accepting the preliminary nature of the model) which include the emphasis on negative life events and the fact that Mann and Beech claim that schemas are 'chronically accessible'. Activation of schema is dependant on affect and they are, therefore, not necessarily always accessible (Marshall, Marshall, Serran, & Fernandez, 2006). In addition Ward and colleagues point to the lack of an explanation of the process by which schemas interact with other factors, and they also question the suitability of using the depression model as a template given the qualitative differences between depression and sexual offending as phenomena.

1.3.2   The Judgement Model of Cognitive Distortions (JMCD)

In 2000 Ward says: "At this point it is unclear whether cognitive distortions are primarily generated by maladaptive underlying schema or are the product of dysfunctional cognitive processing, or both (Ward, 2000, p. 492)." Several years later Ward and colleagues (Ward, Gannon & Keown, 2006; Ward et al., 2007) have attempted to address this point and have expanded on their earlier work on implicit theories to outline the most comprehensive model of offender cognitive distortions to date. They call their model the Judgement Model of Cognitive Distortions (JMCD). The model accounts for many of the inconsistencies and contradictions found in the literature on cognitive distortions. For example, it allows for offenders who make offence supportive statements indicative of implicit theories but who nevertheless, when assessed using other measures, seem not to hold those theories (Gannon, 2006; Gannon & Polaschek, 2005; Gannon et al., 2006). It also seeks to clearly map out the relationship between distorted beliefs and post hoc rationalisations, neutralisations etc. Importantly the model is the first to make a compelling case for how non-sexual

20

implicit theories can impact on the committal of sexual offences. Additionally the model seems to fit neatly into both Ward and Brown's (2004) Good Lives approach to offender treatment and Ward and Beech's Integrated Theory of Sexual Offending (2006, 2008).

The JMCD looks at cognitive distortions in terms of three different ways in which offenders may evaluate, or make judgements about their world: *belief-based judgements*; *value-based judgements* and *action-based judgements.* Belief-based judgements are evaluations of the world that arise out of beliefs or schemas that are held by the individual. These beliefs are about the nature of the individual themselves, and of the world. An offender may, for example, believe that adults are dangerous and to be treated with caution and suspicion. Ward and colleagues (Ward, Gannon & Keown, 2006; Ward et al., 2007) argue that careless reasoning over a sustained period can result in false or distorted beliefs while such reasoning over a shorter period can result in faulty conclusions. Faulty conclusions may explain why temporary distorted thinking may have contributed to an offence without the offender having a more ingrained false belief. Value-based judgements refer to aspects of people or of the world that the individual evaluates as positive or negative. These values are directly related to primary human needs or what Ward often calls human *goods.* Ward believes that offenders do not necessarily value the wrong things, but instead that they may try and meet those primary needs or goods, such as intimacy etc. in an inappropriate way. The reason why these needs are met in an inappropriate way may involve faulty belief-based judgements. Action-based judgements are evaluations that are made as a result of actions that have been committed. This type of judgement may encompass much of the post-hoc rationalisations etc. that has been talked about in the cognitive distortion literature.

Ward and colleagues articulate a framework by which sloppy reasoning can lead to faulty temporary conclusions or more entrenched false beliefs that can interact with values to result in actions. Actions can also result in a re-evaluation of values and beliefs in light of those actions. It is a significant step forward in the theory of cognitive distortions that this framework can account for qualitative differences between post-hoc rationalisations, neutralisations, minimisations etc. and distorted beliefs that have a causal role in offending. It also addresses how a process of sloppy reasoning can

operate at all levels of the framework thus explaining much of the apparent contradictions in previous cognitive distortion literature. For example: an entrenched belief that is a result of sloppy reasoning may play a causal role (belief-based judgement) in offending but sloppy reasoning may also lead someone who has offended to believe that they were in some way justified in that offence (action-based judgement) and thus play a maintaining role in future offending. This framework, though in many ways quite obvious has been lacking from past discussions of cognitive distortions.

Ward and colleagues (Ward, Gannon & Keown, 2006; Ward et al., 2007) look at each of the implicit child molester theories (Ward & Keenan, 1999) and the five implicit rapist theories (Polaschek & Gannon, 2004; Polaschek & Ward, 2002) in terms of the three judgement types (belief-based, value-based and action-based) in what they refer to as thematic networks. For the children as sexual beings implicit theory they outline the following thematic network.

"False Beliefs: Beliefs that may be associated with this thematic network concerns children's sexual interests and ability to make informed decisions about sex. Children are seen to actively seek sex, and the expression of their sexual desires is seen as legitimate.

Values: Value judgements with this thematic network hold that sex is always beneficial and takes precedence over other values or human primary goods. Another related value is autonomy; in this case the value judgement is made that children should be given agency to make their own decisions about sex.

Actions: Child sexual abusers' statements relating to this thematic network are likely to reflect the judgement that children are free and informed agents who can decide for themselves whether or not they want sex. They judge (or ask others to judge) that their actions were justified because their child victims wanted sexual contact" (Ward et al., 2007, p. 63).

The implicit theory of children as sexual beings is the one that is likely to be most closely linked to deviant sexual interest and is subsequently integral to this thesis. The JMCD model has implication for how cognitive products or offence-related utterances are interpreted. Consider the following hypothetical statement "children know about sex, she wasn't naïve; she made a big deal about it because the mother found out". This statement would be typical of one that would fit well into Ward and Keenan's (1999)

"children as sexual beings" implicit theory. However, to interpret whether this is a belief-based judgement, a more temporary faulty conclusion or an action-based judgement becomes more difficult. On one hand the offender could be interpreted as having an entrenched belief about how children think about sex; an offender who states this may consider children to be 'little adults' in terms of their sexual knowledge and apatite. However this could also be more of a temporary conclusion arising out of sloppy reasoning regarding the nature of the pre-abuse relationship with one child (e.g. a misinterpretation of playfulness as having a sexual intention). Third it could be an action-based judgement where the speaker has used sloppy reasoning to misinterpret the child's silence over the abuse as complicity. This example illustrates the necessity for sophisticated experimental techniques to identify the cognitive origin of such a statement. Ward and colleagues stress the need, albeit in an earlier work, for "experimental investigations using less direct and less conscious measures of cognition. Indeed this point is stated in *every* review of cognitive distortions... [yet] the general sophistication of research in this area remains quite low, and hampers theory development (Ward, Polaschek & Beech, 2006, p. 133)."

1.4   The role of sexual interest

Not all men and women who molest children have a sexual preference for children (Seto, 2008). That said, a certain proportion of men who sexually offend against children do have a sexual interest in children. Of those, some have an exclusive preference to children, while some are sexually interested in both adults and children. Most authors concur that gaining some insight into the sexual interest of offenders is an important component in the addressing the problem of child abuse, whether this is in relation to carrying out research, evaluating risk, devising treatments or assessing treatment efficacy (for an alternative view see Marshall & Fernandez, 2003). What is not agreed upon is the importance of sexual interest and its measurement relative to other factors involved in abuse. Neither is the role of the direct treatment of deviant sexual interest in therapy (e.g. Marshall & Fernandez, 2003) nor even the best method to assess sexual interest (for reviews see: Flak, Beech, & Fisher, 2007; Kalmus & Beech, 2005). The importance of sexual interest is more clearly defined in the context of predicting recidivism with deviant sexual interest in children (measured by penile plethysmography) being the single best predictor of recidivism (Hanson & Bussière,

1998). However, given that penile plethysmography (PPG) as a measure is potentially problematic (as will be discussed later) it may be that the true relationship between sexual interest and recidivism is stronger than that found by PPG studies. Indeed, a recent article by Marshall and Fernandez (2003), in which the authors conclude that measurement of deviant sexual interest is only of limited value fails to consider any methods other than PPG when drawing this conclusion. Currently sexual interest of offenders is assessed in several ways, including PPG, clinical interview, self-report, psychometrics, card sorting tasks and viewing time tasks. Each approach has their advantages and drawbacks.

1.4.1   Deviant sexual interest

As mentioned, deviant sexual preference is a clinical symptom or state factor in Ward and Beech's (2008) Integrated Theory of Sexual Offending and is a feature of most other theories of sexual offending. In the Integrated Theory of Sexual Offending deviant sexual interest is hypothesised to be a product of an interaction of the three neuropsychological systems implicated by Ward and Beech in offending. They postulate that it involves an inability to effectively manage mood problems and attachment issues, which relate to the motivation/emotional system. This, combined with dysfunctional schema, relating to the perceptual and memory system may lead to deviant fantasies and a pre-occupation with sex. A failure to regulate sexual drive, implicating the motivational/emotional system again may lead the person to use sex for their emotional needs. These factors combined with problems with sexual control, from the action selection and control system may lead to deviant sexual arousal. Level II theories that seek to explain the underlying causes of neuropsychological dysfunction involved in deviant sexual interest point to an impact of social learning, conditioning and neurobiological factors (Ward, Polaschek & Beech, 2006).

1.4.2   Information processing and sexual interest

The current literature on deviant sexual interest among offenders is lacking a theoretical explanation of the cognitive architecture of such interests. In terms of non-deviant sexual interest, Spiering and Everaerd (2007) explain that sexual memory can be located in explicit long-term memory through recollections of sexual encounters,

fantasies, attitudes about sex, and understanding of sexual costs and rewards and in implicit long-term memory through "innate sexual reflexes, learned (automatized) sexual scripts and classically conditioned sensations" (p. 767). These separate memory constructs interact to result in an individual's cognitive, physiological and behavioural responses to potentially sexually salient stimuli. It is not clearly hypothesised, to the author's knowledge, how and whether deviant sexual interest maps onto such a structure. However, each of the components in Spiering and Everaerd's framework has parallels in theories of sexual offending. Recollection of sexual encounters can quite clearly be problematic if those encounters involve children, either the offender's recollection of being sexually victimised as a child or of committing offences. Ward and Beech's (2006) Integrated Theory of Sexual Offending, among many other theories, implicates the abuse behaviour itself in a kind of feedback loop that may maintain or escalate the offending. The role of fantasy again is clearly implicated in many theories, e.g. Finkelhor's Precondition Model (Finkelhor, 1984), McGuire et al.'s Sexual Deviation Theory (McGuire, Carlisle & Young, 1965; as cited in Ward, Polaschek & Beech, 2006) and also Laws and Marshall's Conditioning Theory (Laws & Marshall, 1990). These last two theories also suggest a role of classical conditioning in deviant interest (Laws & Marshall, 1990). An individual with faulty or distorted cognition could develop deviant attitudes about sex, and a distorted understanding of sexual costs and rewards. Again, learned sexual scripts may be deviant if they are as a result of distorted cognitions involving problematic implicit theories or schemas. Finally, the Integrated Theory of Sexual Offending (Ward & Beech, 2006) supports the idea that even innate sexual reflexes, which are a product of evolution and brain development, could include deviant reflexes. The above is not suggested as a framework by which deviant sexual interest is organised in an offender's memory. Instead it is intended to demonstrate that in many ways deviant interest, if an offender has it, *may* be organised similarly to normal sexual interest. In addition, the above is intended as a starting point to explore how cognitive distortions and sexual interest may interact with one another.

The erotic response of humans is divided by Rempel and Serafini (1995; as cited in Flak et al., 2007) into the physiological (sexual arousal) and the psychological (sexual desire). Singer (1984) splits erotic arousal into a 'trichotomy' of responses. The first, *the aesthetic response*, involves keeping a sexually admired item in range (view, earshot etc.) by eye movements or head turning. This may develop into the second, *approach,*

response. This response involves more pronounced movements towards the target. The final response is *the genital response* involving genital tumescence, heart rate increases etc. Singer posits that these three responses are distinct since one does not necessarily lead to the next and since one is not necessarily a requisite for the 'next' response. It is the genital response that has been the focus for research on deviant sexual interest in the past, through the use of penile plethysmography. However, the approach response has provided other potential research avenues using attentional and information processing paradigms.
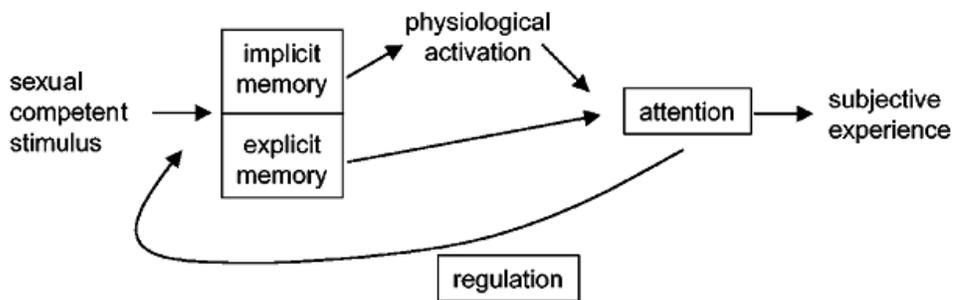
Figure 1.2 An information processing model of the processing of sexual information. Reproduced with permission from Spiering et al. (2004)

Janssen, Everaerd, Spiering and Janssen (2000) proposed an information processing model of sexual arousal that in some ways corresponds to Singer's trichotomy. Their model accounts for the attentional and physiological components of the trichotomy. Information processing models focus on how information is perceived, encoded and filtered; how a response is selected and then how it is executed. Spiering, Everaerd and Laan (2004) went on to further refine the existing model (see Figure 1.2). According to this information-processing model, sexual (or potentially sexual) stimuli are automatically appraised against representations in explicit and implicit memory. If there is a match with sexual elements in memory a physiological response is primed (Janssen et al., 2000). Attention is than brought to bear on the stimulus, which happens, according to Spiering et al. (2004), to allow regulation of the process. They also suggest that it is this triggering that results in the slowing down of responses to sexually salient stimulus in cognitive tasks in what has been termed sexual content induced delay (SCID; Geer & Bellard, 1996; Geer & Melton, 1997). Spiering et al. (2004) contend that the attention a stimulus is able to capture is a product of its ability to elicit arousal,

26

which, in turn is dependent on matches with sexual memory. They argue that once attentional mechanisms have been triggered (which engages regulation modules), conscious appraisal of the stimulus is occurring. It is from this stage that the subjective experience of sexual arousal emerges (Spiering et al., 2004). Spiering and Everaerd (2007) elaborate to an extent on how this unconscious process develops into a subjective conscious experience. They suggest this can occur by two pathways, through 'hot' cognition where awareness of somatic responses yields a conscious sexual feeling, or through 'cold' cognition where increased attentional focus results in the conscious experience. They also suggest that 'cold' cognition is linked to explicit memory and 'hot' to implicit memory. Spiering and Everaerd (2007) propose a network of brain structures that comprises what they propose may be a sex module in the brain. Evidence for this module, for 'hot' and 'cold' cognition and indeed for the information processing model of sexual responses is sparse. Potentially problematic is the linking of attentional processes with explicit memory and somatic processes with implicit memory. It is likely that implicit memory is brought to bear on the attentional process as that process may well contain automatic components that draw on implicit memory. It is worth emphasising however that Spiering and Everaerd (2007) do not state that 'cold' cognition is exclusively related to implicit memory or 'hot' cognition with explicit memory.

While the information processing model of sexual arousal proposed by Spiering and colleagues (Janssen et al., 2000; Spiering & Everaerd, 2007; Spiering et al., 2004) is not complete and has not been sufficiently supported as yet by empirical research, it does give a useful framework for the interpretation of attention-based paradigms for the measurement of sexual interest. The model explains sexual content-induced delay as a delay in the processing of a task since attentional resources have been limited as a result of the demand on those resources by the sexual stimulus in question. This additional demand may be as a result of the regulation process where the stimulus is being matched with representations that exist in explicit and implicit memory.

1.4.3   Penile Plethysmography (PPG)

Penile Plethysmography (PPG) is a measure designed to assess sexual arousal towards a given stimuli by measuring changes in the volume or circumference of the penis. It was

developed first by Kurt Freund and his colleagues in the 1950s in Czechoslovakia (Marshall & Fernandez, 2000). Initially the device was invented to discriminate homosexual men from heterosexual men (Kalmus & Beech, 2005). Differences have been identified between the volumetric and circumferential devices with volumetric seeming to demonstrate more accurate results, especially in the early stage of the erectile response where there can be changes to the length of the penis before there are changes to the circumference (Kalmus & Beech). However, because of cost, availability and ease of use, the circumferential device is the most widely used and reported (Kalmus & Beech). Many authors treat both measures as essentially equivalent (e.g. Flak et al., 2007; Kalmus & Beech, 2005; Marshall & Fernandez, 2000; Marshall & Fernandez, 2003). PPG is widely used in the United States and Canada having become an "integral part of the assessment of sexual offenders in many, if not most, treatment programs" (Marshall & Fernandez, 2003, p. 131). The technique is used to a limited extent in Britain within some prisons and psychiatric facilities (Craig et al., 2008). PPG has not been widely used in Ireland with sexual offenders, if at all. Concern over the use of PPG include the invasiveness of the technique (Beech, Fisher, & Thornton, 2003; Craig et al., 2008), and it could be argued that the procedure potentially dehumanises clients and may damage the client-therapist relationship. Additionally, questions remain regarding the reliability, validity and fakeability of PPG results. The reliability and validity of the evidence provided by PPG is no longer considered to be adequate for use within the legal system in the United States ("United States v. Powers", 1995). Results, however, are still used in determining treatment efficacy and for informing parole decisions. Much of the controversy regarding the reliability and validity of PPG stems from the fact that there have been many conflicting findings in studies using the technique. Murphy and Barbaree (1994) highlight that an inconsistency of method may well be responsible for much of the discrepancies across results. Within the field of sexual offending PPG was used initially as a research device for those studying paedophilia and sexual offending before it found its way into clinical use. However, unlike many clinical tools, the PPG lacks a standardised operating procedure (Murphy & Barbaree, 1994). Many of the different research laboratories used (and still use) different stimuli and different stimulus modalities. Indeed, laboratories were unable to share stimuli in the United States, in many cases, as to transport material that involved pornographic material containing children across state lines would have been a criminal offence.

28

Methods for scoring results have varied and include simply measuring change of circumference of the penis in millimetres, reporting change in circumference as a percentage of full erection and calculating an ipsative z-score, where changes are expressed in terms of a distribution of all responses (Marshall & Fernandez, 2000). Some authors remove participants who show low erectile response (Kalmus & Beech, 2005). However, Harris, Rice, Quinsey, Chaplin and Earls (1992) report that their inclusion did not lessen discriminant validity. On the other hand, Barbaree and Mewhort (1994) demonstrate that using z-score transformations of raw-score data results can be compromised by the inclusion of low variance subjects. Any data that are converted into ipsative z-scores attempts to standardise task-related response variance across participants, by expressing erectile response of individuals to different stimulus categories in standard deviations from their overall mean erectile response. This approach runs the risk of amplifying error variance among participants who have very little true response variance. In addition to using different methods in obtaining a measure of responses for different stimulus categories, these different methods of response treatment have been used to calculate 'deviancy' indices by comparing arousal towards a target stimulus type (e.g. female children, non-consensual sex) with a non-deviant stimulus type (e.g. non-coercive sex with an adult). These indices have been calculated using both ratios and difference scores.

While typically similar, step by step procedures also differ across research laboratories. One of the most critical differences that exists across studies is the selection of study participants. Theory and findings both indicate that different offender 'types', such as incest offenders, offenders against male and female children etc., will be characterised by different sexual interest profiles. Many studies do not separate out these different types and some are not specific about the make-up of their samples. Additionally, as previously mentioned, many studies lose quite a high proportion of participants due to low arousal to any of the material presented which has implications for the generaliseability of the findings. Marshall and Fernandez (2000) raise questions about the internal reliability of PPG testing as well as test-retest reliability.

Other findings may inform predictions of sexual interest among different 'categories' of offender but many should be interpreted cautiously as some of the studies have not been

repeated and therefore the effects found have not been replicated. Some of the more compelling findings are outlined below. Marshall Barbaree and Butt (1988) found that among men who have offended against male children, offenders who are homosexual in their adult sexual preference show arousal to post-pubescent males whereas offenders who have heterosexual adult preferences show arousal to pre-pubescent males. They suggest, in part based on the comments of the offenders who participated, that the heterosexually orientated offenders were looking for female-like features, whereas the homosexual offenders were looking for distinctly male features. The majority of studies that divided incestuous offenders into biological paternal and non-biological paternal (step/adopted fathers) offenders found very little difference between the two groups (Blanchard et al., 2006; Langevin & Watson, 1991; M. E. Rice & Harris, 2002; Seto, Lalumiere, & Kuban, 1999) which suggests that both categories of offender may be combined in terms of their expected sexual interest. However, one study did find differences between these groups in terms of deviant interests (Greenberg, Firestone, Nunes, Bradford, & Curry, 2005). In this study both groups were found to have deviant arousal as measured by PPG. However, stepfathers were found to demonstrate significantly higher deviant arousal than biological fathers. Seto et al. (1999) similarly found lower deviant arousal among biological fathers compared with step fathers, though the difference between the groups was not statistically significant.

Seto, Cantor and Blanchard (2006) found that child pornography[3] offenders (with or without victims) scored highest on deviant arousal. However, comparison groups were divided by number of victims but not by whether they were incest or non-incest offenders (M. Seto, personal communication, September 5, 2008). A sampling method that separated offenders into incest and non incest or mixed offenders would have likely yielded a level of deviance among extrafamilial offenders that was closer to that shown for pornography offenders. Additionally, the use of PPG may be problematic when assessing pornography offenders. The entire method of stimulus presentation in PPG is more similar to pornography offending than it is to contact offending. In a sense what is being asked by PPG is: "do you get aroused by watching images of children while listening to a narrative of a sexual encounter?" Pornography offenders are very likely to

---

[3] The term 'child pornography is' used throughout this thesis instead of 'images and videos of child abuse' for brevity but does not seek to ignore the qualitative difference between pornography that involves consenting adults, and abusive and coercive pornographic material involving children.

respond to this since their offence consists of a similar behaviour. While the results of the task are no less valid as a result of this, it may be that PPG necessarily produces a larger effect for pornography offenders, rendering comparisons with other offenders questionable. By way of further explanation - it makes sense to see whether a non-pornography offender is aroused by what is essentially pornographic material. This may indicate the degree to which that person has a sexual attraction to children. This can then inform treatment targets and risk assessment. It informs risk assessment since the assessor now has two pieces of information, namely, the offender has molested a child in the past and is sexually interested in children. For a contact offender with no history of using child pornography, PPG does not recreate an offending situation. On the other hand using PPG with pornography offenders does approximate an offending situation. The assessor does not have as clear a picture of the pornography offender: this man has viewed child pornography and is aroused by pornographic material involving children. The strength and specificity of that arousal response may indicate risk of porn recidivism, it may even predict future contact offences but the relationship between the PPG measure and future contact offences would seem to be less clear than in the case of the non-pornography offender.

It is apparent that the use of PPG remains controversial and problematic. However, certain results using PPG with child molesters seem robust enough to inform predictions of sexual interest among different sub-groups of offenders. Extrafamilial offenders or offenders with both incestuous and non-incestuous victims typically show the greatest arousal to child stimuli (Abel, Becker, Murphy, & Flanagan, 1981; Frenzel & Lang, 1989; Freund, Watson, & Dickey, 1991; Lang, Black, Frenzel, & Checkley, 1988; Marshall, Barbaree, & Christophe, 1986; Quinsey, Chaplin, & Carrigan, 1979; M. E. Rice & Harris, 2002; Seto et al., 1999) and many studies show that incestuous offenders typically have similar arousal patterns to 'normal' non-offenders (e.g. Frenzel & Lang, 1989; Freund et al., 1991; Marshall et al., 1986; Murphy, Haynes, Stalgaitis, & Flanagan, 1986; Quinsey et al., 1979). However, several studies have demonstrated arousal to child stimuli among familial offenders when audiotaped stimuli were used without visual stimuli (Abel et al., 1981; Lang et al., 1988; Murphy et al., 1986). This may indicate the presence of deviant sexual interest among certain incest offenders that is identifiable when the use of audio stimuli without images allows the offender to imagine his victim(s). However, in the case of Abel et al. (1981) the incest offenders

numbered only 6, one of whom, admitted, albeit after being presented with his PPG results, to have had a long term sexual interest in young girls. While this individual may be representative of some incest offenders it is likely that, as a group, incest offenders have a heterogeneity that cannot be fully represented by a sample of 6. Given the small sample size the influence of this one participant's responses on mean scores may have been quite substantial. However, Murphy et al. (1986) did replicate the findings of Abel et al. (1981) with a larger sample. The finding that stimulus modality may influence the degree of deviance demonstrated by different categories of offender may have implications for any task intended to assess sexual interest among incest offenders. The stimulus modality should be considered when designing such a task, as pictorial stimuli may in fact mask a sexual interest that is towards specific children.

In Ireland exhibitionists who expose themselves to children are typically treated alongside contact offenders against children. PPG has been used to a limited extent with such offenders. Marshall, Payne, Barbaree and Eccles (1991) found no correlation between recidivism and arousal to exposing scenarios. Typically exhibitionists with offences of exposing to adults show similar patterns of arousal to non-offending controls for exposing and intercourse scenes (Marshall & Fernandez, 2003). Rabinowitz Greenberg, Firestone, Bradford, & Greenberg (2002) found that exhibitionists typically showed deviant scores when a paedophilic index (PI) was calculated based on their PPG results. A follow up study reported the same finding that a significant number of the exhibitionists in their study demonstrated paedophilic sexual interests (Firestone, Kingston, Wexler, & Bradford, 2006). However neither Rabinowitz Greenberg et al. (2002) nor the follow up study (Firestone et al., 2006) differentiated in their sample between those who had exposed themselves to children and those who had exposed themselves to adults (P. Firestone, personal communication, September 8, 2008). This may well have implications for mean PI scores of the sample since it is likely that those offenders who exposed themselves to children would have accounted for a large proportion of elevated scores. As with other findings from PPG research, the conclusions that can be drawn from studies that have included exhibitionists are tempered by a heterogeneity of participants that is potentially problematic. However it seems plausible to assume that offenders who expose themselves to young children do so for reasons including paedophilic sexual interest.

To conclude, many of the studies using PPG suffer from a lack of comparability of samples as well as a lack of uniformity of experimental method and analysis of results. What can be concluded with relative confidence is that non-offenders and rapists typically show less arousal to paedophilic stimuli than paternal offenders who in turn show less than non-paternal incestuous offenders and extrafamilial offenders. The most deviant arousal is typically shown by men who have offended both intrafamilially and extrafamilially. Offenders who used child pornography (with or without contact offences) also demonstrate high levels of arousal.

1.4.4 Viewing Time Measures of Sexual interest

Early research established a relationship between the length of time an individual spends viewing sexual images and sexual interest (Rosenzweig, 1942; as cited in Kalmus & Beech, 2005) and that the length of time spent viewing images of naked men and naked women could differentiate between heterosexual and homosexual participants (Zamansky, 1956; as cited in Kalmus & Beech, 2005). In a typical viewing time paradigm, participants are asked to examine slides and to rate how sexually attractive they find the content of the slides. In many ways this is similar to reaction time tasks that purport to measure sexual interest. However referring to the procedure, as several studies have done, as a "visual reaction time measure" (e.g. Abel, Huffman, Warberg, & Holland, 1998; Abel et al., 2004; Barboza-Whitehead, 2001; Letourneau, 2002; Williams, 2003) is somewhat of a misnomer (Maletzky, 2003) since it is different to typical reaction time methods, which are usually rapid. The participant in a viewing time task is not attempting to carry out the task as rapidly as possible and there is no correct answer, therefore they are 'choosing' more so than 'reacting'. This distinction has implications for the clinical utility of this type of task and also for a theoretical understanding of what it is measuring. When a participant conducts a reaction time task they are usually told to "respond as quickly and as accurately as possible". Participants will try and strike a balance between accuracy and speed. Which of these they prioritise will depend largely on their understanding of the task. Some participants will realise that speed of response is being measured and some will not. Typically, however, a reaction time task will be relatively unaffected by whether the participant has prioritised speed or accuracy (i.e. the effect will still show through) as long as the participant has not completely abandoned one for the other. A viewing time task, on the other hand, is

far more vulnerable to distortion if a participant is aware that speed of response is being measured. Such a distortion might happen deliberately or unintentionally. Theoretically, if one adopts Spiering et al.'s (2004) information processing model of sexual arousal it is likely that viewing time measures are tapping into the stage of processing where the attention devoted to the task is sufficient to produce a subjective conscious experience.

Despite some concerns raised above there are two viewing time measures currently commercially available, the *Abel Assessment for Sexual Interest* (AASI; Abel et al., 1998) and the *Affinity* (Glasgow, Osborne, & Croxen, 2003) designed for use with offenders with learning disabilities. In the AASI the results of the viewing time task are combined with responses on a sexual history and interest questionnaire and the individual's response pattern is compared to normed results from a sample of admitting and non-admitting offenders (Abel, Jordan, Hand, Holland, & Phipps, 2001). The exact scoring algorithm for the AASI has not been released so independent evaluation by other researchers is difficult (Laws & Gress, 2004). However, Letourneau (2002) found that viewing time using the AASI materials and technique did accurately identify offenders against male children and female adolescents but not offenders against female adolescents and adult females. PPG results in this study did not prove more satisfactory with a similar pattern of findings emerging. By calculating a deviance index for viewing time similar to that calculated in PPG studies Harris, Rice, Quinsey and Chaplin (1996) were able to significantly discriminate between child molesters and non-offenders. However this discrimination was less than that obtained through PPG. Again though, viewing time yielded similar results to PPG. While the AASI uses non-nude images, Harris et al. (1996) used nude images. Gress (2005) also used nude slides (along with clothed slides) and demonstrated a high rate of consistency when classifying offenders and non-offenders. Viewing time tasks have demonstrated themselves to be commercially viable alternatives to plethysmography. They have clear utility in identifying the sexual interest of those being tested. However it is likely that the transparency of the task may prove to be its downfall. Already there are internet websites that explain in detail exactly what is being measured by the AASI (Sachsenmaier & Gress, in press).

1.4.5 Implicit Cognitive or Information Processing methods of measuring sexual interest

A growing number of cognitive tasks are being applied to the investigation of sexual interest, both deviant and non-deviant. Two of these implicit tasks, the Implicit Association Test and the modified Stroop task will be dealt with in more detail in section 1.6. However, other tasks have demonstrated a utility in discriminating participants based on sexual interest. Wright and Adams (1994) used a choice reaction time (CRT) paradigm, where participants had to identify the location of a dot situated on slides of nude males, nude females or neutral scenes, and found that 95% of participants (male and female) showed increased response latencies in line with their stated sexual preference. The same authors later replicated their findings and demonstrated that clothed images yielded a similar but weaker pattern of results (Wright & Adams, 1999). Giotakis (2005) adapted the CRT paradigm for use with rapists, child molesters and controls and was able to demonstrate the potential of the procedure through identifying group differences between offenders and controls. Gress (2008) found unclear results using a CRT task with adult sexual offenders, youth non-offenders and university students. A viewing time measure administered for comparison yielded clearer group differences that the CRT measure. Gress interprets this result in light of methodological concerns with the CRT.

Hecker, King and Scoular (in press) found using the eye startle probe reflex, that participants looking at aversive stimuli demonstrated higher intensity of these reflexes to aversive images than to neutral or sexual stimuli. They found that the reflexes were attenuated when attending to images of preferred sexual partners (stimuli were paired with short phrases designed to suggest sexual activity with the person depicted). They also found that participants were unable to suppress the response when asked to do so. Janssen, Vorst, Finn and Bancroft (2002) demonstrated that male undergraduate students showed reduced startle reflexes to video clips of consensual sex relative to responses to coercive scenes, perhaps indicating that coercive scenes were aversive or at least less positive to participants (however coercive scenes had no explicit sex, while the consensual scenes did, potentially confounding the results).

Rapid Serial Visual Presentation (RSVP) is a method of stimulus presentation where a sequence of information is presented quickly. When a target stimulus presented within a group of distracter stimuli is salient to the participant (e.g. a sexually salient image for an offender) the participant may fail to or take longer to respond to a second target stimulus if it is presented within 500ms of the first. Beech et al. (2008) found that, under conditions such as those described above (referred to as attentional blink in the literature) child molesters showed greater number of errors when the first target was a picture of a child compared with when the first target was a picture of an animal. This result was significantly different to that of the control group. Interestingly there was little difference in performance between extrafamilial and intrafamilial offenders.

The three paradigms summarised above are included as a demonstration of the variety of implicit task currently demonstrating potential in the investigation of sexual interest among sexual offenders. It is likely that a comprehensive information-processing model of sexual interest/arousal would allow researchers to identify the aspects of that model which different tasks may be tapping into.

1.4.6   The Screening Scale for Pedophilic Interests (SSPI)

Seto and Lalumière (2001) introduced a brief screening scale for identifying paedophilic sexual interest which they called the Screening Scale for Pedophilic Interests (SSPI). The SSPI was designed for the purposes of triage, risk management and research when phallometric data was unavailable. It comprises of four questions: whether the offender has had a male victim, whether they have had more than one victim, whether the victim was aged 11 or younger and whether the offender had any unrelated victims. These questions were chosen, since offender files are likely to contain this information and past research has indicated that a positive result for these factors is related to a greater paedophilic interest (Freund & Watson, 1991; Seto et al., 1999). The scale is measured from one to five with one point given for each positive answer except for the male victim question, with offenders with male victims being given a two. Seto and Lalumière demonstrate the scores on the scale correlate moderately with PPG (deviance index measure) and that offenders with low SSPI scores demonstrate an average PPG pattern similar to those of rapists and non-offenders while offenders with high SSPI scores on average demonstrate a more typical paedophilic'

PPG pattern. In addition using receiver operating characteristics (ROC) yielded an area under the curve (AUC) of between .7 and .73. An ROC curve is a "function that relates hit and false-alarm rates to each other" (Macmillan & Creelman, 1990, pp. 62-63). Chance alone should yield an AUC of .5 while a task that predicted paedophilic interest correctly all the time would have an AUC of 1. While these results are by no means excellent[4] (and the authors caution that the SSPI is simply a quick best guess and should not be used as a replacement for phallometry), they do indicate that the SSPI may have some utility as an indicator of deviant sexual interest. The simplicity of the SSPI causes problem. Consider the following hypothetical examples:

1) An offender who victimised his two teenage sons, aged 12 and 13 would be given an SSPI score of 3

2) An offender who had multiple, unrelated, prepubescent female victims would be given an SSPI score of 3

Given that both these offenders would be given the same SSPI score it is unsurprising that, depending on the PPG cut-off score used, between 28 and 41% of those scoring three were classified phallometrically as having paedophilic interest. This category necessarily has a lot of heterogeneity. Though actuarially it makes sense to have a weighting of two for the male victims question, pragmatically it is problematic since it means that it is impossible for an individual, who has female victims only, to be classified as either a 4 or a 5. It may be that this weighting distorts the picture somewhat. Despite this, the SSPI may prove useful as an easy to apply, if limited, estimate of paedophilic sexual interest. The measure's utility has been further demonstrated by Seto, Harris, Rice and Barbaree (2004) in a study that showed that SSPI scores significantly contributed to the prediction of recidivism.

---

[4]Tape (1999) describes a heuristic for interpreting AUCs with over .9 being excellent, .8-.9 good, .7-.8 fair, .6-.7 poor and AUCs of .5-.6 being described as fails. Rice and Harris (2005), on the other hand, equate AUCs to common measurements of effect size, including Cohen's *d*, equating AUCs of .556, .639 and .714 to effect sizes that Cohen (1988) would consider small, medium and large effects respectively. It is probably best to interpret AUCs not by these conventions but rather by the trade-off of specificity versus sensitivity that best suits the research in question. Whereas tests designed to diagnose illness may be considered too inaccurate at an AUC of .8, such a value would be considered very useful when performing exploratory investigations such as have been reported here.

1.5 What is the relationship between deviant sexual interest and cognitive distortion?

None of the extant theories of sexual interest satisfactorily deal with the question of the relationship between cognitive distortion and sexual interest. Most authors explicitly or implicitly acknowledge the link between the two but at the same time treat them as separate state factors in offending behaviour (e.g. Ward & Beech, 2008). There are two main reasons for this division. First, distorted cognitions in offenders are not limited to distortions with a sexual element. Despite this, one could contend that deviant sexual interest is a component of the *sexual* cognitive distortions. Second, assuming distortions and sexual interest are distinct allows for one to be present in a given offender without the other. Whether this is the case or not is yet to be established in the literature and will require sophisticated approaches to measuring both cognitive distortions and sexual interest.

Mann and Beech (2003) argue that cognitive distortions arise out of problematic schema and interact with other factors such as 'deviant sexual arousal' to result in offending behaviour. This indicates that they believe distortions and sexual interest to be distinct but related. On the other hand Ward and colleagues also highlight the role of schema in cognitive distortions (e.g. Ward & Beech, 2008; Ward, Gannon & Keown, 2006; Ward & Keenan, 1999) but suggest dysfunctional schemas or implicit theories may *lead* to deviant sexual fantasy (Ward & Beech, 2008). Given that a deviant sexual preference in most cases is not thought to be innate, it is necessary that memories are encoded in the offender which result in a distortion of their sexual interests. It is likely that, if originally present, cognitive distortions will impact on the way in which these memories are encoded. For example a non-sexual encounter with a child may be encoded as sexual if the individual holds the implicit theory that children have adult-like sexual interest/urges. This encounter may contribute to the individual developing a sexual interest in children. The reverse of this process is also likely to be true: if an individual has a sexual interest in children it is likely that their world view will attempt to accommodate that interest which may result in distorted schema. For example, if an offender finds himself sexually attracted to children he may begin to see children as being sexually conscious. A sexual interest in children could also lead to the generation of more implicit theories, similar to those proposed by Ward and Keenan (1999). For

example, if an individual has sexual interests that are at odds with society's values a process of cognitive dissonance could result in the individual forming beliefs that sex drive is uncontrollable and/or a sense of being entitled to sexual gratification.

Already mentioned is Spiering and Everaerd's (2007) hypothesis that non-deviant sexual memory is organised in implicit long-term memory to incorporate "innate sexual reflexes, learned (automatized) sexual scripts and classically conditioned sensations" (p. 767) and in explicit long term memory to incorporate recollections of sexual encounters, fantasies, attitudes about sex, and understanding of sexual costs and rewards. It is likely that, in an individual with deviant sexual interest, their sexual memory is organised in much the same way. It is also likely that certain abnormalities in this network would allow for problematic sexual behaviour but not necessarily be sufficient for the individual to hold a marked deviant sexual interest. Having distorted scripts and schemas relating to sex could occur in the absence of deviant fantasy or classically conditioned sensations etc. In this way it is likely that certain individuals could have distorted cognitions about children and sex, and about their own sexual entitlement and the controllability of their sexual urges without having a clear deviant sexual interest. It is less clear how someone could have deviant sexual interest with out having some degree of distorted cognition with it, either as a cause or as a result of that interest.

1.6 Utility of applied implicit cognition to assessment of sexual offending

Implicit cognitive measures are now increasingly being used in research involving cognitive distortions and sexual interest in sexual offenders (Banse et al., 2009; Brown et al., 2009; Gray et al., 2005; Gress & Laws, in press; Mihailides et al., 2004; Nunes et al., 2007; Price, 2005; Smith & Waterman, 2004). The next section will focus on two tasks in particular, the modified Stroop task and the Implicit Association Test (IAT), since these are the two tasks focused on in this thesis.

1.6.1   The modified Stroop task

The word version of the modified Stroop task is a common implicit task used to explore a wide variety of topics. A traditional (therefore non-modified) Stroop task asks participants to name the colour in which words are presented. Some of these words are colour words that are incongruous with the colour of ink in which they are presented. A delay in responding in these cases demonstrates an interference caused by the participant automatically reading the word which makes naming the incongruous font colour more difficult. A modified Stroop paradigm uses a similar method to demonstrate that the content of a word or image that is salient to the participant may impact on the participant's ability to quickly name a target colour. The terms emotional Stroop and modified Stroop are used interchangeably, even by the same authors (e.g. Cox, Fadardi, & Klinger, 2006). Strictly speaking while any adaptation of the Stroop design may be a modified Stroop task the label of emotional Stroop suggests that it is the emotional content of any potentially salient stimuli that is responsible for any attentional bias. To avoid this assumption this dissertation uses the more general term of modified Stroop task. The modified Stroop task, as with the traditional Stroop task, places a to-be-ignored stimulus in conflict with a to-be-attended-to stimulus. The to-be-attended-to stimulus is always a colour, be it the colour of the font in which a word is written (word Stroop) or the colour of either a border or filter of a picture (pictorial Stroop). Participants are asked to name the colour as quickly as they can. Any systematic difficulty in doing this is interpreted as demonstrating that the to-be-ignored stimuli hold some salience for the participant. This type of content induced delay can be used to investigate many research questions, from responses of cocaine addicts towards images related to cocaine use (Hester, Dixon, & Garavan, 2006) to the reaction of violent offenders to words relating to aggression (Smith & Waterman, 2003). An attractive feature of any Stroop type task is its simplicity. The participant needs only to name out loud or press the button that corresponds to the colour[5] of the image or word. In the case of the pictorial modified Stroop, the participant does not even need to be able to read as all instructions can be read to him/her if necessary.

---

[5] Clearly, if a participant were colour blind they may not be able to differentiate between the colours, which would void their results. Between approximately 3% and 8%, depending on ethnicity, of males have some form of colour blindness with the prevalence among women being much lower (Sekuler & Blake, 1994).

The word version of the modified Stroop task is far more commonly used than the newer pictorial adaptation. The word modified Stroop has been mostly used in looking for attentional biases related to the perception of potentially threatening stimuli, often among anxious, phobic or post traumatic stress disorder participants (e.g. Bradley, Mogg, Millar, & White, 1995; Egloff & Schmukle, 2004; Holle, Neely, & Heimberg, 1997; McNally, Riemann, Louro, & Lukach, 1992; Mogg et al., 2000; Wikström, Lundh, & Westerlund, 2003). However, the technique has also been used widely to investigate attention to addiction related stimuli including alcohol (e.g. Carrigan, Drobes, & Randall, 2004; Cox, Yeates, & Regan, 1999; Kramer & Goldman, 2003; Ryan, 2002; Sharma, Albery, & Cook, 2001), smoking (e.g. Mogg & Bradley, 2002; Munafò, Mogg, Roberts, Bradley, & Murphy, 2003; Powell, Tait, & Lessiter, 2002; Waters, Shiffman et al., 2003; Wertz & Sayette, 2001), gambling (e.g. Boyer & Dickerson, 2003; McCusker & Gettings, 1997) as well as heroin, cocaine and marijuana abuse (e.g. Carpenter, Schreiber, Church, & McDowell, 2006; M. Field, 2005; Hester et al., 2006). Other applications of the method have included investigating attention to stimuli related to food, eating disorders, body image and appearance (e.g. Braet & Crombez, 2003; Formea & Burns, 1996; Green, Elliman, & Rogers, 1996; Labarge, Cash, & Brown, 1998; Mahamedi & Heatherton, 1993; Perpiña, Leonard, Treasure, Bond, & Baños, 1998) as well as disease, fear of illness and malingering (e.g. Cannon, 2003; Erblich, Montgomery, Valdimarsdottir, Cloitre, & Bovbjerg, 2003; Fortune et al., 2003; Jessop, Rutter, Sharma, & Albery, 2004; Lecci & Cohen, 2002; Zeitlin, Bradburn, & Lawson-Kerr, 1995). Attention to disorder related material has been investigated in depressed and hypomanic participants (e.g. Dalgleish et al., 2003; French, Richards, & Scholfield, 1996; Gotlib & Cane, 1987) as well as for suicide related words in suicide attempters (Becker, Strohbach, & Rinck, 1999). Processing for aggression words has also been investigated (e.g. Smith & Waterman, 2003, 2005) as has attention to salient stimuli among perpetrators and victims of sexual violence (e.g. Dubner & Motta, 1999; Foa, Feske, Murdock, Kozak, & McCarthy, 1991; Freeman & Beck, 2000; Smith & Waterman, 2004). The methodology adopted across the studies cited above and other modified Stroop studies varies quite considerably but most consistently find evidence of attentional bias towards their target stimuli.

The ability of certain stimuli to divert or compete with the individual's attention allows the experimenter to interpret the salience of those stimuli for that individual. One must

be cautious in extrapolating meaning from what is essentially a momentary increase in attentional demands. For example, two modified Stroop studies have shown victims of sexual violence to show attentional bias towards words relating to the abuse/assault that they suffered (post-traumatic stress disorder was an additional factor in both, Dubner & Motta, 1999; Foa et al., 1991). These studies, taken with the studies that demonstrate an attentional bias towards offence-related words among perpetrators of sexual violence (Price, 2005; Smith, 2006; Smith & Waterman, 2004) demonstrate that the interpretation of any bias must be viewed in the context of how the stimuli have come to be salient.

It has been shown, that performance on a Stroop task *can* be predictive of behaviour; Waters et al. (2003) found that smokers trying to quit showed the strongest attentional bias towards smoking-related words and were more likely to lapse within the follow-up period even when years smoking and self-reported urges were controlled for. Some studies have shown promise in demonstrating that an attentional bias demonstrated by the Stroop task may be malleable through treatment (e.g. Lundh & Öst, 2001; Mattia, Heimberg, & Hope, 1993).

While such studies suggest that the modified Stroop task is capable of tapping into dynamic and predictive cognitive processes, the mechanism by which the task does this is still unclear. Smith (in press) explores this question in more detail. He describes the modified Stroop (with sexual words) as an 'information processing measure of sexual interest' since the response latency is a function of the individual's cognitive processing of the stimuli presented. He emphasises the multiple factors that are likely to influence response bias, including the nature of sexual representations and affect. Smith differentiates between a cognitive-representational model and an information-processing model of sexual interest. He argues that adopting a cognitive-representational model of sexual attention facilitates the conclusion that both cognitive bias (i.e. the Stroop task) and self-report are both consequences of the same representational structures, with bias being simply a more pure measure.

1.6.2 The pictorial modified Stroop task

Some research questions may be best addressed by eliciting a concept using words. Consider the offender whose sexual interest involves committing sadistic acts with specific victims. Images may not be useful with this individual; either because researchers are unprepared or unable to include images depicting sadistic acts, because such an image may become too multi-featured for practical use in a rapid-response task or because too many specific features may be necessary to produce a response from this individual (child age, gender, hair colour, ethnicity, type of sadism). In cases such as these it may be preferable to use words such as 'hit', 'beat', 'gag' etc. over images so that the stimulus does not become too specific to produce a response. On the other hand, the use of images as stimuli in the Stroop task has certain advantages over words. Using images, we can manipulate the age of the individual depicted, their gender, their pose, their ethnicity, the emotion they portray. When trying to assess sexual interest towards children generally, the use of pictures affords a greater scope and flexibility of presenting potentially sexually salient stimuli to the participant. In order to do the same with words one would have to come up with many words that suggest the concept of "sex with children" to the participant, which would be more difficult. As with any measure of reaction time to salient stimuli, the interpretation of such an attentional bias is potentially problematic. One could propose several reasons why a participant would have a slowed response to a given image. Therefore, it is the consistency of response across stimulus category that we may interpret, in this case, as indicative of sexual interest.

The pictorial version of the modified Stroop task has been used in many of the same fields of research where the word version has been used, albeit in a limited way. Studies include the assessment of attention to facial emotion (Ashwin, Wheelwright, & Baron-Cohen, 2006; Heim-Dreger, Kohlmann, Eschenbeck, & Burkhardt, 2006; van Honk, Tuiten, de Haan, van den Hout, & Stam, 2001; van Honk et al., 1998, 2000; van Honk et al., 1999) to cocaine-related images (Hester et al.) to phobia related material (Constantine, McNally, & Hornig, 2001; Côté & Bouchard, 2005, 2009; Elsesser, Heuschen, Pundt, & Sartory, 2006; Kindt & Brosschot, 1997; Kindt & Brosschot, 1999; Kolassa, Musial, Mohr, Trippe, & Miltner, 2005; Lavy & Van den Hout, 1993) as well as towards images relating to eating disorders (Stormark & Torkildsen, 2004; Walker,

Ben-Tovim, Paddick, & McNamara, 1995). The methodology used has varied across studies. Most adopt an unblocked design, where all stimuli are presented randomly as opposed to being presented in groups of matching stimuli (except Hester et al., 2006; Stormark & Torkildsen, 2004). Some studies use pictorial stimuli where the image and the colour are not integrated with the colour forming a border around the image (e.g. Elsesser et al.; Matt Field, Duka, Tyler, & Schoenmakers, 2009), instead of having a coloured version of the image (e.g. Constantine et al., 2001; Côté & Bouchard, 2005).

Two of the studies cited above found similar or reduced bias for pictorial Stroop compared with word Stroop (Kindt & Brosschot, 1997; Lavy & Van den Hout, 1993) whereas a further three studies were unable to find a pictorial Stroop effect at all (Matt Field et al., 2009; Kindt & Brosschot, 1999; Kolassa et al., 2005). In the case of Kindt and Broschott (1997), the researchers did find an attentional bias for pictorial and linguistic stimuli among their spider-phobic participants compared with controls but they found that linguistic stimuli performed better than pictorial stimuli. Interestingly, they also compared non-integrated linguistic stimuli (which were words placed on a coloured circle) which also performed worse than the integrated words. The picture stimuli in this study were created by placing images of spiders on coloured circles. The lack of integration of picture and colour may have been a factor in the relative weakness of the pictorial stimuli to produce results. In their 1999 study, Kindt and Broschott again used non-integrated images. Interestingly three of the five null or poor findings with the pictorial Stroop used the same images in their studies (Kindt & Brosschot, 1997; Kindt & Brosschot, 1999; Lavy & Van den Hout, 1993). Some possible explanations for their own null finding were put forward by Kolassa et al. (2005) and included the fact that they had used only two colours in their version of the Stroop task, that their design was not blocked, and that response modality was manual, whereas oral responses are considered optimal in producing effects (MacLeod, 1991). Since four of the studies were with spider-phobics, a phenomenon specific to attention for phobia related material may underlie these findings (possibly with spider images there, might be a conflict of aversion and attention involved). However, in a second administration of a pictorial Stroop task adopting an unblocked design where image and colour were integrated, Côté and Bouchard (2005, 2009) found a greater post-treatment decrease in response times to spider images than to control images among spider-phobics. The fifth study with null or poor findings, by Field et al. (2009), found no difference in response

times to smoking-related versus control images but again used a coloured border instead of integrating the colours and the images. Despite these particular studies, it is apparent that a pictorial Stroop design can, with certain methodological considerations, identify attentional bias towards salient stimuli.

1.6.3   What does the pictorial modified Stroop task measure?

It is quite likely a pictorial modified Stroop task using potentially sexually salient stimuli is a simple measure of sexual content induced delay, as indexed by reaction time delay, much as the choice reaction time (CRT) task is. If one tries to locate its temporal location on Spiering et al.'s (2004) information processing model of sexual arousal, Stroop interference would probably be occurring at two points. The first would be at a pre-attentive stimulus encoding stage and would be responsible for any so-called fast component in the Stroop. A fast component is an interference effect of stimulus content that is apparent on an individual trial, where the attention has been 'grabbed' (McKenna & Sharma, 2004). It is also possible that this is not a truly pre-attentive process but occurs at early point of the attentive stage where the experience is not yet subjectively available. A slow effect, on the other hand operates between trials (McKenna & Sharma, 2004). It is related to the higher order rumination of the stimuli content and possibly to the activation of associated concepts related to that stimulus category. This is likely to occur during the attentive stage described by Spiering et al. (2004). While involving a conscious and possibly subjectively available experience this process would not be as subjective an experience as that of a viewing time task. A pictorial modified Stroop task might not contain both fast and slow components. If fast components exist a randomised presentation of stimulus types should still yield an effect.

1.6.4   The Implicit Association Test

The Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) is an indirect method that attempts to measure the strength of associations between concepts (Nosek, Greenwald, & Banaji, 2007). It involves sorting individual stimuli into one of four categories with the use of only two buttons. Therefore two categories have to be paired on each button. The logic behind the task revolves around the assumption that it should be easier to carry out the task when the concepts that are paired are associated in

the participant's cognitive structure than when they are not associated. Thus the task will be easier for most people when the concepts of *flower* and *pleasant* are paired than when *insect* and *pleasant* are paired since for most people the former two concepts are clearly more closely associated than the latter two concepts. The IAT has been used recently to explore it's utility in the study of cognitive distortions and sexual interest among offending and non-offending participants (Banse et al., 2009; Brown et al., 2009; Gray et al., 2005; Mihailides et al., 2004; Nunes et al., 2007; Snowden, Wichter, & Gray, 2008). Gray and Snowden (in press) have highlighted the methodological inconsistencies across many of the IATs used to investigate sexual interest/cognitive distortions.

1.6.5   The IAT and sexual offenders

When developing stimuli to assess sexual interest in children using the IAT, there can be an accumulative difficulty in attempting to satisfy two of the principles of good IAT constructions: that stimuli should be easily classifiable into one of the four coherent categories and that categories are formed as contrasting pairs (Nosek et al., 2007). To illustrate this we will consider the following cognitions that might be held by sexual offenders against children:

       1) Children are sexual beings

       2) Children are associated with sex or sexual activity

       3) Children are sexually attractive

Researchers may be interested in investigating whether these cognitions are held by offenders and, if so, whether they are held by all offenders or just a subset of offenders and how common such cognitions are among the non-offending population. While these three cognitions are very similar, the nuances of each would suggest using slightly different IATs to assess them. The first cognition clearly refers to Ward and Keenan's (1999) implicit theory, the third to deviant sexual interest. The second cognition is a little bit vague and could refer to either, the same implicit theory as the first or perhaps to sexual interest. First of all each of these cognitions would most likely be investigated by comparing speed of child-sex associations to adult-sex associations. Child versus adult are opposite and mutually exclusive concepts, and the picking of stimuli that are easily classifiable as one or the other would cause little difficulty. The task becomes more difficult when choosing contrasting sex/not-sex categories. For the first

association mentioned above, "children as sexual beings", a contrasting pair of category labels might be 'sexual' and 'non sexual'. Virtually anything can be non-sexual but there are very few words that would satisfactorily form a coherent 'non-sexual' category. Mihailides et al. (2004) used words that were phonetically similar to their sexual words to form their non-sexual words (e.g. orgasm and sarcasm) resulting in a non-sexual category lacking coherency and a lack of contrast between the two categories. In addition, instead of an adult-sex versus child-sex comparison Mihailides et al. (2004) also used a 'child' versus 'not child' comparison with words that were phonetically similar to the child words forming the 'not child' words (e.g. kids and lids). The lack of coherent contrasting categories potentially makes the interpretation of what exactly was being measured by this task less clear than for other IAT designs.

Looking at the second association one might want to measure, that of "children being associated with sex or sexual activity", the creation of 'sex' vs. 'not sex' categories again causes difficulty. Gray et al. (2005) deal with this association by creating a coherent set of words not commonly considered sexual that were all related to the body as their 'not sex' category. It could be argued that they don't create a contrasting category so much as an unrelated category with a contrasting label. If one accepts that the association of 'children' and 'sex' is asking whether children are associated with sex or sexual activity it is likely that this will incorporate more than just a "children as sexual beings" implicit theory. However, this is unsurprising given that there is likely to be considerable overlap in how sexual schemas and sexual interest are represented in the brain. It is important to note that Grey et al. (2005) do not set out to explore cognitive distortion with their IAT. Instead they treat an association between children and sex (as measured by their IAT) as analogous with sexual interest (as measured by PPG).

Of the three studies looking at offenders against children, possibly the most successful in their creation of categories to assess the third association mentioned above (children as sexually attractive) is that of Nunes et al. (2007). They chose 'sexy' and 'not sexy' as the titles of their categories and were able to create a coherent category of 'not sexy' words that provide a real contrast to the 'sexy' words. 'Sexy' versus 'not sexy' was chosen instead of 'sexually attractive' and 'sexually unattractive'. However it could be argued that sexy captures a different association than 'sexually attractive' with 'sexually

attractive' being more passive compared to the slightly more active role the target stimuli (adult or child) may take in being 'sexy'. This distinction is not necessarily a criticism of the method used as the inclusion of sexy may incorporate some of the "child as sexual being" implicit theory into the "children as sexually attractive" association that Nunes et al. (2007) set out to explore. Nunes et al. chose 'sexy'/'not sexy' as they felt this was clearer, more concise as well as being easier to differentiate between categories compared to 'sexually attractive'/'sexually unattractive' (K.L. Nunes, personal communication, December 13, 2007).

Any study involving offenders against children must take into account that some of their participants may suffer from intellectual and/or learning difficulties (Langevin & Curnoe, 2007). This necessitates that researchers ensure that instructions, stimuli and categories are as clear as possible when using cognitive tasks. However, without this constraint Snowden, Wichter and Gray (2008) were able to carry out an IAT that had an excellent ability to discriminate gay men from straight men using *sexually attractive* versus *sexually unattractive* as their contrasting pair labels. As the study was interested in sexual interest towards adults only, they were also able to use images of adult males and females as their male and female stimuli (it is worth noting that the images used in this study were of nude adults). However, for use within an offending population, a design such as that used by Snowden et al. (i.e. using images instead of words, 2008) would not have the potential to elicit the concept of males and females of any age in participants and also there is a risk that the category labels used by that study would make the task a little more complicated.

Two very recent studies have used images in the design of their IATs. Brown et al. (2009) using a pictorial IAT design where images belonged to the categories of child and adult, and the second concept pair were sex and non-sex words were able to demonstrate a significant difference between paedophilic and ephebophilic type offenders. As with Gray et al. study (2005) the words used as non-sex stimuli were neutral words relating to the body or bodily functions. Banse et al (2009) created three different pictorial IATs with Man-Woman, Girl-Woman, and Boy-Man object categories respectively. For each of the three IATs *sexually exciting* (comprising of the words erotic, exciting, lustful, sensual, orgasm) or *unexciting* (dull, bland, indifferent, unexciting, boring) were the second pair of concept categories used. The IATs were set

up in such a way so that sexual orientation, measured by the Man-Woman IAT, was disentangled from 'sexual age preference' (Banse et al., 2009). In this study viewing time was also measured using the same images as were used for the IATs. Both viewing time and IATs demonstrated an ability to discriminate between offenders and controls that was comparable to the explicit sexual interest questionnaire used. Viewing time outperformed the IATs in terms of their ability to discriminate but both were correlated for most stimulus types. Perhaps most importantly a combination of implicit and explicit measures yielded a very high area under the curve value (.95) when ROC analysis was used indicating a very good ability to discriminate between offenders and controls. Again it is worth considering whether the *sexually exciting* and *unexciting* category labels in Banse et al. (2009) made the task a little bit more difficult for some of the offending participants. It might be that simpler labels would have yielded stronger IAT effects.

An alternative way of conducting the IAT that departs from the previous studies might use a control category instead of a contrasting category. Obviously, this decision would be at the expense of having an "opposite to sex" category, though in the case of Mihailides et al. (2004) and Gray et al. (2005) it could be argued that the opposite category was opposite in name only. It should be stressed that all three of the published studies, using the IAT with offenders against children, were able to identify group differences using their respective methodologies suggesting that any effect is robust enough to transcend methodological variations. On the other hand, movement towards a consistent methodology when conducting sexual interest IATs would allow for a better understanding of what exactly is being measured by this task and would also allow direct comparisons of studies carried out by different researchers. Given the success of the Snowden et al. (2008) study and the appeal of its simplicity, future IATs for the assessment of sexual interest may follow their method. Whether or not it is necessary to present nude images and whether 'sexually attractive' versus 'sexually unattractive' are suitable labels for participants of different abilities should be addressed by research. Another possibility is to do away with the not-sex category completely and to adopt a method like the Go/No-Go Association Task (Nosek & Banaji, 2001).

1.6.6   What does the IAT measure?

Establishing what the IATs mentioned above are measuring is potentially more complicated than determining what the Stroop is measuring. Since they are attempting to measure the strength of associations between concepts it is likely that they are measuring schemas. However since schemas related to sexual orientation are likely to be complicated and multifaceted in their associations it remains unclear whether an IAT designed to measure a component of that network of associations might not be discreet enough to measure anything other than the orientation. Consider the following example: An IAT designed to examine Ward and Keenan's (1999) *children are sexual beings* implicit theory would be seeking to see whether there were associations in the offender's thematic network of children's ability to decide about sex, their interest in sex etc. In order to test this, the IAT would have the concepts of child and also sex or sexual. An offender who held the cognitive distortion might show an IAT effect. However an offender who doesn't have the *children as sexual beings* distortion but does have a sexual interest in children along with an *entitlement* implicit theory (i.e. the offender does not see the child as sexually curious, receptive or capable but will use them for sexual gratification regardless) could still show an association between sex an children. It is probable therefore that the IATs will measure an overlap between cognitive distortions and sexual interest, though it is likely that much of this overlap is a result of how sexuality, both deviant and non-deviant is represented in the brain.

1.6.7   The relationship between sexual IATs, Stroop and CRT

It has already been suggested that the pictorial modified Stroop task (and the choice reaction time task) is a measure of sexual interest based on the fact that sexually salient stimuli will induce a delay or an increase in attention as a component of the sexual arousal process. In addition it has been suggested that the IATs used in this thesis will demonstrate a facilitation effect when concepts that are associated with sex in the sexual schema of an individual are paired together in the tasks. Given that cognitive distortions/implicit theories relating to sex and sexual interest are likely to overlap considerably there is likely to be considerable overlap between a Stroop task designed to assess sexual interest and an IAT designed to assess sexual associations. There is also likely to be overlap between both types of measure since neither is likely to be a 'pure' measure of association or interest. In the case of the IAT, it has been argued that using

50

children and *sex* as potentially associated concepts in an attempt to see if children are associated with sex or sexual activity may capture sexual interest, to an extent, as well as schematic associations. However if "a children as sexual beings" implicit theory (Ward & Keenan, 1999) can exist without deviant sexual interest it should be possible for an individual to demonstrate a Child-Sex IAT effect in the absence of a Stroop effect for images of children or other indicators of deviant sexual interest. Depending on whether or not the IAT does tap into sexual interest or not, and if it is possible to have deviant sexual interest without holding a distorted belief around children as sexual beings it should be possible to demonstrate a Stroop effect but not an IAT effect. Among non-offending controls the degree to which the results of the pictorial Stroop task and a Gender-Sex IAT correlate and whether they contribute separately to the prediction of sexual orientation may indicate the degree to which these two tasks are measuring distinct constructs.

1.7   Research Objectives

This thesis sets out to explore the clinical utility of several implicit tasks in the assessment of sexual interest. In addition the research is concerned with the theoretical implications of the presence or absence of agreement between different implicit tasks. The Implicit Association Test (IAT) should, in theory, be accessing an associative network and thus schematic representations of sex and gender or sex and age. On the other hand the pictorial modified Stroop task (and also the choice reaction time (CRT) task) should be a simpler measure of attention to sexually salient stimuli. Theoretically, therefore, the IAT should, at least in part, tap into cognitive structures represented as schema while the attention based paradigms of the pictorial modified Stroop task and choice reaction time tasks should, access sexual interest only. The convergence and divergence of association-based and attention-based implicit tasks are proposed to be a product of three things:

a) The degree to which schematic representations of sex and associated concepts overlaps with the process of sexual attention/arousal.

b) The degree to which association-based paradigms are capable of accessing schematic representations of sex and associated concepts, and the degree to which attention-based paradigms are capable of accessing sexual attention/arousal.

c) The sensitivity and specificity with which the particular methodologies used in the current series of studies successfully access the schematic representations and processes that are intended.

This thesis seeks to address several broad questions:

1) Can a pictorial modified Stroop methodology and IATs be shown to be accessing sexuality and sexual orientation among a sample of non-offending participants?

2) If they can, are they correlated or do they seem to be measuring distinct constructs? If they are distinct, does a combination of results from both paradigms give a clearer picture of sexual orientation?

3) If the pictorial Stroop and IAT tasks have been shown to have potential in accessing sexual cognitions and processes among non-offending participants do they also have utility in accessing deviant sexual interest and/or cognitive distortions among sexual offenders? Again do the converging or diverging results indicate that both paradigms are addressing the same or distinct aspects of sexual cognition?

4) Is it possible to increase the relationship between attention-based and association-based paradigms by varying the methodology used (i.e. using pictorial stimuli in both etc.)? If so, what are the implications a) for what each test is measuring and b) for the relationship between deviant interest and cognitive distortion?

The thesis, therefore aims to contribute to the literature in three broad ways. First the thesis will further the development of implicit methodologies for the assessment of sexual offenders. Second it will explore the convergent and divergent validity of those various implicit measures. Third the thesis will examine the safety of the *a priori* contentions that the IAT is measuring schematic representations of sex and related concepts and the attention-based paradigms of Stroop and CRT are accessing the attentional component of the arousal process. As part of this the thesis will comment on the appropriateness of interpreting previous findings using implicit methodologies with offenders as measuring either sexual interest or cognitive distortions.

Chapter 2: Methodological Issues in the application of Implicit Association Tests and a Pictorial Modified Stroop Task

This chapter deals with some additional methodological issues that were too specific for inclusion in chapter 1. The chapter looks at some of the factors that might influence responses and the measurement of responses for the implicit tasks adopted in the thesis. It also outlines the process of stimulus creation for the experimental tasks.

2.1 Potential factors influencing response latencies and scoring in reaction time studies

2.1.1 Volunteer bias

Volunteers for the five studies conducted through Trinity College were recruited using advertisements (via email, college notice-boards or internet forums) that clearly stated the sexual nature of the study. The information sent to prospective volunteers regarding the two studies that involved presenting child stimuli indicated the intention to also test sexual offenders. While volunteer bias is a concern for all experimental research in psychology, it can be especially problematic in studies involving sexuality (Trivedi & Sabini, 1998; Wiederman, 1999). Trivedi and Sabini (1998) conclude that studies involving sexuality may involve both selection bias and presentation bias. Wiederman (1999) also found differences among volunteers and non-volunteers for sexuality research using a population of undergraduates studying a psychology module. These differences related to level of sexual experience, sexual attitudes, sexual sensation seeking and other variables. While neither of these studies approximated the type of studies conducted as part of this research they do highlight the caution needed in generalising results to the larger population. The fact that the studies in this dissertation were studies, not just about sexuality but also involving sexual associations to children, is likely to have appealed to individuals that varied systematically on some traits of personality etc. from those who had no interest in participating. All the studies in this dissertation relied on self-report questionnaires to ascertain the sexual interests etc. of

participants[6]. In most cases this information was uncorroborated by other sources. If volunteers (especially control volunteers) for these studies were individuals who were, perhaps, drawn to such research because they had some confusion around their own sexuality then this may have impacted on their willingness or ability to clearly identify their true sexual interests in these questionnaires. In interpreting findings of these studies it was therefore important to consider whether volunteers were truly representative of the larger population and whether their self-reporting would be any more prone to inaccuracy than is normally the case with self-report.

2.1.2 Stimulus characteristics

Good methodology for the development of cognitive tasks dictates that stimuli used across different trial types should be as similar as possible in many aspects. Inevitably the experimenter will need to make subjective decisions about the appropriateness of stimuli when some of these aspects come into conflict. For trials using words, the length, frequency in common usage and valence (if valence isn't part of what's being tested) should be kept constant across conditions (e.g. Cox, Fadardi, & Pothos, 2006; McKenna & Sharma, 1995; both studies using a modified Stroop paradigm). Researchers also should attempt to keep many features of pictorial stimuli constant between trial types in order not to confound results (e.g. Constantine et al., 2001). In the present series of studies these features could include the size of image on the screen, the amount of target colour present in images coloured for use in a Stroop task, the attractiveness of the person depicted, how realistic the person was (given that the images were morphed, see paragraph 2.2), the demeanour and facial expression of the person, their pose or the amount of their bodies that are covered by clothing. Essentially these features all contribute to two conditions that a pictorial Stroop design attempting to discriminate people based on their sexual interest should satisfy:

1)   The structural makeup of the image should not affect the ability of participants to pick out the target colour.

2)   That all stimulus categories are capable of eliciting sexual interest in individuals orientated towards those categories.

---

[6] Self-reported sexual interest was unavailable for study 4 which consisted of participants from the Sand Ridge Secure Treatment Center in Wisconsin, USA.

54

Study one of the dissertation sought to address the first condition above. However since two of the categories in the Stroop tasks used in studies 1, 2 and 3 comprised of images of children it was not possible to address condition two above. This would have involved asking sexual offenders against children to rate the images of children for attractiveness. Including this in the proposed procedure may have been less likely to result in the granting of access from host organisations. In addition, the available population of offenders was so small that it was deemed more important that all available participants would take part in the Stroop and IAT study. Participants who were used to validate the stimuli would not be able to take part since they might demonstrate habituation to those stimuli. Finally, only offenders who admitted a sexual interest in children could have been used to validate the images and these represented a very small proportion of our final sample. As study 6 was not using the images of children it was decided to revisit the issue of the sexual attractiveness of the images when preparing new stimulus for the tasks involved in that study.

2.1.3 Blocked versus random design for modified Stroop task

In tasks such as the modified Stroop, stimuli can be presented either randomly across the entire task or with stimuli of a certain type blocked together. While conventional wisdom among experimental psychologists usually dictates that components of experiments should be randomised or pseudorandomised, wherever possible, several modified Stroop studies have shown a blocked design to elicit stronger attentional biases than non-blocked designs (Ballesteros, Reales, & Manga, 2000; Cox, Fadardi, & Pothos, 2006; Holle et al., 1997; Richards, French, Johnson, & Naparstek, 1992; Waters & Feyerabend, 2000). This finding is supported by a meta-analysis of 70 studies that showed a greater interference effect when the Stroop design was blocked (Phaf & Kan, 2007). Cox et al. (2006) recommend that categorical stimuli are grouped together in order that any carry-over effect (further processing of a given stimulus, after the participant has moved onto another trial, which might interfere with attention on the new trial) would be carried over onto similar trials as opposed to unrelated trials. This would mean that any additional cognitive interference caused by a stimulus would contribute to the overall reaction time for the stimulus category. The modified Stroop effect may therefore be more a product of slow carry-over effects than fast interference effects (Algom, Chajut, & Lev, 2004; McKenna & Sharma, 2004; Phaf & Kan, 2007).

These slow effects may be a result of rumination that occurs over the course of the block of trials (Foa et al., 1991; Holle et al., 1997).

It is likely some stimuli have the ability to 'grab' a significant portion of attention quite instantaneously. In those instances a randomised presentation of trials would detect the effect. It is also likely the attentional demand created by this grabbing would not dissipate immediately after that stimulus has been removed from the computer screen. Therefore a reduction in attentional resources would quite likely be present for one if not a couple of subsequent stimuli. In such a scenario, a task design where potentially salient stimuli are followed by 'filler' stimuli (e.g. Waters, Sayette, & Wertz, 2003) would find an increased reaction times for these 'filler' words when they are preceded by an attention grabbing word compared with a neutral word. It is likely that the number of filler words effected and the amount of interference (i.e. the degree to which it was more difficult to name the colour in a Stroop task trial) caused by the presentation of the initial salient stimuli would be an indicator of exactly how salient the stimuli were. For stimuli such as these, a blocked design, where the block contained many similarly salient/attention grabbing stimuli would again show an effect. However it would be difficult to judge the duration of any carry-over effects for single stimuli (since filler stimuli would now be potentially salient) and there would be the potential for higher order rumination effects to occur. To use the example of an arachnophobic participant, a block of spider images could produce the following components:

1) There may be an immediate reaction to an image of a spider which interferes with the participants ability to name the colour in which that image was presented

2) There may be a residual reduction of attentional resources available to carry out the colour-naming task spanning subsequent trials thus slowing down response times. This may be compounded by an additional grab of attention if the subsequent trial is again highly salient/threatening (note that all images in a block may not be salient to all participants).

3) There may be additional rumination of the stimuli by the participant contributing to a general block-wide slowdown.

As implied some stimuli may not be quite as attention grabbing as the spider stimuli to the arachnophobic in the above (hypothetical) example. Stimuli that do not capture

56

attention immediately or that do not capture sufficient attention to significantly impede colour naming for the trial on which they are presented, would not produce significant differences between trial types on tasks where stimuli were not blocked. It is possible that a Stroop effect for stimuli that are appetitively salient, as opposed to threateningly salient, is influenced more by rumination than by faster effects that are related to a 'grabbing' of attention. It makes sense that when a threat is perceived that attention would be more sharply focused. Since rumination is likely not only to cause slowdown in carrying out the low-priority (given the presence of a threat) colour naming task it is also possible that it might impede threat monitoring. It is unlikely that rumination is responsible for the entire modified Stroop effect for appetitive stimuli since randomised studies, and studies that had 'filler' trials but that were otherwise randomised, have produced effects. It would be logical to assume that fast effects in studies of appetitive stimuli would be more pronounced depending on the degree to which participants had addictions or compulsions relating to those stimuli. It is interesting to consider whether fast effects are, therefore, more strongly related to pathology than slower effects.

In conclusion, while it is possible that different research questions may be best served by adopting blocked or randomised Stroop designs, a blocked design will theoretically give the most chance in an exploratory study to determine whether any effect of stimulus content exists.

2.1.4 Habituation and novelty

Habituation in cognitive tasks such as those used in this series of studies refers to the fact that "interference effects [may] diminish with stimulus repetition" (McKenna & Sharma, 1995, p. 1596). McKenna and Sharma (1995) demonstrate that, in a modified Stroop paradigm, that the effects of emotional stimuli can habituate over several blocks but that this habituation can be ameliorated, though not eradicated, by the inclusion of novel stimuli in each of the blocks. In a task investigating sexual interest Santtila et al. (2009) demonstrate that only their first phase of trials in a choice reaction time (CRT) task has utility in discriminating participants by sexual orientation. A Receiver Operating Characteristic (ROC) curve is a "function that relates hit and false-alarm rates to each other" (Macmillan & Creelman, 1990, pp. 62-63). Chance alone should yield an AUC of .5 while a task that predicted sexual orientation correctly all the time would

have an AUC of 1. Santtila et al. (2009) find that using the difference in reaction times between male and female sexually explicit stimuli to predict sexual orientation yields an area under the ROC curve of .82 in the first phase which falls to .63 and .60 in the second and third phases respectively. Interpretation of this result is potentially problematic, however. Target images were primed with explicit or non explicit images of males and females in order to test a second hypothesis. This may have impacted on the habituation of the stimuli. Stimulus types were randomised, in each phase, so any carry over effects affects may have increased reaction times to non-salient stimuli, thereby reducing the size of the effect. The task contained 480 trials over three phases, it is not surprising that there would be a reduction in effect in a task that typically took over 21 minutes to complete. This paper contains inconsistencies when stating number of phases and images[7] so it is difficult to be confident in the accuracy of the methodology section. However, it seems that novel images were not used in each phase. As a result, it is not clear from the study, whether the habituation effect occurred because of habituation to the stimulus categories or to the stimuli themselves.

Arousal, measured by penile plethysmography (PPG) and self-report, to erotic stimuli has been found to habituate with repeated viewings of the same material (Koukounas & Over, 1999, 2001). In addition repetition of erotic stimuli were found to increase the eyeblink startle response, indicating a reduction of appetitive interest in the stimuli (Koukounas & Over, 2000). This pattern mirrored a reduction in arousal (PPG) and both trends were reversed with the inclusion of novel stimuli. The erotic stimuli used in the cited studies were explicit film clips. It remains unclear whether the same would apply to non-pornographic still images. Therefore the findings may be limited in their applicability to the present research.

2.1.5 Age, intelligence and cognitive ability

Many studies including meta-analytic studies have demonstrated that ageing is associated with a general slowing down of response times on both lexical and non-lexical tasks (Faust, Balota, Spieler, & Ferraro, 1999). In addition to again showing group differences in various response time tasks between older and younger adults

---

[7] Different sections refer to there being 3 or 4 phases. The method section indicates that there were 12 versions each of 40 male and 40 female images used.

Sheppard and Vernon (2008) in a large review of studies on intelligence and speed of processing demonstrate robust correlation between measured intelligence and response times. Interpreting differences in response latencies between participants groups based on raw millisecond reaction times alone must therefore be done with caution. Methods of accounting for individual differences in speed of processing are considered when analysing the studies presented in this thesis. Individual differences in age and intelligence are more likely to be pronounced in the studies in which participants and controls are drawn from outside the college population. For practical reasons it was not possible to apply a measure of intelligence for those studies. Level of education is recorded as an, admittedly rough, indicator of comparability of groups in this regard. In addition several of the studies in this series incorporated a traditional Stroop task into the design of the study. This was done for two reasons; first to give the participant some practice in how to conduct the subsequent pictorial Stroop task and second to produce at least one measure of cognitive ability. In this vein the Stroop task can be used as a measure of selective attention and cognitive flexibility (Strauss, Sherman, & Spreen, 2006). While the Stroop is too limited a measure to assess whether different samples of participants are comparable in terms of cognitive ability it should be an adequate measure of task-specific cognitive ability.

2.1.6 Function of the distribution

Reaction time data do not typically follow a normal, or Gaussian, distribution. This is because it is unlikely that there will be many extreme values in the fast-response portion of the distribution. Typically responses under 100 or 200 milliseconds are errors or are responses that were inadequately processed but happened to be correct, i.e. fast guesses (Whelan, 2008). In addition, the opposite tail of the distribution can stretch out quite far. It is a challenge to researchers to determine which of these slower responses represent true responses, where the cognitive processing of that trial required the length of time, and which should be considered responses where there was some interruption to the cognitive processing that means that the inclusion of that particular response time would not contribute to the research question of interest. Distributions in reaction time studies therefore are typically positively skewed. Many authors consider them to be best approximated by the ex-Gaussian distribution which is a distribution that comprises of a combination of the Gaussian and exponential functions (Ratcliff, 1993; Whelan, 2008).

When looking for within- or between-group differences using variables that follow a Gaussian distribution, methods of parametric analysis typically look at the mean or mu (μ) and standard deviation or sigma (σ) of the distribution. In addition to these components group differences in ex-Gaussian distributions may be located in tau (τ). Differences in τ would indicate differences in the exponential tail of the distribution. Some authors argue that such differences can be indicative of the cognitive processes involved in the task though this contention remains controversial. For example Balota and Spieler (1999) tenuously suggest that the Gaussian component of the distribution represents automatic or non-analytical processes, while the exponential component represents attention demanding or analytical processes. While it may not be necessary to explore the distributions in the current study in such depth in order to identify group differences, it is important to consider the overall impact of the independent variables on the distribution of responses, not just its impact on the mean and standard deviation. In addition, the sample sizes in the current studies are relatively small and there are limited numbers of trials in each Stroop trial type category, which would make an in depth analysis of the distributions difficult.

2.1.7   Scoring issues with the IAT

Treatment of extreme data proposed by Greenwald and his colleagues (e.g. Greenwald, Nosek, & Banaji, 2003) involves the removal or replacement of data if certain thresholds are passed. Depending on which algorithm is used this may involve the removal of trials if reaction time is greater than 10 seconds (new *D* algorithm) or the replacement of reaction times with 3 seconds if actual reaction time is over that (conventional algorithm). Since participants in reaction time studies typically have quite short average responses it seems inadvisable to potentially include reaction times of 10 seconds (or even 3 seconds, if one was using the conventional scoring algorithm) in the data set. A method of removing outliers based on the interquartile range of responses of each participant seems much more prudent. However, it follows from section 2.1.6 that any method that removes extreme reaction time data, especially from the portion of the distribution that includes longer response times runs the risk of excluding genuine responses or including true outliers. Determining an appropriate cut point is a challenge to all such research.

Since the goal of research using the cognitive methods outlined in this volume is to identify procedures with clinical uses, we must consider how these tests perform on an individual level. Extreme value treatment must attempt to best capture the individuals true typical response while minimising the noise introduced into the data by distraction, lapses of attention, cognitive impairments, fatigue or loss of interest. Selecting arbitrary cut-points beyond which data are removed or recoded is a method that may suit the improvement of global data but should not be applied when attempting to best interpret the results of one individual. Blanton and Jaccard (2006) raise several criticisms of the IAT including questions about the meaningfulness and arbitrariness of the scoring method used. Questions raised include issues surrounding the potential of distorted results for individuals with small standard deviations in their reaction times and also the danger of interpreting effect sizes using Cohen's (1988) definitions without also looking at unstandardised scores. Given these concerns and the exploratory nature of the current research and also given the fact that the methodology adopted differed somewhat from standard IAT methodology several different methods of scoring the IAT were attempted. Interestingly three previous published studies that looked at sexual interest/cognitions among offenders using the IAT (Gray et al., 2005; Mihailides et al., 2004; Nunes et al., 2007) each used different methods of data assessment, again suggesting that the phenomenon may be robust enough to materialise despite differences in consensus regarding the best analysis method.

2.2 Stimulus creation

The choosing of word stimuli for all experimental tasks was carried out by compiling lists of previously used words (e.g. Banse et al., 2009; Gray et al., 2005; Mihailides et al., 2004; Nunes et al., 2007; Smith & Waterman, 2004), expanding on those lists using the authors own judgement and, in the case of the stimuli for study 6, collaboration with other researchers carrying out similar research, and finally excluding words from those lists on the grounds of word length, frequency and also on other *a priori* grounds.

Stimuli used in the pictorial modified Stroop tasks, pictorial Gender-Sex Implicit Association Test and Choice Reaction Time task consisted of images of adult males,

adult females, female children, male children[8] and control images of large cats (lions, tigers etc.). All images of people were clothed in bathing suits or shorts. Child images were taken from the commercially available Not Real People (NRP) image set created by Laws and Gress (2004). Some of the adult images were also taken from the NRP. The NRP is a set of males and females images that have been computer modified for use in the assessment of and research involving sexual offenders (Laws & Gress, 2004). "Each finished image was produced by compiling and morphing three or more images, plus additional modifications such as hair, eye and body colour, simple pose modifications and clothing" (Laws & Gress, 2004, p.188). Images in the NRP are classified according to their secondary sexual characteristics, using Tanner stages (Tanner, 1973). While the NRP does contain some images of adults it was necessary to create more images of adults. Not only were there not enough adult images in the stimulus set but some were also deemed to not be of high enough quality for inclusion. Subjective inspection of the NRP set suggests that some of the image categories (especially, Tanner four and five males and Tanner five females) have images that are not as realistic as images in some of the other categories. Perhaps this is a result of the developers having fewer (or lower quality) original[9] images in those categories. Unique adult images were created by the author using a similar technique to those adopted by Laws and Gress. It was decided to attempt to mirror the process by which the NRP images were created as closely as was possible since morphed images are likely to vary systematically from non-morphed images in eliciting reaction times.

Images were created using Adobe Photoshop Elements 3.1© and two image morphing programs[10]. Morphed images comprised of two faces morphed together onto one body. The backgrounds of all images (NRP and novel morphs) were removed and the images coloured using the 'Colour Replacement Brush' tool in Photoshop©. It was found that using this technique to alter the colour of the images instead of simply placing a coloured filter over the image produced images that were much clearer since the colour replaced images retained a greater deal of shading, thus minimising the changes to the original pictures. The removal of the background ensured that in order to respond to the

---

[8] Images of children were only used in the pictorial Stroop tasks used in studies 1, 2 and 3. All other tasks using pictorial stimuli had images of adults and control images of large cats only.

[9] These original images came from a set of nude images of children recovered from child pornography investigations, which were used in PPG laboratories in North America and were therefore possibly limited by the number of high quality original images in some categories.

[10] Morpher© and WinMorph© (Fujimiya, 2001; Kumar, 2002)

image participants would have to attend to the image itself and not be able to rely on any area of tinted background. Large cats were chosen as control images, as it was important that the control images form a coherent category (Cox, Fadardi, & Pothos, 2006) and also that these controls should be animate. Cats were chosen over other potential animate controls, apes for example, mainly for the reason that the colour replacement method used worked better with images containing a good amount of contrasting colours. Unfortunately, in the clothed version of the NRP, the clothing worn by the female images varies depending on the Tanner stage of the image with Tanner stages one to three wearing one-piece bathing suits, while the four and five Tanner stage images are wearing two piece bathing suits. To ensure the same ratio of two-piece to one-piece bathing suits between child female and adult female images (in tasks where child images were being used) two versions of some of the adult female images were produced (i.e. a one-piece and a two-piece bathing suit version). In total 39 novel images of adults were created by the author. 15 images of large cats were also modified. Fifteen of these adult images and 12 of the cat images were initially created and were used in studies one, two, three and four. The remaining 24 adult and 3 cat images were created for study five. Example images are included in Appendix 9.

Chapter 3: Developing implicit tasks that tap into sexual interest and associations

The two studies in this chapter explore the utility of a pictorial modified Stroop task and two Implicit Association Tests (IAT) in assessing sexual interests and sexual associations. The first study looks specifically at a version of the pictorial Stroop task where the images used are scrambled. This is to test whether the colour content of the various images used is consistent enough to produce uniform responses when the content of those images is not discernable. The second study then uses the unscrambled version of the pictorial modified Stroop task in combination with two IATs to compare the responses of gay and straight participants across all three tasks.

3.1   Study 1: comparing reaction times to scrambled versions to images from various trial types

3.1.1   Introduction

The process of creating images for use with a pictorial modified Stroop task has already been described in chapter 2 as has the process of colouring those stimuli. In order to determine whether the content of the images used in a pictorial Stroop can induce a delay in reaction time it is necessary to know that the coloured images themselves don't vary to any great degree in how easy or difficult it is to identify the target colour. In the case of the images designed for use in the pictorial Stroop task it was possible that there might be differences in the amount of dark and light areas and also shading and tone present in the original which could influence the amount of the target colour that was present in the finished image. There could also be differences in the area of the screen taken up by the person or large cat depicted. To test this, the coloured images were all scrambled so that the content (cat, adult or child) was no longer discernable. It was then possible to use these scrambled images in place of the unscrambled images in a blocked Stroop task to investigate whether differences in the amount of target colour present in the images alone could result in systematic differences between response times and error rates for different trial types (i.e. adult female, adult male, child female, child male and cat). Also explored was whether reaction times and errors differed across the four

different stimulus colours. In addition the design allowed investigation of whether there would be an influence of order of block presentation on reaction times and errors. The study had three hypotheses. The first hypothesis was that there would be no difference between reaction times and error rates across the different trial types. Second, there would be no difference between reaction times and error rates across the four stimulus colours. Third there would be a significant effect of order on reaction times, with the first presented block being slower than subsequent trials.

3.1.2 Method

3.1.2.1 Design

The study adopted a repeated measures design with all participants responding to each of five trial types: scrambled images of male adults, female adults, male children, female children and large cats. Each image was presented in four colours: blue, green, red and yellow. Trials were presented in blocks, where each image in the block belonged to one image category (e.g. all male adult images were presented together in a block). There were 48 images in each block (12 different images coloured in each of four colours and then scrambled). Order of block was randomised by the stimulus presentation software used, as was order of trials within each block. Reaction times, in milliseconds, were recorded for each trial as well as whether the trial was answered correctly.

3.1.2.2 Participants

Eleven males completed the scrambled image task. Participants consisted of undergraduate students participating for course credit.

3.1.2.3 Apparatus/Materials

Computerised tasks were presented on a Gateway Pentium III computer with a Gateway EV9108 19" CRT monitor. Participant responses were made via a Cedrus response pad (model RB-620) with four coloured buttons (red, green blue and yellow)[11]. Tasks were

---

[11] While section 1.6.2 states that oral response is considered most accurate for Stroop tasks a button-box method was chosen here for the reason that some testing was anticipated to take place in situations

run using the SuperLab4© stimulus presentation software. Stimuli consisted of scrambled images of male and female children and adults in bathing suits along with control images, also scrambled, of large cats. Images of people were taken from the NRP image set (Laws & Gress, 2004) or were created by the author (see section 2.2). All images had blue, green, red and yellow versions (colouring of images followed the procedure outlined in section 2.2). Images were coloured prior to scrambling. Images were scrambled using a piece of bespoke software. The images were divided into a grid with 32 columns and 42 rows. The squares of the grid were then jumbled by the program's algorithm. Figure 3.1 shows an example of a scrambled and unscrambled image.

Figure 3.1: Scrambled and unscrambled green version of an image of a male child taken from the NRP set (Laws & Gress, 2004).

3.1.2.4 Procedure

Participants were seated approximately three feet from the monitor and responded during the task via the button box which was placed on the desk at a comfortable distance in front of them. Participants were instructed to use the index and middle finger of each hand to respond. All instructions for the task were presented on the computer

---

without the guarantee of absolute silence (background noise could impact on the accuracy of response times based on oral responses),

screen. Participants were directed to identify the colour of the scrambled images and to respond as quickly and as accurately as possible.

3.1.3   Results and Discussion

3.1.3.1   Impact of trial type

A repeated-measures ANOVA indicated that there was no significant impact of trial type on reaction times; $F (4, 40) = .3$, p = .876, partial $\eta^2 = .029$. A further repeated-measures ANOVA indicated that there was no significant difference across trial types on the amount of errors committed; $F (2.304, 23.035^{12}) = .17$, p = .872, partial $\eta^2 = .017$. Results therefore support the hypothesis that there would be no significant differences in reaction times or error rates across trial types. This suggests that any future differences in means between trial types, when unscrambled images are used, are likely to be due to the content of the image rather than any additional difficulty in colour naming.

3.1.3.2   Impact of stimulus colour

An ANOVA was carried out comparing the reaction times to trials in each of the four colours. The result of the ANOVA was approaching significance; $F (3, 27) = 2.772$, p = .061, partial $\eta^2 = .235$. Means indicated that the longest reaction times were to green stimuli (Mean = 598.8ms, SD = 128.6ms) compared with red (Mean = 560.4ms, SD = 120.5ms), yellow (M = 559.1ms, SD = 108.8ms) and blue (M = 555.8ms, SD = 101.9ms) stimuli. The non significant result here supports the hypothesis that there would not be a significant difference between reaction times for trials of the four different colours. However, the result in this case was approaching significance, with green stimuli yielding the highest mean reaction time. It is worth noting that green stimuli yielded the highest scores across all trial types. The interaction between trial type and colour was not significant; $F (12, 108) = .563$, $p = .867$, partial $\eta^2 = .059$. Therefore, if the green stimuli take slightly longer to identify the effect should be the

---

[12] For this ANOVA Mauchly's test indicated that the assumption of sphericity was unsafe, $\chi^2 (9) = 20.887$, p = .015, therefore the degrees of freedom were changed accordingly using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .576$).

same across all unscrambled trial types and should, therefore, not confound future results.

Two participants had incorrectly identified all or almost all of the yellow trials as green. After removing the scores of these two participants, an ANOVA was conducted, comparing errors across trials in each of the four colours which indicated no significant effect of stimulus colour on error rate, $F(3, 24) = .629$, p = .603, partial $\eta^2 = .073$. While there is no significant difference between the error rates across stimulus colour, the necessity to remove two participants indicate that there was a systematic error in those two cases. It is not clear why this was the case. Participants were asked beforehand whether they were colour-blind. However it may be that these two participants were somewhat colour-blind but unaware of it. Alternatively, as yellow is not a primary colour, the process of colouring the stimulus yellow (as outlined in Chapter 2) may not produce stimuli that are quite as vibrant as the blue, red and green stimuli. The participants may have simply felt that the stimuli looked more like green than yellow. As with the potentially elevated reaction times for green stimuli the apparent difficulty for two participants to correctly identify yellow stimuli should not cause any problem in future unscrambled studies, should participants face similar difficulties, since the errors were committed across all experimental trial types.

3.1.3.3   *Impact of block order*

A repeated-measures ANOVA was carried out to test whether block order (regardless of trial type) had an effect on reaction times. Results indicated that mean reaction times did not differ significantly as a result of order of block presentation*; F* (1.924, 19.236[13]) = 2.537, p = .107, partial $\eta^2 = .202$.   While no significant differences were found between reaction times based on block order, visual inspection of the mean reaction times indicate that the first presented block did yield the slowest response indicating that order could potentially have an influence on reaction times. The hypothesis that order of block presentation would influence reaction times was not supported. While

---

[13] Mauchly's test indicated that the assumption of sphericity had been violated, χ^2 (9) = 21.676, p = .011, therefore the degrees of freedom were adjusted using Greenhouse-Geisser estimates of sphericity (ε = .481).

clearly non-significant the slower reaction times for the first presented block followed the pattern predicted by the hypothesis.

An ANOVA comparing error rates across the order of block presentation indicated that error rates did not differ significantly across the blocks; F (2.021, 20.214[14]) = 1.267, p = .299, partial η^2 = .112. This indicated that errors were no more likely as a result of order of block presentations. No specific hypothesis had been made regarding this.

3.1.4   Conclusions

Results of the study indicated that no one set of images (male adult, female adult, male child, female child or cat) was more difficult to classify by colour once the images had been scrambled. Results also indicated that any difficulty in identifying individual colours, though approaching significance, were equally likely to occur across trial types. In addition findings suggested that the effect of order may not be as large as had been hypothesised. Overall results indicated that differences found between trial types in any future study are likely to be a product of the content of the images and not of differences in the amount of target colour present in the image.

3.2   Study 2: the measurement of sexual interest and sexual associations among non-offending participants using a pictorial modified Stroop and two Implicit Association Tests

3.2.1   Introduction

The results of study 1 indicated that any delays in the amount of time it takes to name the colour of the stimuli prepared for a pictorial modified Stroop task are likely to be due to the content of those images. If the content of those images is salient enough to the participant to co-opt some of the attentional resources that should be allocated to carrying out the task of colour naming, some image categories may produce a

---

[14] Mauchly's test again yielded a significant result, χ^2 (9) = 20.042, p = .02. Greenhouse-Geisser estimates of sphericity (ε = .505) were used to adjust the degrees of freedom for the ANOVA.

slowdown in reaction time and/or an increase in wrong responses. In the case of the Implicit Association Test (IAT), of which there are two in this study, there should be a speeding up of responses when words are paired together that are compatible with the participants cognitive associations. Conversely, there should be a slowing down of response time when the participant is faced with a pairing of responses incompatible with the associations that they hold. In the current study, the results of the modified pictorial Stroop task and both IATs are hypothesised to relate to participants' sexual orientation. Specifically, as theorised in Chapter 1 a pictorial Stroop effect for sexually salient material may be caused by both pre-attentive and attentive processes in sexual arousal. In essence, this would yield an indirect measure of sexual interest. On the other hand an IAT effect for sexually relevant material is theorised to indicate the strength of the associations that the participant makes between these sexual concepts. As explored in chapter 1, the IAT is therefore likely to be a measure of sexual schemas or scripts. In the case of gay and straight participants taking part in this study, there is no reason to suspect that there would be a discrepancy between their sexual schemas and their sexual interests. It is hypothesised therefore that both the Gender-Sex IAT and the pictorial Stroop task would correlate with one another and that both would be indirect measures of sexual orientation. 'Indirect measure' of orientation is used here since the Stroop task is thought by the author to measure one component of the arousal process, which is indicative of sexual interest and thus orientation, while the IAT is thought to measure sexual schema and therefore arrive at an estimate of orientation by a slightly different route. The Age-Sex IAT should bear a relationship with the results of the Stroop task, but only if participants hold a strong association between the concepts of adulthood and sex[15]. The current study had two overarching (and related) goals. The first was to examine the relationship between the three experimental tasks and sexual orientation and their relationship with each other. The second goal was to pilot the three tasks for future use with men who had committed sexual offences. Because of the exploratory nature of the study several different methods of data treatment were attempted at each stage of the analysis.

In addition to a general exploration of the data there were some specific hypotheses made. Before completing the pictorial Stroop task and both IATs, the participants

---

[15] Alternatively some participants may hold a child-sex association or schema. See section on volunteer bias in chapter 2 (section 2.1.1).

carried out a traditional Stroop task. For this task it was hypothesised that congruent trials would be significantly faster than control trials, and both would be significantly faster than incongruous trials. Further, there would be no difference between gay and straight participants in their results on the traditional Stroop task. Age would significantly correlate with overall response times on the traditional Stroop task.

For the pictorial modified Stroop task it was hypothesised that participants would differ in their patterns of response times to the five image types depending on their sexual orientation. For adult images, gay participants would have higher reaction times to male images relative to female images, while the opposite pattern would be true for straight participants. It was hypothesised that images of children would produce a lower reaction time relative to response times for adult stimuli for both gay and straight participants and that there would be no difference between the two groups in their reaction times to cats. There was no prediction made as to whether relative differences in reaction times to images of male and female children would be in line with participant's adult orientations. It was also hypothesised that age would correlate with pictorial Stroop reaction times.

Since the participants in this study were non-offenders, it was hypothesised that participants would demonstrate a stronger association, and therefore quicker reaction times, when adult and sex words were paired together compared with when child and sex words were paired in the Age-Sex IAT. It was further hypothesised that there would be no interaction between sexual orientation and age category-sex associations. For the gender-sex IAT it was hypothesised there would be no main effect of gender of the concept-sex pairing but that there would be an interaction between gender of concept-sex pairing and sexual orientation of participants. There was no specific prediction made about the degree to which both the Stroop and the IATs would converge or diverge. It was expected that the measures would correlate but that the strength of that correlation might be indicative of the extent to which both measures are tapping into the same cognitive processes.

3.2.2 Method

3.2.2.1 Design

A mixed factorial design was used with gay and straight participants completing all experimental tasks; the pictorial modified Stroop task, the Gender-Sex IAT, the Age-Sex IAT as well as a version of the traditional Stroop task.

3.2.2.2 Participants

Participants were recruited through university and community internet notice-boards along with university postering. Recruitment materials called for 'males of all sexual orientations'. Some participants took part for course credit, while others were offered €10 to compensate them for their time. Twenty one males were recruited, eight of whom classified themselves as having an exclusive sexual preference towards males while eleven stated an exclusive sexual preference towards females. One participant identified himself as having a preference for males with some sexual interest in females, while another stated that he had a preference for females with some sexual interest towards males. In the absence of a large number of individuals with non-exclusive sexual preferences these two participants were included in the gay and straight categories respectively yielding an overall sample of 9 gay and 12 straight participants. All participants were college educated and had fluent English (however 3 gay and 3 straight participants were not native Anglophones). Gay participants had a mean age of 25.2 years (SD = 4.7 years) and straight participants had a mean of 27.8 years of age (SD = 10.2 years).

3.2.2.3 Apparatus/Materials

Tasks used the same apparatus and presentation software as study 1. A 2-item questionnaire was also administered asking the participants about age and sexual orientation (appendix 3). Stimuli included images of male and female children and adults in bathing suits along with control images of large cats. As already mentioned in chapter 2 all images of children and some of the adults were taken from the NRP image set (Laws & Gress, 2004).

3.2.2.4 Procedure

Participants first filled in the questionnaire and then carried out the Stroop section. Responses were given via the response pad. All instructions for the task were presented on the computer screen. The Stroop section contained several sub-sections. First, participants completed a set of practice trials where words were presented in the different colours. Feedback was given in the form of a red X appearing if the wrong colour was selected. Participants could not continue to the next word until they gave the correct answer (this was to ensure that participants properly understood the procedure; no feedback was given on the subsequent sections and participants had no opportunity to correct mistakes). Second, participants completed a traditional Stroop task with congruent, control and incongruent font-colour to word pairings. The three trial types were presented in separate blocks with order of block presentation randomised across participants. In the final part of the Stroop section participants were presented with the modified pictorial Stroop condition. Stimuli were grouped according to stimulus type resulting in five blocks: adult males, adult females, child males, child females and large cats (controls). Chapter 2 (section 2.2) describes the creation of these stimuli. Each block contained 12 images in four colours, yielding a total of 48 images per block. Blocks of child stimuli contained three images in each of four Tanner stages. As with the traditional Stroop task, participants had to identify, using a button box, the colour with which each image had been tinted. Block order was randomised across participants and trials were randomised within blocks. Response time was recorded for every trial along with whether that response was correct or not.

After completing the Stroop task participants had a short break before proceeding on to the first IAT. Order of IAT type (Gender-Sex, Age-Sex) presentation was randomised across participants with the procedure for both IATs being exactly the same. All instructions were presented on the computer screen. Participants were instructed to use the two outer buttons (left and right) on the response pad. In the first block of trials participants categorised words that appeared on the screen one-by-one into one of two categories (e.g. male names versus female names in the Gender-Sex IAT, child words versus adult words in the Age-Sex IAT) using the left and right buttons. In block 2, participants classified words as being sex words or furniture words in the same way. Once this had been practiced, categories were combined in blocks 3 and 4 so that

categories were paired with each other (e.g. push the left button if the word that appears on the screen is either a sex word *or* a child word; push the right button if the word is either a furniture word *or* an adult word). The procedure and stimuli for blocks 3 and 4 were identical except that block 4 had double the amount of trials. Next, the order of male and female names, in the case of the Gender-Sex IAT and child and adult words in the Age-Sex IAT were reversed. Participants practiced this new order in block 5. Categories were again paired in blocks 6 and 7 but this time with the new order practiced in block 5 was used so that, in the case of the Age-Sex IAT, child was now paired with furniture words, while sex and adult were paired. As with block 4, block 7 had twice the amount of trials as block 6. For both IATs participants could only move on to the next word once they answered correctly. Time to initial response was recorded as well as timing of every subsequent response until the correct answer was given. All participants completed the Gender-Sex IAT and the Age-Sex IAT.

3.2.3   Results

3.2.3.1   *Results of the traditional Stroop task*

A one-way repeated-measures ANOVA was carried out comparing reaction times to stimulus in the traditional Stroop part of the task. Outliers over 1.5 times the interquartile range beyond the 25th and 75th percentiles had been removed. Reaction times to congruous, incongruous and control words were compared and were found to differ significantly from each other*; F* (1.45, 29.002[16]) = 27.915, *p* < .001, partial η^2 = .583. Reaction times were quickest for words where font colour and meaning were congruous (Mean = 720.1ms, *SD* = 166.35ms). Next came times for words where meaning had no relationship to font colour (Mean = 747.1ms, *SD* = 164.62ms). Reaction times were slowest for words where font colour and meaning were incongruous (Mean = 845.83ms, *SD* = 198.36ms). When post-hoc tests were carried out using the Bonferroni method[17], the incongruent word reaction times were found to be significantly slower than both the control words (*p* < .001) and the congruent words (*p* < .001). Congruent and control words did not differ significantly (*p* = .098). The

---

[16] Degrees of freedom were adjusted using Greenhouse-Geisser estimates of sphericity since Mauchly's test of sphericity indicated that sphericity could not be assumed.
[17] The Bonferroni method instead of the Tukey method was used since sphericity was violated. However the Bonferroni method can be too conservative and therefore a, less conservative, Dunn-Sidak method was also conducted which yielded the same result.

74

inclusion of sexual orientation as a between-subjects factor (creating a 3x2 mixed factorial ANOVA) indicated no interaction between sexual orientation and the interference effect of the traditional Stroop task. Age correlated positively and significantly with each of the three conditions; congruous, $r = +.435$, $n = 21$, $p = .049$, two-tailed; control, $r = +.448$, $n = 21$, $p = .042$, two-tailed; and incongruous, $r = +.511$, $n = 21$, $p = .018$, two-tailed. Taken as a whole, a traditional Stroop effect was found, where words with meanings that were incongruent with font colour resulted in significantly longer reaction times compared to both words where meaning and font colour were congruent and words where there was no relationship between meaning and colour. There was no interaction of sexual orientation and traditional Stroop effect but age did correlate positively with reaction times across all three trial types.

3.2.3.2 *Initial analysis of the pictorial modified Stroop task*

A 2x2x2 mixed factorial ANOVA was carried out on the results of the modified Stroop task with sexual orientation as the between-subjects factor and gender of stimulus images as the first within-groups factor and age of stimulus images (child vs. adults) the second. Again reaction times that were more than 1.5 times the interquartile range beyond the 25th and 75th percentiles were removed. There were significant main effects of age of stimulus as well as for gender of stimulus along with a significant interaction of gender of stimulus with sexual orientation. However, since there was a significant three-way interaction of gender of stimulus, age of stimulus and sexual orientation of participant all findings were interpreted in light of this; $F(1,19) = 8.616$, $p = .008$, partial $\eta^2 = .312$. In order to identify the source of the interaction further analyses were carried out. First, separate 2x2 mixed factorial ANOVAs were carried out, looking at child and adult stimuli separately (see Figure 3.2 and Figure 3.3). A Bonferroni conversion was applied to these results to reduce familywise error (α set at .025). For child stimuli no interaction was found between gender of stimulus and sexual orientation, however there was a main effect of gender of stimulus with participants, regardless of sexual orientation, typically taking longer to react to images of male children (Mean = 728.4ms, SD = 111.2ms) than images of female children (Mean = 680.4, SD = 82.9); $F(1, 19) = 7.25$, $p = .014$, partial $\eta^2 = .276$.
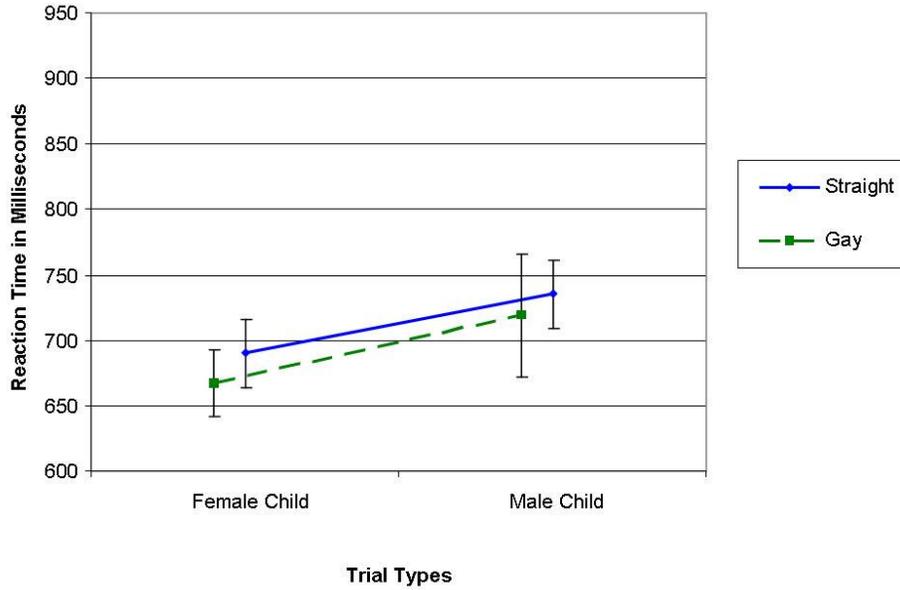
Figure 3.2: Mean reaction times of gay and straight men to images of male and female children from the pictorial modified Stroop task (error bars indicate one standard error).
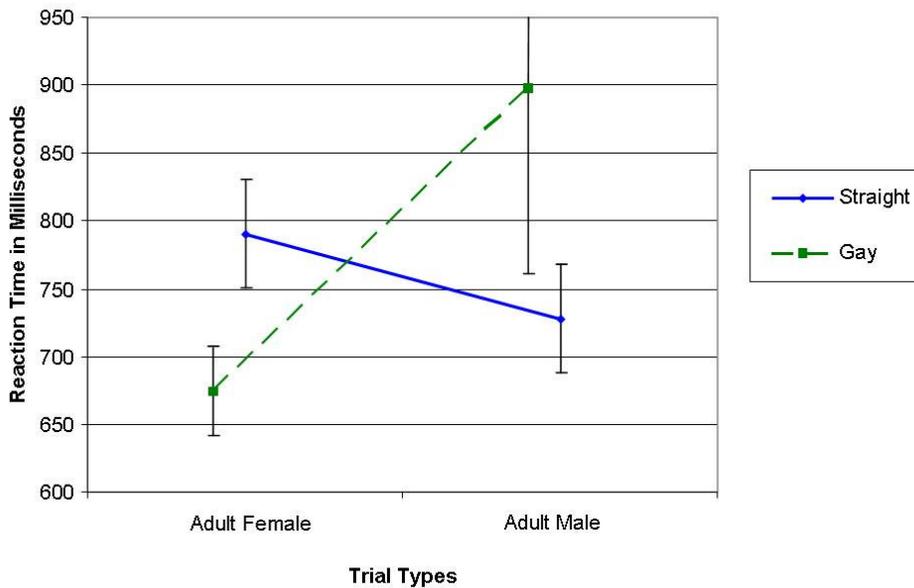
Figure 3.3: Mean reaction times of gay and straight men to adult male and female images using the pictorial modified Stroop task (error bars indicate one standard error).

For adult stimuli, the main effect of gender of stimulus did not produce a significant result; $F(1, 19) = 2.541$, $p = .127$, partial $\eta^2 = .118$. There was a significant interaction of gender of stimuli and sexual orientation; $F(1, 19) = 8.067$, $p = .01$. Pairwise comparisons were carried out in order to investigate the source of the interaction.

76

Graphing the interaction indicates a disordinal interaction between gender of image and sexual orientation of participant. Four *t*-tests (two independent samples and two paired samples) showed a significant difference only between reaction times for adult female images between gay and straight men with straight men taking longer to respond: $t(19) = -2.153$, $p = .044$, two tailed, $r = .443$. The difference between adult female and adult male images was approaching significance among gay men: $t (8) = -2.079$, $p = .071$, two tailed, $r = .059$. The same comparison among straight men yielded a non-significant *t*. However when a further Bonferroni conversion was applied to allow for familywise error none of the results from these *t*-tests were significant (α set at .00625). However at this *alpha* level for the independent samples t-test, at least 90 participants would have been required in each group to have acceptable power (.8) to identify a moderately sized effect. Reaction times to control stimuli (large cats) were consistent across gay and straight participants (mean for gay participants was 671.8ms, *SD* 97.8ms; mean for straight participants was 679.4ms, *SD* 113.9ms) with no significant differences found between them, $t (19) = -0.16$, $p = .874$, $r = .037$, two-tailed.

A Receiver Operating Characteristic (ROC) curve was plotted to assess how well a score based on the difference between mean reaction time to adult female stimuli and mean reaction time to adult male images was capable of predicting self reported sexual orientation. This analysis shows an area under the ROC curve (AUC) of .898 which represents a predictive ability that differs significantly ($p=.002$; $SE=.073$) from .5. As mentioned, difference scores were based on reaction times where outliers beyond 1.5 times the interquartile range were removed. A second (less conservative) score was also calculated where only reaction times that were more than 3 times the interquartile range beyond the 25th and 75th percentiles were removed. If you base the calculation on Stroop scores using this approach to extreme value removal the AUC increases to .926 ($p=.001$; $SE=.061$). An examination of the coordinate points generated to draw the ROC curve indicated that the optimum cut-off point (if one were looking for a point on the Stroop difference scores that sees the best trade off between correct classification of participants into one or other of the sexual orientation groupings) for either Stroop difference score is very close to zero. This suggests that the practice of subtracting an adult male mean from an adult female mean produces a score with a meaningful zero point. Put another way, we would expect the cut-off point to be zero since a score greater than zero indicates a bias for female images compared with male images and

that a score below zero indicates the opposite. Using this cut-off point for the less conservative Stroop difference score (actual value: 0.53 milliseconds) gay participants were correctly identified 83% of the time while straight participants were identified correctly 89% of the time. Other measures of central tendency yielded AUC that were promisingly high but none were as high as the value for the mean using the less conservative method of outlier removal. The results of these ROC analyses are summarised in Table 1.

Table 1: Summary of AUC values for difference scores calculated by different measures of central tendency

Difference score based on:	AUC	*p* value	Standard Error
Mean	.843	.009	.086
Median	.917	.001	.064
Mean (outliers removed based on 1.5x IQR)	.898	.002	.073
Mean (outliers removed based on 3x IQR)	.926	.001	.061
Ipsative *z*-scores (see section 3.2.3.3)	.917	.001	.068

Taken as a whole the results of the Pictorial modified Stroop task indicated that the sexual orientation of participants impacted on their response times to adult stimuli. The task identified group differences in response times and also demonstrated that a difference score based on response times to male and female adult could categorise participants as gay or straight with a high degree of sensitivity and specificity.

3.2.3.3   *Impact of Age on reaction times and the use of ipsative z-scores*

As with the Stroop reaction times, age of participant correlated significantly with reaction times to most of the different image types (except adult males): Adult females, $r = +.511$, $n = 21$, $p = .018$; child females, $r = +.548$, $n = 21$, $p = .01$; child males, $r = +.451$, $n = 21$, $p = .04$; cats, $r = +.466$, $n = 21$, $p = .033$. This pattern of significance is similar for scores calculated using the alternative, less conservative method of outlier removal (though the correlation for cats is no longer significant). When the correlations with age are calculated for reaction times to all trials (pictorial or traditional Stroop), using either outlier removal method the results, are similarly positively and highly

78

correlated for both; conservative, $r = +.561$, $n = 21$, $p = .008$; less conservative, $r = +.552$, $n = 21$, $p = .009$.

Both viewing time and PPG measures of sexual interest sometimes use ipsative z-scores to standardise responses across individuals (Barbaree & Mewhort, 1994; Sachsenmaier & Gress, in press), i.e. to account for individual differences in response times in the AASI and individual differences in percentage erection in PPG. It was decided to investigate whether a similar approach would influence the pattern of pictorial modified Stroop results. For each participant, a z-score was calculated for each of the 5 experimental conditions (adult female, adult male, child female, child male and cat) by subtracting the mean overall reaction time to images from the mean reaction time for each block and dividing by the overall standard deviation for reaction times to images. The block mean, overall mean and overall standard deviation were calculated having removed outliers more extreme than three times the interquartile range beyond the 25th and 75th percentile[18]. After converting each of the experimental trial type means to z-scores in this way, age and values for trial types were no longer significantly correlated. In all cases except the z-score for adult female images correlations were below were ± .2 and had associated p values over .4. The correlation for adult females was the only one that now approached significance, $r = +.406$, $n = 21$, $p = .068$.

As shown in Table 1 calculating a difference score based on these ipsative z-scores, when subjected to an ROC analysis, yielded an AUC of .917 ($p=.001$; $SE=.068$), which indicates that this difference score is only slightly less accurate than the mean with outliers removed (less conservative method) when used to discriminate straight from gay men. Using these z-scores to carry out the 2x2 mixed factorial ANOVA comparing gender of adult stimuli across gay and straight participants instead of the means with outliers removed (conservative method), used in section 3.2.3.2, strengthened the results in line with the hypothesis that there would be no main effect of gender but that there would be an interaction effect. The effect size of the non-significant main effect was further reduced using z-scores; $F (1, 19) = .226$, $p = .640$, partial $\eta^2 = .012$, compared with means (outliers removed by conservative method); $F (1, 19) = 2.541$, $p = .127$,

---

[18] This method of outlier removal was chosen since it was found to yield results that best discriminated participants by sexual orientation. Ina addition, this method involves the retention of more of the original data than the more conservative method calculated using 1.5 times the interquartile range. This approach to calculating the ipsative z-scores is consistent across the entire thesis.

partial η^2 = .118. The interaction effect was increased using z-scores. Both mean values with outliers removed (less conservative method) and median values also decreased the main effect of gender and increased the interaction effect but neither produced as large a decrease or an increase respectively as the z-score method. A comparison of the 5 measures (raw mean is also included) is included in Table 2.

Table 2: impact of different measures on the main effect and interaction of the ANOVA comparing reaction times to gender of adult stimuli across gay and straight participants

Measure	Main effect of gender			Interaction of gender and sexual orientation		
	F	p value	partial η^2	F	p value	partial η^2
Mean (raw)	1.870	.187	.09	6.983	.016	.269
Median	1.395	.252	.068	10.414	.004	.354
Mean (outliers removed based on 1.5x IQR)	2.541	.127	.118	8.067	.01	.298
Mean (outliers removed based on 3x IQR)	1.925	.181	.092	10.932	.004	.365
Ipsative z-scores	.226	.640	.012	16.318	.001	.462

To summarise, an investigation of the impact of age on the pictorial Stroop results indicated that age may confound the results of the task. Therefore a method designed to reduce the impact of individual differences in overall response times was attempted where reaction times to different trial types were converted to ipsative z-scores. This approach had slightly less discriminant ability than means using the less conservative method of outlier removal when classifying participants as gay or straight. However the approach did strengthen the results in line with the hypothesis that there would be no main effect of trial type while there would be a significant interaction of trial type and sexual orientation.

3.2.3.4 *Graphical representations of the pictorial Stroop data: frequency distributions and order of presentation plots.*

The most effective measure of central tendency in discriminating between straight and gay participants is the one that excludes outliers but only very extreme ones (remember that the z-score method is also computed using outlier removal only when those outliers

are 3 times the interquartile range beyond the 25th and 75th percentiles). Given that most reaction time data is positively skewed (Ratcliff, 1993), it is likely that the improvement of the less extreme method of outlier removal over no outlier removal and the more conservative approach stems from the fact that genuine data from the slower end of the response range is being retained in the data. In reaction time studies it is not only the mean that may change based on the experimental intervention but there may also be a change in the shape of the distribution across experimental tasks. It could be the case that a similar mean reaction time across different trial types could mask the fact that a genuine change has taken place in the shape of the distribution. While the previous analyses indicate that there is an observable difference between responses to different trial types, it is still worth exploring, by visual inspection at least, the distributions of reaction times for trials where the image was potentially sexually salient (i.e. images of adult males for gay participants and images of adult females for straight participants) and where the image should not have sexual salience (all other images). In addition to looking graphically at the distributions of the reaction times it was also decided to plot reaction times by order of presentation for each experimental block.

All histograms and order plots are calculated by converting reaction times for individual trials to z-scores[19]. These z-scores are computed using that participant's overall mean for all image trials that has been trimmed of outliers that are over 3 times the interquartile range beyond the 25th and 75th percentile. This cut-off was chosen in order to remove only extreme outliers and therefore to retain as many genuine data points as possible. The standard deviation was also computed based on data that fell within the same range. The outliers that were identified for this purpose were removed from the order graphs, since they would be likely to distort any pattern that existed. The same would not be true for the histograms so such data points were, therefore, retained. Figure Figure 3.4 shows the z-score distribution of reactions to image trials averaged across all participants. In other words it is an average of the histograms that would be produced for all participants. This distribution has a skewness of 1.072 and a kurtosis value of 2.254. The kurtosis value is outside the ±2 range typically considered acceptable A Kolmogorov-Smirnov test of normality indicates that the distribution is

---

[19] While Vincentising (Vincent, 1912) was also an option to create histograms of the averaged participant distributions, the method adopted here, though novel, made more intuitive sense to the author.

not normal, $D(229) = .106$, $p < .001$.
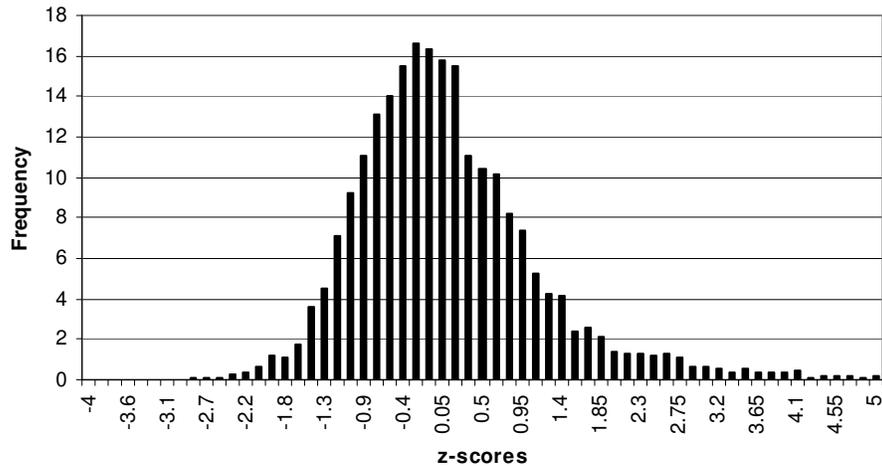
Figure 3.4: Averaged frequency distribution of z-scores calculated from reaction times to all image trials

Frequency distributions were also plotted for five different trial types. These were:

a) Trials where the images presented were of an adults (age appropriate) and consistent with the participant's sexual orientation (male images for gay participants and female images for straight participants).

b) Trials that were age appropriate but inconsistent with the participants sexual orientation.

c) Trials that were age inappropriate (i.e. of children) but consistent with the participants sexual orientation.

d) Trials that were age inappropriate and were orientation inconsistent

e) Control trials (i.e. images of large cats)

As with the values plotted in Figure 3.4, reaction times converted to z-scores were calculated for all five types and frequency distributions were plotted. These distributions are overlaid in Figure 3.5. It is clear from the graph that the only *z*-score distribution that differed noticeably from the rest was the one associated with trials where images were of adults and were consistent with the participants' sexual preference. All other distributions mirrored each very closely, while the age appropriate, orientation consistent distribution was less positively skewed[20], had less kurtosis[21] and

---

[20] The age appropriate, orientation consistent distribution had a skewness of .854 and an aggregate distribution based on the other four trial types had a skewness of 1.133.

possibly had a greater amount of data points in the tau (τ) region of the distribution (if one argues that the distribution approximates an ex-Gaussian distribution).
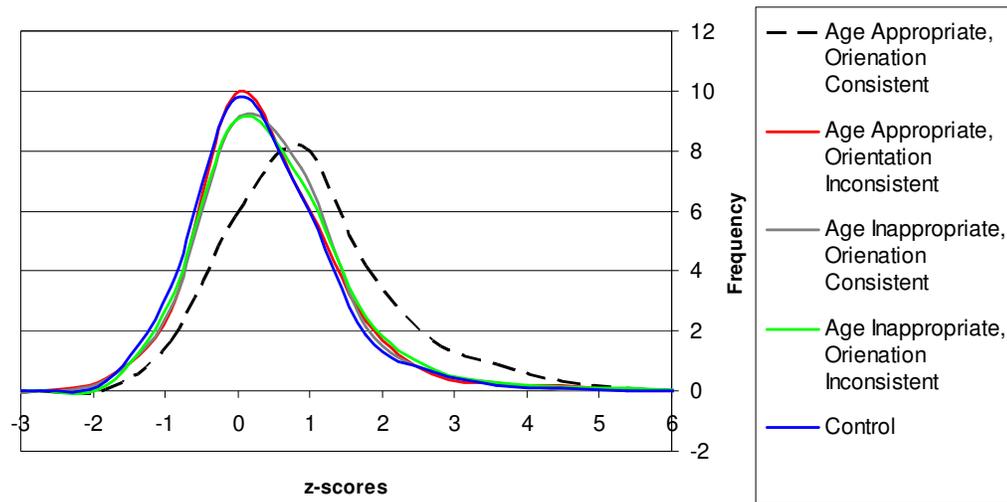
Figure 3.5: Averaged frequency distribution of z-scores calculated from reaction times to image trials based on whether images are consistent with participants' sexual orientations, whether they are of children or adults, or whether they are control images.

Figure 3.6 represents the average reaction times by order to the five different stimulus types mentioned above. Reaction times were again converted to *z*-scores, but were graphed by order of presentation. Only reaction times that fell between ±3 times the interquartile range beyond the 25th and 75th percentile were included in the graph. To simplify the graph and to aid interpretation *z*-scores were averaged across six trials at a time (i.e. a mean was calculated for the first six trials presented, the second six and so on). No clearly interpretable pattern of order emerged from the data, though it seemed that response times for the age appropriate, orientation consistent trials (i.e. the trials where a sexual content induced delay was expected) was reducing towards the end of the 48 trials. In addition it was apparent that there was not a clear accumulative effect, where response times would build over the course of the task for the age appropriate, orientation consistent trials.

---

[21] The kurtosis of the age appropriate, orientation consistent distribution was 1.152, while the kurtosis for the aggregate distribution containing the rest of the data was 3.183.
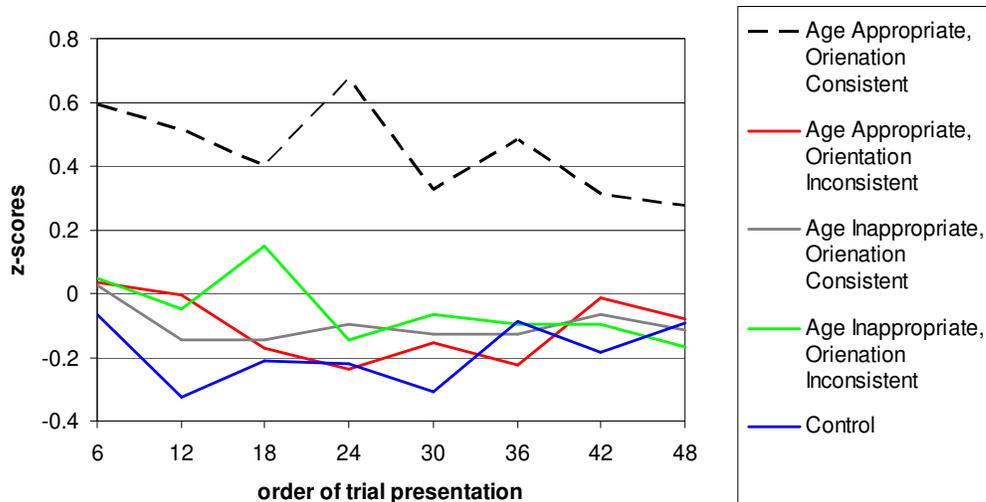
Figure 3.6: Plot depicting average reaction times by order of trial for five trial types. Individual reaction times were converted to *z*-scores and then averaged across participants. Means for six trials at a time are presented.

Taken together, the depiction of the data in this section suggested that the overall distribution of the reaction time data in the pictorial modified Stroop task followed an ex-Gaussian distribution. When the distributions for different trial types were plotted, all trials except those that involved age appropriate, orientation consistent images had similar distributions. Age appropriate, orientation consistent stimuli seemed to produce a distribution that differed not just in mean but possibly also in shape to the other trial types. When trials were plotted by order of presentation, there was no clear evidence of an accumulative effect of trial presentation on response time.

3.2.3.5 *Errors and Outliers in the pictorial Stroop task*

A 2x5 mixed factorial ANOVA was conducted to compare the number of errors committed by gay and straight participants across each of the five image types. Two participants were removed from the analysis as they had both responded incorrectly to almost all yellow trials. There was no interaction of orientation and image type and there was no main effect of image type. A second 2x5 mixed factorial ANOVA was conducted, again with sexual orientation as the between-subjects factor, but with number of outliers (calculated using 1.5 times the interquartile range) across image types as the within-groups factor. While, there was no interaction, the main effect of image type was approaching significance*; $F_{(4, 76)} = 2.207$, $p = .076$, partial $\eta^2 = .104$).

Overall, adult female images produced the most outliers (Mean = 3.05, *SD* = 1.88) followed by child female images (Mean = 2.67, *SD* = 2.35). Third came cat images (Mean = 1.9, *SD* = 2.05), with child male (Mean = 1.86, *SD* = 1.35) and adult male (Mean = 1.81, *SD* = 1.6) in fourth and fifth position respectively.

3.2.3.6 The influence of order on the observed data

The pictorial modified Stroop task used here incorporated a blocked design to maximise group differences when reacting to potentially sexually salient stimuli. However, the disadvantage of this technique is that it also may introduce order effects. This does not cause a problem for between-groups comparisons as any order affects should be counterbalanced by the random presentation of blocks across participants. However if one takes the next step and analyses data at the individual level the resultant confounding variance may reduce the tasks sensitivity in identifying sexual orientation. Therefore, statistical consideration was given to the effect of block presentation order on results. Mean scores with outliers removed (1.5 times the interquartile range) were calculated for blocks based on presentation order across each participant. On inspection of boxplots one block mean for one participant was removed from analysis as it was considered an extreme outlier (6.44 times the interquartile range). Less extreme outliers (less than 3 times the interquartile range) were not removed as their removal would involve too great a loss of data. This was also deemed acceptable since these values were not single observations but block means (with outliers already removed) for each participant. A repeated measures ANOVA was carried out on these presentation order block means, which found a significant effect of block order; $F(4,76) = 4.826$, *p*= .002, partial η^2 = .203. Pairwise comparisons show significant differences between block 1 (Mean = 777.2ms, *SD* = 189.66ms) and each of blocks 3 (Mean = 706.07ms, *SD* = 120.3ms), 4 (Mean = 687.1ms, *SD* = 103.27ms) and 5 (Mean = 659.41ms, *SD* = 94.08ms) and also between blocks 2 (Mean = 720.53ms, *SD* = 114.69ms) and 5 and between blocks 3 and 5. All differences are in a direction that shows a decrease in reaction times across successive blocks. However when a Bonferroni correction is applied (α = .005) only one of these differences continues to be significant, that between block 1 and 5; $t(19) = 3.785$, *p*= .001, two-tailed. When between-group differences between gay and straight participants were investigated no significant interaction was found between sexual orientation and block order.

3.2.3.7 Initial analysis of the Implicit Association Tests

A 2x2 mixed factorial ANOVA was carried out on the results of the age-sex IAT with responses to child-sex concept pairings versus adult-sex pairings as the within subjects factor and sexual orientation as the between subjects factor. No significant main effects or interactions were found. A 2x2 mixed factorial ANOVA was carried out on the results of the gender-sex IAT with responses to the pairing of male names with sex words versus female names with sex words as the within groups factor and sexual orientation as the between subjects factor. No significant main effect was found. However, there was a significant interaction between gender of name-sex pairing with sexual orientation: F (1, 19) = 5.876, p = .025, partial η^2 = .236. Pairwise comparisons were carried out in order to find the source of the interaction. Four t-tests (two independent samples and two paired samples) showed a significant difference only between reaction times for female-sex word pairings between gay and straight men with gay men taking longer to respond: t (19) = 2.341, p = .03, two tailed. However when a Bonferroni conversion was applied to allow for familywise error no results were significant. Though non-significant, results were in the direction expected with gay men taking longer to respond to female-sex pairings than they did to male-sex pairings which was the inverse of the pattern for straight men (see Figure 3.7). At least 78 participants would have been required in each group to for the independent t-tests to have acceptable power (.8) to identify a moderate effect given the Bonferroni conversion applied.
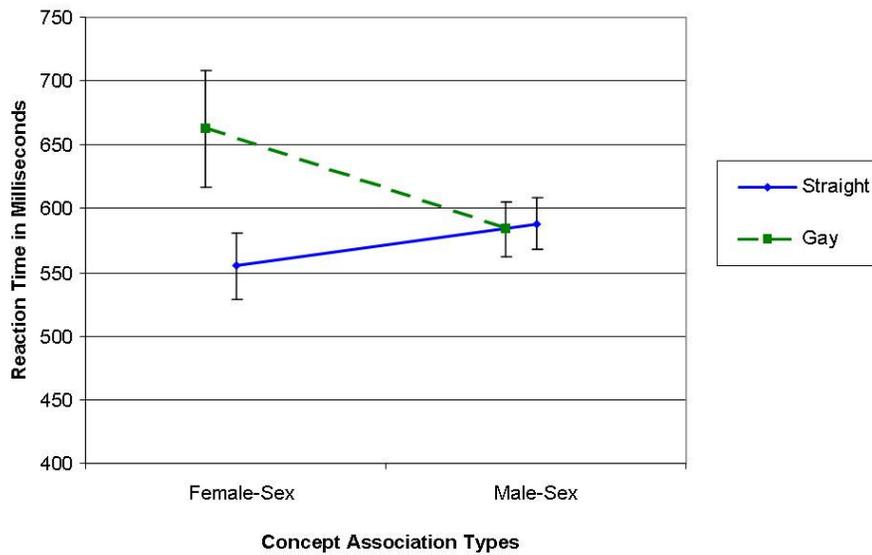
Figure 3.7: Mean reaction times to female-sex and male-sex associations for gay and straight participants in Gender-Sex IAT (error bars indicate one standard error).

As with the Stroop task several different methods were available for calculating a difference score between the two trial types. In total, for the Gender-Sex IAT, 13 different difference scores were calculated. These were based on different iterations of three approaches:

1) Greenwald et al. (2003) recommend using response times from each of the blocks where concepts are paired (i.e. blocks 3, 4, 6 and 7). In addition, they recommend removing responses over 10 seconds and removing participants who have more than 10% of responses that are quicker than 300ms. The mean reaction[22] times for blocks 3 and 6 are subtracted from each other and divided by the pooled standard deviation of both blocks. Blocks 4 and 7 are treated the same way. The mean of the two difference scores is then computed to give the *D* score. Three versions of Greenwald et al.'s (2003) *D* algorithm were calculated. The first calculated a score following their algorithm but did not include incorrect responses or any method of error penalty. This method is not suggested by Greenwald et al. (2003). The second method followed their algorithm exactly, including replacing each

---

[22] Greenwald et al (2003) suggest two approximately equivalent versions of the algorithm where the mean includes the replacement of errors with 'error penalties' or the inclusion of the total time it took for the participant to respond correctly, regardless of how many errors that included.

87

error with the mean for that block plus 600ms. The third approach, which Greenwald et al. (2003) claim to be approximately equivalent involved including the total time to correct response for trials where the participant initially responded incorrectly.

2) Two further versions of each of the three *D* scores mentioned above were calculated that removed outliers based on the interquartile range, not on cut-off points as suggested by Greenwald et al. (2003). There were two versions of each score where one had outliers removed if they were 1.5 times the interquartile range beyond the 25th or 75th percentiles and the second had outlier removed if they were 3 times the interquartile range beyond the percentiles. In all other aspects the methods were identical to the three *D* algorithms.

3) The third approach was to calculate a simple difference score between either the blocks of interest (i.e. the difference between the mean reaction time to blocks 3 and 4 compared with blocks 6 and 7) or to calculate the score only based on trials where the category of male or female was paired with sex (i.e. leaving out any furniture associations). There were two versions of each of these approaches, one based on conservative outlier removal, the second based on the less conservative method, yielding a total of four scores in this approach.

An ROC curve was plotted using all these difference scores to determine how well they discriminated between participants based on sexual orientation. All three *D* measure scores performed poorest, all yielding AUCs under .7. The remaining measures were all quite similar with two scores yielding the same highest value. The first of these was the alternative method to Greenwald et al's (2003) *D* where the time it took for participants to record a correct response was retained, but where total reaction times more extreme than 3 times the interquartile range were removed. The second method that discriminated best was the difference score that simply subtracted mean reaction times to female-sex pairings from mean reaction times to male-sex pairings (with outliers more extreme than 1.5 times the interquartile range removed). The AUC for both methods was .815 (*p*=.016, *SE*=.095/.096). Since the second of these best performing difference scores was expressed in milliseconds the coordinates of the ROC curve were inspected. As with the Stroop difference score, that inspection revealed that the

88

optimum cut-off point, if one was looking for the best trade-off between sensitivity and specificity, is again very close to zero. At this score (actual value: 9.7 milliseconds) gay participants were correctly identified 75% of the time while straight participants were identified correctly 88% of the time. A value close to zero, like the Stroop difference score, would be predicted given that theoretically a score greater than zero indicates a facilitatory effect when the concepts of sex and female are paired and that a score of less than zero indicates a facilitatory effect when male and sex are paired. The AUC values for all difference scores calculated are reproduced in Table 3.

Table 3: Summary of AUC values (demonstrating discrimination between straight and gay participants) using 13 different methods of computing a difference score for the Gender-Sex IAT.

Method of difference score calculation	AUC	Standard Error	*p* value
D with no errors	.667	.122	.201
D errors replaced with block mean +600ms	.694	.124	.136
D including time-to-correct responses	.657	.127	.227
Alt. version, with no errors (IQR x 1.5)	.787	.100	.028
Alt. version, with no errors (IQR x 3)	.787	.105	.028
Alt. version, errors replaced with block mean +600ms (IQR x 1.5)	.796	.098	.023
Alt. version, errors replaced with block mean +600ms (IQR x 3)	.796	.101	.023
Alt. version, including time-to-correct responses (IQR x 1.5)	.778	.101	.033
Alt. version, including time-to-correct responses (IQR x 3)	.815	.095	.016
Mean RT from blocks 6&7 minus mean RT from blocks 3&4 (IQR x 1.5)	.759	.107	.047
Mean RT from blocks 6&7 minus mean RT from blocks 3&4 (IQR x 3)	.750	.111	.055
Mean RT for male-sex pairs minus mean RT for female sex pairs (IQR x 1.5)	.815	.096	.016
Mean RT for male-sex pairs minus mean RT for female sex pairs (IQR x 3)	.787	.103	.028

In summary, analysis of the IAT results found that there was no main effect of trial type across participants for the Age-Sex IAT. No interaction was found either. The Gender-

Sex IAT again identified no main effect of trial type but a significant interaction was found. Further investigation of this interaction suggested that gay men took longer than straight men in allocating sex words or female names to the correct category when those concepts were paired. Thirteen methods were used to calculate a difference score between male-sex and female-sex associations. Most showed a trade-off between sensitivity and specificity that was significantly better than chance in discriminating between gay and straight participants. The difference scores that followed Greenwald et al.'s (2003) *D* algorithm preformed poorly relative to the other methods.

For the sake of simplicity only three difference scores were included in further analyses of the IAT results which compared the IAT and pictorial Stroop results. These were the two highest performing difference scores, in terms of discrimination of sexual orientation and also the best performing of the three *D* scores: the one that included a penalty of the block mean plus 600ms for incorrect responses (this was included for the sake of completeness). For brevity, the three methods will be referred to as the '*D* score', the 'alternative score' and the 'mean difference score'.

3.2.3.8   Influence of age on IAT

Age of participant did not correlate significantly with any of the three difference scores calculated from the IAT results, either for the Age-Sex or the Gender-Sex IATs. There were no significant correlations either between age and mean reaction times across all of the experimental trial blocks in both IATs. The IAT, therefore seems to be more resilient to the effects of age compared with the Stroop task.

3.2.3.9   Comparing the pictorial modified Stroop task and the IAT

Pearson correlations were calculated between three pictorial modified Stroop difference scores (conservative method of outlier removal, less conservative method and the difference score based on ipsative z-scores) and the three IAT difference scores. The highest correlation was between the pictorial Stroop z-score method and the IAT *D* score, $r = +.559$, $n = 21$, $p = .008$, two-tailed. The pictorial Stroop z-score method correlated significantly with all three IAT difference scores and the IAT *D* score correlated with all three pictorial Stroop difference scores. The remaining results were

90

not significant but some were approaching significance and all were correlations over .3 and in the direction expected. All correlations are reproduced in Table 4.

Table 4: Correlations between different methods of difference score calculations for Stroop task and the Gender-Sex IAT

		Stroop z-score	Stroop mean (conservative outlier removal)	Stroop mean (less conservative outlier removal)
Mean RT for male-sex pairs minus mean RT for female sex pairs (IQR x 1.5)	*r*-value	.519(*)	.329	.386
	p-value	.016	.146	.084
Alt. version, including time-to-correct responses (IQR x 3)	*r*-value	.547(*)	.374	.427
	p-value	.01	.095	.054
D errors replaced with block mean +600ms	*r*-value	.559(**)	.493(*)	.534(*)
	p-value	.008	.023	.013

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

The correlations indicate that there is a relationship between results of the Gender-Sex IAT and the responses to adult male and female stimuli in the pictorial Stroop task. In addition, scores on both show group differences between gay and straight participants. Both can also categorise participants with a sensitivity and specificity that is between good and excellent[23] according to ROC analysis. It may be that both tasks, while overlapping to some degree, are tapping into separate components of sexual interest. If this was the case both together might better predict sexual orientation. To explore this question, two forced entry binary logistic regressions were carried out to assess whether a combination of both the IAT and pictorial Stroop scores yielded a better categorisation of participants as either gay or straight than either did individually. The difference score based on the less conservative Stroop means was first used as it was

---

[23] If using Tape's (1999) heuristic for interpreting AUCs: over .9 being excellent, .8-.9 good, .7-.8 fair, .6-.7 poor and AUCs of .5-.6 being described as fails.

shown by the ROC curve analysis to have a better trade off between hit and false alarm rate than the difference score based on the more conservative means and z-scores. For the same reason the difference score used for the Gender-Sex IAT was the one calculated from the means for male-sex and female-sex associations (though the IAT alternative score could also have been used here). In the first regression analysis the IAT difference score served as the predictor variable in the first block and the Stroop difference score was the predictor variable in the second block. This order of inclusion in the analysis was reversed in the second regression. The analysis was conducted in both directions to see how each score performed on their own and then when the other was included. Taken together both scores yielded an overall classification rate of 90.5% with 88.9% (8/9) of gay and 91.7% (11/12) of straight participants being correctly classified. Individually the Stroop score outperformed the IAT score in terms of categorisation (85.7% compared with 71.4%). Both scores explained large amounts of the variance with the Stroop model ($R^2 = .68$) being higher then the IAT ($R^2 = .41$). The combination of both predictor variables yielded a 10% increase in total variance explained ($R^2 = .78$). The Stroop score accounted for 39% of unique variance while the IAT only accounted for 10%. Twenty nine percent of the variance in categorisation was attributable to overlapping variance among the IAT and Stroop. The addition of the IAT score to the regression model including the Stroop score increased the classification rate by 6% which corresponded to the correct classification of one additional gay participant. From a practical perspective, an increase of 6% of correct classification is noteworthy given the implications for large samples. This increase also suggests the utilisation of both IAT and Stroop in combination is a fruitful research strategy. No other method of difference score calculation yielded a higher percentage of correct classification, though using the Stroop difference score based on z-scores and the IAT alternative score also yielded an overall correct classification rate of 90.5%.

To see if there was any relationship between the Age-Sex IAT and the pictorial modified Stroop task, correlations between the two were examined. Two difference scores were calculated between the response times to child images and adult images for the pictorial Stroop task. Both approaches used the ipsative z-scores since they had previously been shown to have the highest correlation with the Gender-Sex IAT. The first approach calculated the difference between the average reaction time to adult images and the average reaction time to child images. The second approach was to take

each individual's slowest mean reaction time to adults (i.e. either RTs to males or females, whichever was slower) and subtract their slowest mean reaction time to children from it. Neither of these measures correlated with any of the Age-Sex IAT difference scores.

To summarise, a comparison of the pictorial modified Stroop task and the two IATs found that the pictorial Stroop and the Gender-Sex IAT were significantly correlated but that both also explained a limited amount of unique variance. The results of the Age-Sex IAT on the other hand were not correlated with any of the pictorial Stroop results.

3.2.4 Discussion

Results of the traditional Stroop task indicated that participants took significantly longer to respond to incongruent trials than to congruent and control trials. This was in line with the hypothesis. Also in line with hypotheses was the finding that age and response time across all trial types were positively and significantly correlated and that sexual orientation has no impact on response times. The hypothesis that congruent trials would be significantly faster than control trials was not supported, though the difference between them was in the expected direction.

The data from the pictorial Stroop task were initially examined using a conservative method of outlier removal. A mixed factorial ANOVA indicated an interaction of orientation of participant, gender of stimulus and age of stimulus (adult or child). A much larger sample would be needed to properly tease out the source of this interaction but results did indicate that adult stimuli did produce significantly longer reaction times than child stimuli. In addition further testing indicated that participants did, on average, have higher reaction times to sexual orientation consistent adult images. Though non significant, this pattern was in-line with the hypothesis. Following this support for the main research question (i.e. whether the pictorial Stroop task would be in some way tapping into sexual interest) the data were explored to determine how predictive they were of sexual interest and how well they discriminated gay participants from straight ones. Several different measures of central tendency were compared. A mean based on a less conservative method of outlier removal was calculated. Also calculated was a value based on z-scores where each participant had their mean responses to each of the

five trial types represented in terms of number of standard deviations away from their overall mean reaction times to any pictorial stimuli. This method was chosen for two reasons. First, results indicated that Stroop response times and age were correlated, so it was hypothesised that this method would reduce some of the influence of age on the responses. The correlation between Stroop performance and age is likely to be due to a general effect of age on cognitive speed. Typically there is a large individual variability in the impact of age on cognitive speed so it is likely that using an individual's own mean and standard deviation to attempt to control for variability in cognitive speed would be a better solution to the issue than using age as a control (or covariate). Correlations between age and pictorial Stroop performance were reduced when this method was used. The second reason for using this method is that a similar "ipsative z-score" approach is used in both PPG measurement and viewing time measurement. The approach is problematic however. When used with individuals with very large or very small standard deviations, the size of an effect may be distorted. Caution must be exercised in the interpretation of interpretation of the results. In relation to ipsative z-scores for PPG, Barbaree and Mewhort (1994) demonstrate that using z-score transformations can compromise estimates of type one error and influence power when individuals with high or low variability of arousal are in a sample. They recommend the alternative approach used by many PPG practitioners of expressing arousal as a percentage of full erection. This alternative is not without its drawbacks, either. Similarly, Sachsenmaier and Gress (in press) raise similar concerns about the use of ipsative z-scores with viewing time measures, mainly regarding the potential of large or small individual standard deviations to minimise or exaggerate an effect. However, the issues raised by those authors are primarily concerned with the risk of false positives in the clinical application of PPG and viewing time. In establishing the construct validity of the pictorial modified Stroop task a z-score method may prove useful. In addition to those already mentioned, two further measures of central tendency were calculated: means with no outlier removal and medians.

Regardless of the measure of central tendency used, the pictorial Stroop task demonstrated a good ability to discriminate between gay and straight participants. The difference score based on means with outliers removed (less conservative method) had the best discriminant ability with an AUC of .926. Re-running the ANOVA that looked at reaction times to adult stimuli for both gay and straight participants using the

different measures of central tendency (and the ipsative z-score method) yielded no change in the pattern of results but brought the results more in line with the original hypothesis: that there would be no main effect of gender but that there would be a significant interaction. The ipsative z-score method gave the smallest main effect with the largest interaction effect. The finding that different measures of central tendency give better classification of participants in terms of sexual orientation suggests that the treatment of extreme data is important in determining how best to interpret the data from the pictorial Stroop task. Since extreme values in reaction time data typically occur on the right hand side of the distribution (i.e. slow responses), the fact that the greatest accuracy in discriminating occurred when only extreme outliers were removed (i.e. those calculated based on 3 times the interquartile range) indicates that part of the effect observed was occurring in the tau (τ) region of the distribution. A graphical comparison (Figure 3.5) of the distributions of orientation-consistent and orientation-inconsistent responses suggests that reaction times to sexually salient stimuli forms a distribution that differs not just in the mean but also in the shape to the distribution of reaction times for stimuli that are not sexually salient. It is striking in Figure 3.5 that all non-sexually salient stimuli distributions map onto each other very closely. Chapter 2 mentions the suggestion by Balota and Spieler (1999) that the Gaussian component of the reaction time distribution may represent automatic processes and that the exponential component may represent more analytical processes. This ties in well with the current author's suggestion that a pictorial Stroop effect for sexual stimuli is as a result of different stages of the arousal process. The pre-attentive and attentive stages stage hypothesised by Spiering et al (2004) in their model of arousal, and introduced in Chapter 1, may correspond to fast effects and slow/rumination effects respectively in the pictorial Stroop task which may then, in turn, manifest as Gaussian and exponential effects on a frequency distribution.

The pictorial Stroop task produced an unexpected main effect of gender for the child stimuli when an ANOVA was carried out looking at response to child images across sexual orientations. Images of male children produced significantly slower reaction times than did images of female children. This pattern was true for both straight and gay participants. It is not clear why this might have been the case. It is possible that images of male children held more sexual salience for some of the control participants. However, it seems more likely that this difference was due to chance or that the images

of children used were salient (or attentionally demanding) for another reason. It has already been stated in Chapter 2 that some of the male images from the NRP seem to be less realistic than the female images. While the NRP set has undergone independent rating in terms of classifying the images by developmental stage (Laws & Gress, 2004), they have not been validated in terms of attractiveness and how realistic the images are. This criticism could also be levelled at the novel images developed for this study. Chapter 2 outlines why it would have been problematic to address this issue in a study using images of children but a later study (that uses only adult images) in this dissertation will seek to address the matter. While there was a clear main effect of age of stimulus in the pictorial Stroop task, some participants had higher response times to images of children than they did to adult images. Again this may have been due to a lack of precision in the task but it may also have been that child images held more salience for that participant than did adult images. This may represent a percentage of the population that finds children sexually salient. On the other hand it may indicate that there were issues with volunteer bias, as was suggested in Chapter 2, with participants who had some sexual interest towards children self-selecting for the study. Another alternative might be that some participants, aware that sexual interest in children is being investigated, but unaware that it was speed of response and not amount of errors that was the main variable of interest, may have slowed down their responses on trial types that they were most concerned with being seen as having a sexual interest towards. It must also not be discounted that for these participants, images of children were salient to them for reasons that had nothing to do with sexual interest. Word Stroop studies have shown that stimuli that are negative or threatening can in result in a content induced delay (McKenna & Sharma, 1995) as can words that are salient for many other reasons. It is difficult to hypothesise a clear reason why images of children might be salient.

Given the presence of an order effect in the data it seems likely that the removal of that order effect would have produced data with an even better ability to discriminate the sexual interest of participants. The presence of this effect may have impacted on the ability to detect correlations between IAT results and Stroop results and may also have impacted on the ability to see whether Stroop results were on a continuum that might allow us to measure strength of interest. Two methods may reduce the impact of order. The first, quite obviously, is to do away with the blocked design, and completely

randomise all stimuli. This would insure that any effects of order would be randomly distributed across all trial types. The cost of this design may be to miss phenomena that are driven more by rumination than by instantaneous attentional capture as argued in Chapter 2. The second method is an attempt to strike a balance between the random and the blocked design: in a 'clustered' design matching stimuli would be presented in small blocks or clusters with several clusters of each trial type being spread across the task. One could even adopt a pseudorandom presentation that would ensure that a cluster of each trial type must be presented in random order before moving on to a second cluster of any trial type and so on. This design could go some way to removing the impact of order and therefore improving classification at the individual level. An inspection of Figure 3.6 suggests that age-appropriate orientation-consistent stimuli typically yielded consistently longer response times across the entire duration of the block which would indicate that splitting the block into several clusters should not eradicate the effect.

The Age-Sex IAT yielded no main effect of age category-sex pairing (i.e. child-sex versus adult sex) and no interaction between age category-sex pairing and orientation of participants. While it was hypothesised that there would be no interaction between sexual orientation and age category-sex pairing it had been hypothesised that there would be a main effect of age category. As hypothesised there was no main effect of gender of category-sex pairing for the Gender-Sex IAT. Also in line with the hypothesis, there was an interaction between gender of name in the name-sex pairings and sexual orientation. As with the Stroop results, there would need to be more participants in the study to be able to fully explore the interaction found. However, as predicted, gay men typically responded quicker to male names paired with sex words than to female names paired with sex, whereas the opposite pattern was found with straight men. With the exception of the three difference scores based on Greenwald et al.'s (2003) *D* algorithm, all calculated difference scores were able to discriminate between gay and straight participants, using ROC analysis, at a level that was significantly better than chance. This supported the conclusion that, as with the Stroop, the Gender-Sex IAT is tapping into cognitive processes related to sexual interest.

The mixed performance of Greenwald et al.'s (2003) *D* algorithm is interesting, given that much testing has gone into its development. In the current study the different versions of algorithm performed worst when it came to classifying people by sexual

orientation. On the other hand *D* outperformed the other measures when it came to correlations with the pictorial Stroop task. While this discrepancy is hard to understand, there are several reasons why the method may not have performed as well as it might have. The algorithm was developed mainly for IATs that used *positive* and *negative* or similar as concept categories. The algorithm may not be as suited to tasks that look at concepts that are not directly related to valence. Perhaps more crucially the *D* algorithm is designed for an IAT where there are opposite concept categories whereas, in this case there was a concept category (*sex*) and a control category (*furniture*). The presence of a control category means that the *D* score now contains data from half the responses where there isn't hypothesised to be either a strong or a weak association. This would potentially dilute the IAT effect by including what is, essentially, random data with data that may contain an effect. This was illustrated by the fact that methods of scoring the IAT that included only –sex associations outperformed identical measures that included –sex and –furniture associations.

Taken separately, both the pictorial Stroop task and the Gender-sex IAT showed considerable potential in being able to access sexual interest or sexual interest-related cognitions. There was a large amount of convergence between the results of the pictorial Stroop task and those of the Gender-Sex IAT, indicating that both are measuring related cognitive processes or processes that are correlated with one another (at least in the case of our sample). If, as hypothesised in Chapter 1, the IAT is measuring sexual associations or schema and the pictorial Stroop task is measuring sexual arousal-related attention it is likely that both would overlap greatly, especially in a sample of participants without problematic or disordered sexual interests. Put another way, the fact that the two measures agree does not prove that they measure the same thing. Indeed, the fact that both combined are better able to correctly categorise participants as either straight or gay may indicate that they are not measuring the same underlying construct. It should be stressed, however, that the inclusion of the second measure only slightly improved categorisation.

As mentioned the Age-Sex IAT showed no significant differences between trials where child and sex were associated and trials where adult and sex were associated. Neither was there a relationship between the pictorial Stroop and this measure. There are two possible reasons for this. First, the Age-Sex IAT as constructed may not be capable of

eliciting an association of age-category and sex. Second, non-offending participants may not typically hold a strong association between adulthood and sex and may therefore not be likely to respond much faster to adult-sex associations than they would child-sex ones.

Regarding the first reason, that because of design difficulties the Age-Sex IAT was incapable of detecting an association that does in fact exist, there are several reasons to suspect that this was not the case. The design of the Gender-Sex IAT was identical to the Age-Sex IAT, except for the words used in the *adult* and *child* categories, and yet it was able to demonstrate a sexual association quite well. It could be argued, therefore, that the words chosen as stimuli for the *adult* and *child* categories did not adequately represent those categories. However, previous IATs (De Houwer, 2001) have found that the expected associations were found between concept pairs (e.g. British-good, foreign-bad tested with British participants) even when the stimulus words were at odds with that association (i.e. infamous Britons and popular foreigners). De Houwer argues that the IAT effect reflects attitudes to the target concepts more-so than the individual stimuli that form those concepts (2001). Given that the procedure used in this study differs significantly from that typically adopted for an IAT, it is possible that individual stimuli are more important in this instance. Finally, however, the majority of the *adult* and *child* stimulus words were identical to ones successfully used by Mihailides et al. (2004) or Grey et al. (2005); therefore these stimuli should have been able to elicit a similar effect to those found by both those studies.

The alternative possibility, as to why there was a null effect of Age-Sex IAT, is that control participants simply do not have a strong association of sex and adults. It may be that the concept of adult, or indeed childhood, does not get incorporated into an individual's sexual schema unless it is problematic. The results of Grey et al.'s (2005) study found a smaller average *D* score in the hypothesised direction for incarcerated non-paedophilic offenders (control) than they did for paedophilic offenders. What this means is that while both groups performed as expected, with speed of responses being facilitated towards offence-compatible word pairings that facilitation was larger (almost 3 times) for the paedophilic offenders than it was for rapists. It is possible that if there was a small IAT effect in the current study that it could have been obscured by the fact that all participants completed the condition where child and sex were paired first and

where adult and sex were paired second. This means that any difficulties the participants had in adjusting to the new arrangement of words on the screen could negate somewhat any impact of their ease of adult-sex associations. This design was deliberately chosen to maximise the possibility to identify child-sex associations and reduce the necessity to include IAT block order as an additional independent variable when exploring findings (this was important, given that recruitment of participants was difficult).

While the null finding of an Age-Sex IAT effect does run contrary to the initial hypothesis, it does raise an interesting argument. As implied in the introduction it was going to be difficult with the current study to support or discount the thesis that IATs measure schema and that the pictorial Stroop measures arousal-related attention (essentially sexual interest) since in the case of non-offending controls the net result of both sexual schema and sexual interest should be the same: sexual orientation. Therefore if both the Gender-Sex IAT and the pictorial Stroop correlated with each other and predicted sexual orientation (which they do) this could mean that they are both measures of sexual interest, both measures of schema or measuring a different component of orientation entirely. However, if there was nothing methodologically wrong with the Age-Sex IAT, then the fact that there was a null finding and that there was no relationship between this and the child-adult component of the Stroop task (which did demonstrate lower reaction times to child images) would suggest that both instruments are measuring different things and that the participants in the study simply don't have a strong adult-sex association. Clearly this is a very tenuous argument but one which is interesting nonetheless.

3.3   Chapter conclusion

Study 1 indicated that there were no significant differences between image types in terms of the ease by which the target colour in them could be identified when the content was not a factor. In addition, any differences caused by the colours themselves (certain colours being more easily identifiable) were evenly spread over trials. In study 2, the traditional Stroop performed as hypothesised and indicated no significant difference in cognitive performance between gay and straight participants. Results of the analysis on the pictorial modified Stroop task showed an interaction of gender of

adult stimuli and sexual orientation in line with the hypothesis that gay and straight men would have different reaction times to adult male and female images. Though the directions of the group differences were as predicted the analysis lacked sufficient statistical power, due to a small sample size, to reach definitive conclusions about the full nature of the interaction. Results supported the prediction that there would be no significant differences in reactions to child images. Examination of the differences between reaction times to male and female adult images suggests that the task has the potential to discriminate between gay and straight participants at an individual level. As hypothesized and in line with the pictorial Stroop findings the Age-Sex IAT found no significant differences between gay and straight participants in their reaction times to child-sex associations versus adult-sex associations. Contrary to what was hypothesised participants overall demonstrated no Age-Sex IAT effect, indicating that the task was unable to elicit the intended associations or that participants simply didn't have a strong adult-sex schema or association. Results of the Gender-Sex IAT were in line with the stated hypothesis that gay and straight men would perform differently on the task. As with the pictorial Stroop task a significant interaction was found between gender of name-sex association and orientation of participant. When this interaction was investigated further, differences were in the directions predicted, but again the analysis suffered from a lack of power. The IAT did not appear to be as good as the pictorial Stroop task at predicting orientation based on comparisons of the area under the ROC curve. Results of the pictorial Stroop task and the Gender-Sex IAT were in the direction expected with reaction times to orientation-consistent stimuli in the Stroop task being longer (demonstrating an attentional bias) than orientation-inconsistent stimuli. The IAT on the other hand showed a facilitative effect when orientation consistent stimuli were associated with sex compared to when the association was between orientation-inconsistent-stimuli. Results suggested that, within the modified Stroop task, reaction times to child male stimuli were significantly longer than child female stimuli across participants. The reason for this is unclear. Results supported the conclusion that these two implicit tasks have the potential to tap into sexual interest and associations. Difference scores calculated from the results of the pictorial Stroop task and the Gender-Sex IAT correlated highly and significantly with each other. Logistic regression analysis did indicate that the inclusion of both measures in the model did marginally improve its ability to correctly categorise participants as either gay or straight. The findings justify the next step of the process, which is to use the techniques within an

offending population. The end goal of this vein of research is to examine the utility of these measures as potential new tools to add to the arsenal of those involved in the assessment of people who sexually abuse. To that end, group differences are not sufficient to establish utility. One way in which the Stroop may be improved is to reduce the potential for order effects by presenting trials not in blocks but in smaller clusters.

Chapter 4: The use of implicit measures with sexual offenders

In this chapter, two studies explore the utility of implicit tasks in the assessment of sexual interest among sexual offenders and non-offending control participants. Study 3 uses the same IATs that were used in study 2 in the previous chapter and a slightly altered pictorial modified Stroop task. The results of sexual offenders against children and non-offending control participants are compared. Study 4 uses a different version of the pictorial modified Stroop task and compares the results of that task with the penile plethysmography results of a sample of participants consisting of incarcerated sexual offenders against children and sexual offenders against adults.

4.1   Study 3: Testing the utility of the pictorial modified Stroop task and two Implicit Association Tests with offenders from the Granada Institute and Arbour Hill Prison, Dublin.

4.1.1   Introduction

Having demonstrated in the previous study (study 2, chapter 3) that both the pictorial modified Stroop task and the Gender-Sex IAT were able to tap into the sexual interests/associations of non-offending participants, and after arguing that a lack of an effect for the Child-Sex IAT may in fact make sense for a non-offending sample, the next goal of the thesis was to explore whether the measures could demonstrate differences between men who had committed sexual offences against children and men who hadn't. The IATs used in the current study were identical to those used in study 2, as was the pictorial Stroop task except that images were presented in clusters of 16 images of the same trial type instead of blocks of 48 images of the same type. This was done in order to reduce potential order effects of the larger block design.

 As has already been mentioned, not all offenders against children have a sexual preference for, or even a sexual interest in, children (Seto, 2008). Therefore any study that explores a measure's ability to identify sexual interest in children among a sample of offenders must attempt to preselect those offenders most likely to hold such

preferences. There are many ways in which researchers could attempt to categorise those offenders who are likely to have a sexual interest in children or not. This could include number of victims, gender of victims, age of victims etc. However, given the small sample size of the current study it was decided to use the relationship of the offender to the victim as an indicator of the likelihood of that offender having a sexual interest in children. Studies have shown incest offenders to typically demonstrate lower arousal to (visual) child stimuli than offenders with extrafamilial victims (Freund & Watson, 1991; Murphy et al., 1986; Seto et al., 1999). In addition, admitted sexual interest in children was also considered.

The study set out first of all to explore whether there would be a group difference in response times on the pictorial modified Stroop task between offenders and non-offenders. It was then planned to look at the response times for those deemed more likely to demonstrate a sexual interest in children relative to adults. It was hypothesised that extrafamilial offenders and offenders with an admitted sexual interest in children would differ in their response times to children relative to adults when compared with non-offending controls and most incest offenders (i.e. those without an admitted sexual interest in children). It was intended to carry out similar comparisons using the Age-Sex IAT, comparing offenders (and subgroups of offenders) with control participants. However, given that it is suggested in this thesis that the IAT is likely to be activating schema related to a possible implicit theory of children being associated with sex, it may not follow that extrafamilial offenders would be more likely than incest offenders to have that association. In fact, Hanson, Gizzarelli and Scott (1994) found incest offenders to have associations between sex and children.

The current study also set out to investigate whether the Stroop and Gender-Sex IAT were able to tap into the adult sexual preferences of participants and whether they were reliably able to discriminate between participants based on sexual orientation. It was hypothesised that participants would have longer reaction times to orientation-consistent images in the pictorial Stroop task and shorter reaction times to orientation consistent word pairings in the Gender-Sex IAT. It was also hypothesised that both tasks would discriminate between the orientation of participants with an accuracy significantly greater than chance. It is intended to compare the results of both IATs and the pictorial Stroop task to see if there was a correlation between the measures. It was

hypothesised that there would be a correlation between the results of the Gender-Sex IAT and the pictorial modified Stroop task as this had been found by study 2 looking at non-offending participants. There was no hypothesis about the relationship between the Age-Sex IAT and the pictorial modified Stroop results. Receiver Operating Characteristic (ROC) analysis were to be used to explore the sensitivity and specificity of the different measures in discriminating between offenders and controls, groups of offenders and also participants based on their sexual orientation. While it was hypothesised that ROCs would significantly discriminate between gay and straight participants, all other comparisons were exploratory.

A traditional Stroop task was included in the study to 'train' participants in carrying out a Stroop-type task and also to be able to compare the selective attention and cognitive flexibility of offenders and non-offenders (Strauss et al., 2006). While the Stroop is possibly too limited a measure to assess whether different samples of participants are comparable in terms of general cognitive ability it should be an adequate measure of task-specific cognitive speed.

4.1.2 Method

4.1.2.1 Design
A mixed factorial design was used with offending and non-offending control participants completing all experimental tasks; the pictorial modified Stroop task, the Gender-Sex IAT, the Age-Sex IAT as well as a version of the traditional Stroop task.

4.1.2.2 Participants
Twenty four men who had committed sexual offences involving children and 24 non-offending control participants took part in the study. Ten offenders were recruited from the Granada Institute which is a community treatment setting, while fourteen were incarcerated offenders recruited from Arbour Hill Prison. Both Institutions are in Dublin. Control participants were recruited by a combination of college notice-board and poster recruitment, use of participant lists and recruitment of individuals attending public lectures in psychology.

All Granada Institute clients who participated in the study were attending group therapy. Clients were first told of the study by their therapists. The researcher then spoke to the clients within their treatment groups. Those interested communicated their interest to the therapists and appropriate participation times were arranged. These offenders were at various stages of treatment and were attending for contact offences (i.e. child molestation), non-contact' (e.g. child pornography, exhibitionism) offences or both. Testing of these participants took place in the Granada Institute.

Potential participants recruited from Arbour Hill prison were first approached by the Governor of the prison or a chief officer and were given the option of meeting the researcher[24]. Eighteen inmates met the researcher who explained what the research entailed. Of those 18 inmates, 14 eventually participated (two declined to take part and two were unable to due to illness or court appearances). The incarcerated participants were all convicted of sexual assault, rape, indecent assault, gross indecency or a combination of these offences. All offences involved child victims. One of the inmates was serving a life sentence, the remaining 13 participants had a mean sentence of 6.6 years (SD = 2.78 years). The total number of eligible inmates at the time of testing, according to the prison authorities, was 49. This included three inmates serving life sentences, one on remand and two where duration of sentence was not available to the author. The remaining 43 inmates had a mean sentence of 6.3 years (SD = 3.9 years). Testing took place in the school building within Arbour Hill Prison. The testing room overlooked a prison courtyard where inmates were occasionally talking loudly during testing sessions.

Offending (incarcerated and community) participants had a mean age of 50.13 years (SD = 15.92 years) while control participants had a mean age of 42.26 years (SD = 20.61 years). Twenty one of the twenty four control participants had completed Leaving Certificate or higher. Fourteen offenders had completed their Leaving Certificate or greater, while ten had not. The Leaving Certificate is the state exam completed by students in the Republic of Ireland at the end of their secondary education and before entering university.

---

[24] While the prison authorities were asked to try to recruit any inmate whose conviction involved a sexual crime involving a child it is likely that some selection bias occurred and that only the more compliant inmates were asked.

Background data on offending participants comes from a combination of questionnaire responses and information from therapeutic files for community sample participants and from the longer offence history questionnaire in the case of the prison participants. Of the total sample of offenders there were 21 contact offenders and three non contact offenders. Of those three, two had child pornography offences and one had exhibitionism offences. Two of the contact offenders also had pornography offences. All non-contact offenders had offences involving female children only. The contact offenders included nine with offences against female children, 10 with offences against male children and 2 with offences against both. Ten contact offenders had only unrelated victims, 10 had related victims and one had both related and unrelated victims. Of the 11 offenders with related victims, six had a paternal relationship with their victims (father/step-father/guardian). Seventeen offenders had offences (molestation, exposing or child pornography) involving children under the age of 13 or admitted having a sexual interest in children younger than 13. The remaining seven had victims that were all 13 years of age or over. Eleven of the offenders abused more than one child.

On completion of the Stroop tasks, one offending participant indicated that he had adopted a strategy to minimise his response times on the task (most likely squinting or not looking directly at the screen). His Stroop task results were therefore excluded. One participant did not carry out the IATs because of reading difficulties while another participant was only able to complete one IAT in the available time.

4.1.2.3   Apparatus/Materials

Computerised tasks were presented on a Gateway Pentium III computer with a Gateway EV9108 19" CRT monitor or on a Gateway Solo 9300 Laptop. Participant responses were made via a Cedrus response pad (model RB-620) with four coloured buttons (red, green blue and yellow). Tasks were run using the SuperLab4© stimulus presentation software. A questionnaire was also administered to participants. There were three versions of the questionnaire. All versions contained questions asking the participants about age, colour blindness, education and sexual interest in adults. Offending participants were also asked about sexual interest in children and about their offences

(offence type, victim age, relationship to victim). Offenders from the community treatment setting and the prison setting received slightly different versions of the questionnaire. There were more detailed questions about treatment for those in the community setting. Prison participants were asked more detailed offence questions since detailed files were available to the authors for the community participants. All questionnaires are reproduced in appendices 4, 5 and 5. Stimuli for the Stroop task and the IATs were the same as those used in Study 2.

4.1.2.4 Procedure

Participants in this study followed the exact same procedure as study 2 (chapter 3) except that the pictorial modified Stroop task adopted a clustered, rather than a blocked design and that the questionnaires were different and were filled in at the end of testing. Stimuli in the pictorial Stroop task were grouped in clusters according to stimulus types; therefore there were three clusters each of five trial types: adult males, adult females, child males, child females and large cats (as control images). Each cluster contained four images in four colours, yielding a total of 16 images per cluster. There were novel images in each cluster and the clusters of child stimuli contained images from each of four Tanner stages. As with the traditional Stroop task, participants had to identify, using a button box, the colour with which each stimulus had been tinted. Cluster order was randomised across participants, with the condition that a cluster of each trial type had to be presented before a second cluster of any type could be presented and so on. Trials were randomised within cluster. Response time was recorded for every trial along with whether that response was correct or not. Traditional Stroop task and IAT procedure did not differ from study 2.

4.1.3 Results

Several different analyses were conducted on the data. First the results of the traditional Stroop task were compared across the offending and control groups. The pictorial modified Stroop task results were then analysed in several ways: comparing the results of gay and straight participants, comparing the results of offenders and non-offenders and finally comparing the results of offenders hypothesised to be the most likely to demonstrate deviant sexual interest with remaining offenders and non-offenders. The

108

results of the IATs were used to carry out similar comparisons. The Gender-Sex IAT was used to compare gay and straight participants and the Age-Sex IAT was used to explore offender/control differences using the same groupings as for the Stroop task. ROC analyses were conducted in addition to the between groups analysis to explore the ability of the IATs and pictorial Stroop task to accurately differentiate between groups of participants. Finally correlations were explored between the pictorial modified Stroop and the IATs.

4.1.3.1 The traditional Stroop task

A 2x3 mixed factorial ANOVA was carried out, comparing mean reaction times (with outliers three times the interquartile range beyond the 25th and 75th percentiles removed), for both offending participants and non-offending controls to congruous, control and incongruous trials on the traditional Stroop task. There was a significant main effect of trial type; $F (1.191, 52.388^{25}) = 25.799$, $p < .001$, partial $\eta^2 = .37$. Pairwise comparisons (using a Bonferroni adjustment for multiple comparisons) demonstrated that response times for incongruous trials were significantly slower than both other trial types ($p < .001$) demonstrating a traditional Stroop effect. There was no interaction of trial type and participant group indicating that both groups had results that followed the same pattern; $F (1.191, 52.388) = .6$, $p = .47$, partial $\eta^2 = .01$. However, there was a significant between groups effect indicating that the offending group (Mean = 1040.53, $SE = 53.34$) performed slower overall than the control participants (Mean = 793.3, $SE = 53.34$); $F (1, 44) = 10.725$, $p = .002$, partial $\eta^2 = .196$. This result was taken to indicate that there was a difference in cognitive ability/speed between both groups, suggesting that the offending participants could have slower reaction times across cognitive tasks. Ipsative z-scores were, therefore used when comparing the results of the pictorial modified Stroop task for two groups, in order to potentially minimise the impact of individual differences in cognitive speed.

---

[25] For this ANOVA Mauchly's test indicated that the assumption of sphericity was unsafe, $\chi^2 (2) = 48.964$, $p < .001$, therefore the degrees of freedom were changed accordingly using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .595$).

4.1.3.2 The pictorial modified Stroop task

Ipsative *z*-scores are sometimes used in PPG analysis and the analysis of viewing time tasks to reduce individual differences in arousal/response time (Murphy & Barbaree, 1994; Sachsenmaier & Gress, in press). Study 2 (chapter 3) demonstrated that ipsative *z*-scores may reduce the impact of individual variability in overall response times on results. Therefore ipsative *z*-scores were again calculated for the pictorial Stroop data in this study. Values were calculated by taking the mean reaction time for one type of image, subtracting the overall mean for images from it and dividing it by the standard deviation calculated for all image trials. Outliers that were over 3 times the interquartile range beyond the 25th and 75th percentile were removed before these means and standard deviations were calculated. The ipsative *z*-score method yielded values that indicated how many standard deviations an individual's trial type mean was away from their overall mean. Negative values indicated trial type means that were quicker than the grand mean while positive scores were slower.

4.1.3.3 Comparing offenders with non-offending controls

A 2x2x5 mixed factorial ANOVA was carried out on the results of the pictorial modified Stroop task, where both sexual orientation and offending status (i.e. offending or control) were between subjects variables and trial type was the within subjects variable. Trial types in this case referred to the image categories of adult female, adult male, child female, child male and large cat. Sexual orientation was based on self-report for the control participants and for the prison offenders. For community offenders self-reported sexual interest was compared with file information and in two cases, where file and self-report were at odds, the orientation was taken to be that reported in the file. One prison participant was removed from the analysis as his self-reported sexual orientation, sexual history and offence history did not give a clear indication of sexual orientation[26]. The ANOVA demonstrated a significant interaction between trial type and sexual orientation; $F(4, 168) = 3.163$, $p = .015$, partial $\eta^2 = .07$. There were no interactions involving offender status indicating that both offending and non-offending participants yielded similar patterns of responding.

---

[26] Therefore, in this analysis there were twenty two offending participants, nineteen of whom were classified as straight and the remaining three classified as gay. There were twenty four control participants where twenty one identified themselves as straight and the remaining three identified themselves as gay.
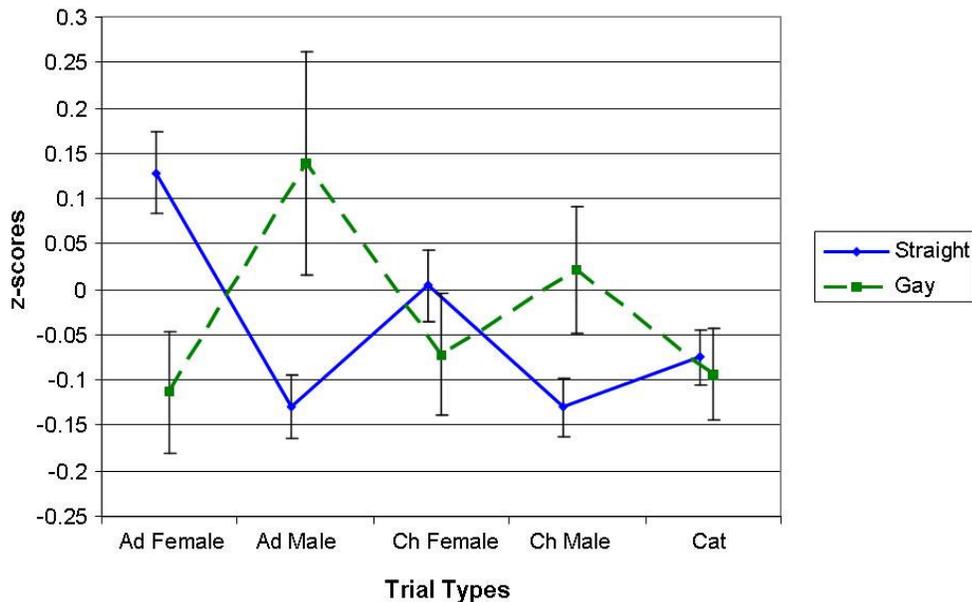
Figure 4.1: Reaction times (as ipsative z-scores) across trial types for gay and straight participants (error bars indicate one standard error).

Figure 4.1 graphically depicts the interaction of trial type and sexual orientation. There is a clear pattern in the results where orientation-consistent, age-appropriate trial types yield the slowest reaction times followed by orientation consistent, age-inappropriate trial types. However, given the fact there were only three gay participants in the offending sample, only three gay participants were recruited to the control sample and, therefore, comparisons between task results based on orientation, while indicative of a strong effect, should be interpreted cautiously given that the small number of gay participants may not constitute a representative sample.

4.1.3.4 Comparing familial offenders, extrafamilial offenders and control participants

As stated there was no interaction between offenders and controls in their responses to different trial types. However it was hypothesised that all offenders would not demonstrate a sexual interest in children so this result is not surprising. As stated, extrafamilial offenders are typically demonstrated, by PPG using visual stimuli, to demonstrate higher levels of deviant arousal, usually considered indicative of higher deviant sexual interest. Therefore, the sample of offenders was divided into those who

had exclusively offended against family members and those who had extrafamilial victims. Nine of the offenders admitted to having a sexual interest in children. When asked what age child they found sexually interesting, two of the nine referred to children over the age of 16. The remaining seven participants admitted a sexual interest in children that were all 15 years of age or younger. Consistent with the expectation that extrafamilial offenders would show greater deviant arousal, all bar one of the offenders with an admitted sexual interest in children had extrafamilial victims. The intrafamilial offender who admitted a sexual interest in children was included in the group of offenders considered likely to demonstrate Stroop results indicative of deviant sexual interest. Offenders with pornography only and exhibitionism only offences were also included in this group, referred to as 'extrafamilial' for brevity despite the inclusion of one intrafamilial offender (and the fact that one of the participants had extra- and intrafamilial victims). There were fourteen participants in the extrafamilial group, eight in the intrafamilial group and twenty four in the control group.

Given the size of the sample and to simplify the analysis, the response times of participants were re-organised to remove the need to add sexual orientation and gender of victim as variables in the analysis. A new variable was produced consisting of reaction times across participants for orientation-consistent adult images regardless of whether those images were of males or females. The same was produced for orientation-inconsistent adult images. For control participants, child images were treated identically with orientation-consistent and orientation-inconsistent trials being re-coded into two variables. For offenders the new child variables contained offence-consistent or offence-inconsistent reaction times. For offenders with victims of both genders, the participant's stated preference for gender of child was used as a guide. Figure 4.2 graphically shows the pattern of responses across the four trial types for the three groups of participants: control participants, extrafamilial offenders and intrafamilial offenders. All three groups show a similar pattern of responses with reaction times to orientation consistent (or offence consistent) images being higher than to orientation inconsistent images
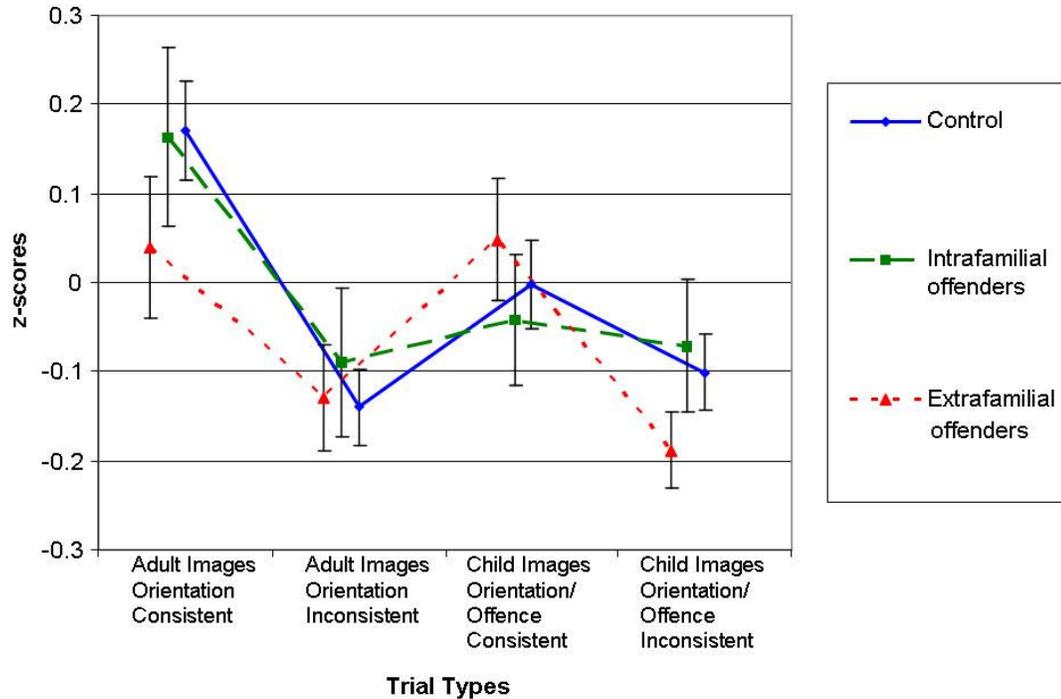
Figure 4.2: Response times to stimulus types for control participants, for offenders with exclusively familial victims and for offenders with at least one extrafamilial victim or an admitted sexual interest in children (error bars indicate one standard error).

A 3x4 mixed factorial ANOVA was carried out comparing the response times of the three groups (control, intrafamilial and extrafamilial). There was a main effect of trial type; F (3, 129) = 7.668, p < .001, partial η^2 = .151. There was no interaction of group and trial type; F (6, 129) = .605, p = .726, partial η^2 = .151. Given that this was an exploratory study several planned comparisons were carried out. First, paired samples t-tests were carried out to compare response times to adult and child orientation/offence-consistent stimuli. Control participants had indicated significantly slower reaction times to adult stimuli, t (23) = 2.076, p = .049, r= .4, two-tailed. Offenders with extrafamilial victims did not differ significantly in their response times to age appropriate and age inappropriate stimuli, t (13) = -.066, p = .948, r= .02, two-tailed. Offenders with only incest offences also had a result that indicated no significant difference between reaction times to adult and child stimuli, t (7) = 1.268, p = .245, r= .43, two-tailed. However, this difference represented a larger effect size than there had been for the control group and it can be seen from Figure 4.2 that the response pattern of the incest

113

offenders and the control participants mirror each other quite closely. A small sample size in the incest group is likely to have contributed to the lack of significance.

Next a series of t-tests were carried out to explore whether control participants had significantly longer reaction times to preferred versus non-preferred adult stimuli and whether this pattern was the same for child stimuli. These results were then compared with those of intrafamilial and extrafamilial offenders since it was hypothesised that extrafamilial offenders would differ from controls while intrafamilial would not. For control participants, response times to orientation consistent images of adults were significantly longer than those for orientation inconsistent adult stimuli; $t(23) = 4.099$, $p < .001$, $r = .64$, two-tailed, while the difference between orientation consistent and inconsistent child images was not significantly different; $t(23) = 1.505$, $p = .146$, $r = .3$, two-tailed. The intrafamilial offenders show a similar pattern to control participants with a large effect size for the adult stimuli difference and a small effect size for the child stimuli difference. Unlike the control participants the difference between orientation consistent and inconsistent adult images is not significant; $t(7) = 1.595$, $p = .155$, $r = .51$, two-tailed, though this is likely due to size of the sample of intrafamilial offenders. The difference between offence consistent and inconsistent child images is also non significant but, as mentioned, has a small effect size; $t(7) = .229$, $p = .826$, $r = .1$, two-tailed.

In contrast to the results of the intrafamilial offenders and the control participants, extrafamilial offenders demonstrated larger differences between response times to offence consistent and offence inconsistent child images than they did to orientation consistent and inconsistent adult images. The difference between child images were significant; $t(13) = 2.427$, $p = .03$, $r = .56$, two-tailed, while they were not for adult images; $t(13) = 1.447$, $p = .172$, $r = .37$, two-tailed.

As multiple analyses were carried out which increases the risk of familywise error a correction should have been applied to the critical p-value in the above t-tests. However, given that the analysis was exploratory it was deemed appropriate to simply exercise caution in extrapolating from the results. The pattern of results and magnitude of the effect sizes for the extrafamilial offenders, seemed to indicate that there was not a very large difference between responses to child and adult stimuli. On the other hand

114

intrafamilial and control participants showed longer responses to adult stimuli. While, as a group, the extrafamilial offenders seem to demonstrate more deviant interest than the incest offenders and non-offending controls there is considerable heterogeneity in patterns of responding in each of the three groupings. Importantly, over one third of control participants show reaction times that are slower to child images than either of their adult image response times.

4.1.3.5 *The ability of difference scores to discriminate between offenders and non-offenders and to differentiate between participants based on orientation*

Receiver Operating Characteristic (ROC) analysis was used to explore the sensitivity and specificity of a measure based on the difference between reaction times to orientation-consistent adult images and orientation/offence-consistent child images in predicting whether a participant belonged to the offending group or not. The area under the ROC curve (AUC) indicated that the measure was not very successful in discriminating between offenders and non-offenders, AUC = .591, p = .296, SE = .086. The measure faired slightly worse when discriminating between the offenders with at least one unrelated victim (including, child pornography offenders, exhibitionists and one incest offender with an admitted sexual interest in children) and all other participants, AUC = .588, p = .352, SE = .099. These results are included in Table 5.

Study 2 (chapter 3) explored the utility of an ROC curve in establishing whether the results of the pictorial modified Stroop task could differentiate between participant based on sexual orientation. To demonstrate whether this was the case with the current data, an ROC was plotted to see how well the Stroop task classified participants as either gay or straight. The ipsative z-score method yielded a significant AUC value, AUC = .808, p = .016, SE = .075. However since the number of gay participants in the sample was very low, this value should be interpreted with caution. Given that one argument for the use of implicit tasks is that it may give insight into phenomena that individuals are unwilling or unable to report accurately it seems problematic to validate these measures against self-report. Therefore, an alternative orientation variable was created where participants, both control and offending, who had had male sexual partners, had stated some level of sexual interest in males or had sexually offended against male children were classed as having a sexual interest in males, while all others

were classed as having a sexual interest in females. The difference between the mean response times to adult males and adult females (converted to z-scores) gave an AUC of .894 ($p < .001$, *SE* = .048) when discriminating between groups using this new classification measure. The analysis indicated a cut-point of .0474 that yielded the best trade-off between sensitivity and specificity. At this cut-point all but two participants (of 11) with a hypothesised sexual interest in males were correctly classified as such. Of the two that weren't, one was an offending participant and had higher reaction times to male children than any other stimulus type. At the same cut-point 7 of 36 participants hypothesised as having a sexual interest in females were incorrectly classified. Five of these were offenders and three of those had higher reaction times to female children than any other trial type. Of the control participants, one had very little variability across all trial types, suggesting that the Stroop may not have invoked a content induced delay for this participant at all.

4.1.3.6 The Gender-Sex IAT
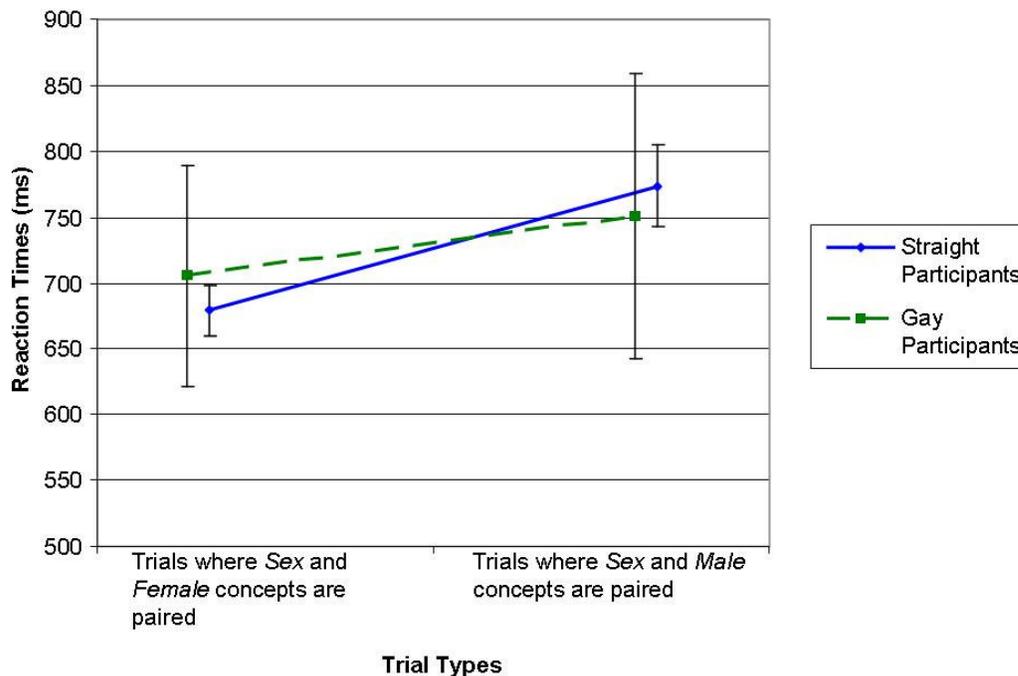
Figure 4.3 Reaction times to male-sex and female-sex concept pairs in the Gender-See IAT for gay and straight participants (error bars indicate one standard error).

116

Study 2 (chapter 3) did not counter-balance order of concept-pairing presentation in the IAT tasks. Given that Gender-Sex IAT in that study was demonstrated to tap into the sexual associations of participants, no counterbalancing was adopted in study 3[27]. However, from Figure 4.3 it is clear that response times were quickest for the first concept pairing that was presented, in this case the sex-female pairing. The greatest slowdown in response times to the second-concept pairing, male-sex, occurred for the straight participants. The difference between mean reaction time to female-sex trials and male-sex trials was significant, t (37) = -4.915, p<.001, r = .63, two-tailed. On the other hand, for gay participants, the difference between mean reaction time to female-sex trials and male-sex trials was not significant, t (5) = -.774, p = .474, r = .32, two-tailed. Clearly the second t-test lacked statistical power due to the small number of gay participants in the study. Additionally the standard error for gay participants is large, as can be seen from the figure, again due to the small sample size. However, the difference in effect size between the two analyses seems to suggest that orientation *might* have moderated a slowdown effect that occurred as participants changed from one type of concept pairing to a second.

For the Gender-Sex IAT, study 2 (chapter 3) found, using ROC curve analysis, that an identical Gender-Sex IAT had an AUC .806 in discriminating between gay and straight non-offending men. This was calculated by subtracting the mean reaction time for trials where *male* and *sex* were paired from trials where *female* and *sex* were paired after outliers had been removed. The current study did not find a significant AUC value for the results of the Gender-Sex IAT, AUC = .711, *p* = .101, *SE* = .135. While this AUC could be considered 'fair' (Tape, 1999) it has already been stated that the small number of gay participants in the sample may have impacted on the results of the ROC analysis. In addition, three of the gay participants were offenders and offenders were typically slower in speed of responding overall which may have impacted on the results. Given, the small numbers of gay participants in the study, looking at the results of the Gender-Sex IAT may be of limited value.

---

[27] It was deemed useful to avoid counter-balancing so that the small number of available offending participants in this study would complete identical tasks. Nosek, Greenwald, & Banaji (2005) found that designing the IAT in such a way that there were additional trials in the 5th block (where the participants 'learns' the new stimulus configuration) reduced or eradicated any order effect. Such a design was adopted in each of the IATs in this thesis.

4.1.3.7 The Age-Sex IAT
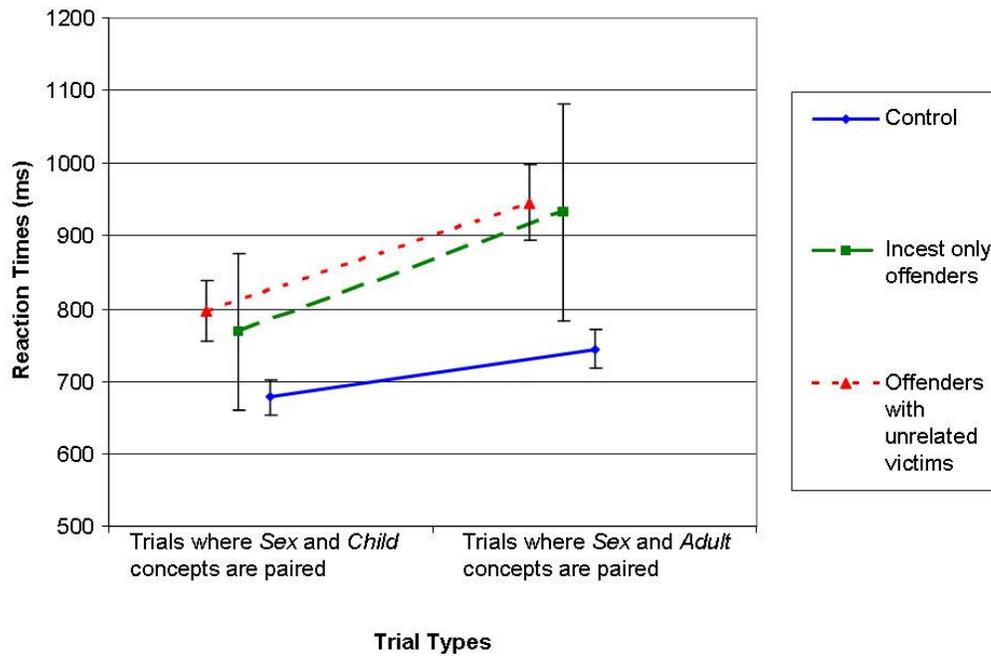
Figure 4.4: Reaction times to adult-sex and child-sex concept pairs in the Age-See IAT for control participants, participants with incest only offences and participants with at least one extrafamilial victim (error bars indicate one standard error).

The analysis of the results of the pictorial modified Stroop task gave a certain amount of support to the hypothesis that incest offenders would show responses consistent with those of non-offending controls, while extrafamilial offenders would demonstrate responses indicative of potential deviant sexual interest. The Age-Sex IAT results of the same three participant groups were therefore compared first. Figure 4.4 shows mean reaction times to different trial types from the Age-Sex IAT across three different participant types. As with the Gender-Sex IAT all participant groups showed faster reaction times to the first concept pair than the second. Contrary to the study's hypothesis, control participants showed significantly faster responses in trials where the concept of *child* and *sex* were paired than when *adult* and *sex* are paired, t (23) = -3.211, p = .004, r = .56, two-tailed. Both offending groups also showed significantly faster responses to *child* and *sex* pairings. Incest only offenders, t (7) = 3.146, p = .016, r = .77, two-tailed, and offenders with at least one extrafamilial victim, t (14) = 3.561, p = .003, r = .69, two-tailed, both had considerably larger effect sizes than that found in non-offenders.

However visual inspection of Figure 4.4 suggested that there may have been an overall difference in response times across trial types between offenders and non-offenders similar to that observed with the traditional Stroop task. To explore this, a 2x2 mixed factorial ANOVA looking at trial type across offenders versus controls was carried out. The ANOVA found a significant main effect of group, F (1, 45) = 6.562, p = .014, partial η^2 = .127, and also of trial type; F (1, 45) = 34.044, p < .001, partial η^2 = .431. There was also an interaction between trial type and group; F (1, 45) = 5.169, p = .028, partial η^2 = .103. Main effects should be interpreted in light of the interaction. However, given the presence of a significant between groups difference, it is possible that the greater difference between trial types in offenders is due to differences in cognitive speed. That is to say that those participants who were slower in responding to the first-presented concept pair might have been differentially disadvantaged when the concept pairings were swapped.

Given the concerns about the impact of possible difference in cognitive speed among offenders and control participants a method of data treatment that would minimise this impact was desirable. A t-test was carried out comparing the Age-Sex IAT responses of offenders and control participants using Greenwald et al.'s (2003) D scoring algorithm, which calculates a difference score between blocks of trials containing *child-sex* concept pairings and blocks containing *adult-sex* concept pairings, which is argued to reduce the impact of speed of responding on the IAT effect. Similar to the ipsative z-score method used for the pictorial modified Stroop task, the method uses standard deviations to minimise differences due to individual variability in response times. A difference was found between the average D score of both groups that was approaching significance, t (45) = 1.999, p = .052, r = .29, two-tailed. Concerns have already been raised about the appropriateness of Greenwald et al.'s D method given that extreme outliers are potentially retained by the analysis (i.e. that only outliers greater than 10 seconds are removed). With this in mind an algorithm was used that mirrored that of Greenwald et al. (2003) except that outliers were removed if they were 3 times the interquartile range beyond the mean. This method yielded a slightly smaller t value, t (45) = 1.848, p = .071, r = .27, two-tailed.

Two independent sample t-tests using Greenwald et al.'s (2003) D score and the alternative D algorithm as described above method found no significant difference

between Age-Sex IAT response times for those offending participants who admitted a sexual interest in children and non-offending controls. Additionally, two one-way ANOVAs comparing the *D* and alternative scores across extrafamilial, incest and control participants also found no significant difference between groups. Scores though non-significant, were lower for controls relative to offenders for all comparisons, indicative of weaker associations of children and sex.

The offending group and the control group did seem to differ in terms of cognitive speed. It was, therefore, important that any ROC curve that would attempt to quantify the ability of a measure to discriminate between offenders and non-offending control participants would use measures relatively unbiased by cognitive speed. Greenwald et al.'s *D* was used, as was the alternate version calculated with outliers removed. These two measures were used to calculate an AUC for discriminating between offending and control participants. They were additionally used to calculate an AUC to quantify their ability to discriminate between offenders with any extrafamilial victims (including pornography and exhibitionism offenders and an incest offender with an admitted sexual interest in children). The results of the analyses are summarised in Table 5. The ability of difference scores from the pictorial modified Stroop task to discriminate these groups is also summarised in the table. Only the Age-Sex IAT score measured using the modified version of the *D* algorithm had an AUC that differed significantly than one that might have been achieved by chance for both group classifications.

Table 5: AUCs demonstrating the ability of three different measures to discriminate between offenders and control participants.

Difference score	AUC	*P* value	Standard Error
All offenders versus non-offending controls			
Age-Sex IAT Greenwald et al. (2003)*D* algorithm	.664	.059	.081
Age-Sex IAT alternative *D* algorithm (outliers removed)	.696	.025	.079
Stroop mean RTs converted to *z*-scores	.591	.296	.086
Extrafamilial offenders versus all other participants			
Age-Sex IAT Greewald et al. (2003)*D* algorithm	.611	.239	.094
Age-Sex IAT alternative *D* algorithm (outliers removed)	.700	.033	.091
Stroop mean RTs converted to *z*-scores	.588	.352	.099

4.1.3.8   *The relationship between the IATs and the pictorial modified Stroop task*

Study 2 (chapter 3) found a significant correlation between the blocked version of the pictorial modified Stroop task and the Gender-Sex IAT ($r = .559$, $p = .008$). In the current study, there was no such correlation between the two measures, even among control participants. In addition there were no significant correlations between the results of the Age-Sex IAT and a difference score based on response times to child images subtracted from a response time to adult images in the pictorial modified Stroop task. This correlation was again lacking for both offenders and control participants.

4.1.4   Discussion

The study found that offenders overall did not differ from non-offenders in their patterns of responding on the pictorial modified Stroop task. However, when subsets of offenders were compared, offenders with an admitted sexual interest in children or with extrafamilial victims showed less difference between response times to adults and children, than did incest offenders or non-offenders. ROC analyses showed some utility

in differentiating between gay and straight participants based on their pictorial modified Stroop results but not in differentiating between offending and non-offending participants. All IATs showed slowdown of response times across participant groups between the first and second concept-pairs. In the case of the Gender-Sex IAT this slowdown *seemed* to be moderated or exaggerated by the sexual orientation of the participants. Similarly the slowdown observed in the Age-Sex IAT *seemed* to be exaggerated or moderated by whether the participants belonged to the offending or non-offending groups. Using the D score the results of the Age-Sex IAT did differ between offenders and non-offenders but not between subgroups of offenders. ROCs based on the ability of the slightly modified version of the D algorithm were able to discriminate with a 'fair' amount of accuracy (Tape, 1999) between offenders and non-offenders. There were no significant correlations between the results of the pictorial modified Stroop task and the IATs.

The results of the traditional Stroop task suggested that offenders were slower in general in carrying out the cognitive tasks. The same pattern was found across the subsequent pictorial modified Stroop task and Implicit Association Tests (IATs). The potential for an influence of group differences in cognitive speed were therefore taken into account in the analysis an interpretation of results. The results of the pictorial modified Stroop task supported the hypothesis that the task taps into sexual interest. Taken as a whole, the sample of participants (both non-offending and offending) responded in a pattern consistent with their sexual orientations. Interestingly the average pattern of response showed the longest reaction times to orientation-consistent stimuli, regardless of whether those stimuli were age-appropriate or not. This will be discussed later. For control participants, and as would be expected, reaction times for orientation-consistent stimuli were significantly longer for adult stimuli than child stimuli, though both were longer than for orientation-inconsistent or control images. This pattern was mirrored in incest offenders, though the sample lacked sufficient size to achieve significance in comparing the difference between the age-appropriate and age-inappropriate orientation/offence consistent stimuli. Extrafamilial offenders had a group mean reaction time for images of children (offence-consistent) that was almost identical to that for adults (orientation-consistent). This result was in line with the hypothesis that this group of offenders would show the most deviant sexual interest.

The difference between reaction times to orientation-consistent adult images and orientation/offence-consistent child images did not reliably differentiate between offenders and non-offenders. This could be for several reasons. First, the task itself may not tap into sexual interests. This explanation seems unlikely, given the patterns found in this study and also in study 2 (chapter 3), albeit using a different version of the pictorial modified Stroop task. Second, it is possible that child stimuli grab participants' attention in a way that is less attributable to sexual interest than might be the case for adult stimuli. It is also possible that sexual interest in children, or at least the capture of attention in evaluating the sexual salience of children, is a phenomenon not limited to offenders. Third, it could be that the presence of older children in the stimulus sets had a disproportionate influence on the increase in reaction times (a quarter of child stimuli were images of children corresponding to Tanner stage 4). Additionally, if a rumination effect is implicated in the pictorial modified Stroop task for sexually salient stimuli there is not necessarily a direct relationship between the images presented and that which is ruminated about or the associations activated. This has implications for the utility of the pictorial modified Stroop as a clinical tool and therefore highlights the need for further research and validation

In line with previous PPG research there was a difference between extrafamilial and intrafamilial offenders in their responses to the pictorial modified Stroop task. However, the study lacked a sufficient sample size to further subdivide incest offenders into those with a paternal relationship with their victim(s) and those with other familial relationships. Previous research using PPG suggests that there is an important distinction within incest offenders when it comes to the degree of deviant sexual interest present (Blanchard et al., 2006; Seto et al., 1999). In addition Murphy et al. (1986) report that stimulus modality can impact on the demonstration of deviant interest among incest offenders in PPG studies, with audio descriptions eliciting a deviant response pattern in that group. This may be related to the fact that an audio description may allow an offender to imagine their own victim (Murphy & Barbaree, 1994). A pictorial modified Stroop design may also be unable to identify patterns that another stimulus modality could. Research on the development of a battery of cognitive tasks is therefore recommended.

The current study only treated incest offenders as less likely to demonstrate deviant sexual interest. There were many other comparisons that could have been carried out, such as comparing results of those with male victims, multiple victims, prepubescent victims or by looking at their SSPI classifications etc. These comparisons would have been possible with a larger sample. However the goal of the current study was to determine if the pictorial modified Stroop task had initial promise in exploring deviant sexual interest in children. The finding that extrafamilial offenders and those admitting a sexual interest in children did differ from control participants supports the further exploration and validation of the measure. To properly establish if the task has a real clinical utility, it will be necessary to compare the results, not just with demographic and offence details but also with PPG results and with the results of other indices of sexual interest.

In line with the stated hypothesis, ROC analysis indicated a good trade-off between sensitivity and specificity in using the pictorial modified Stroop difference score to discriminate between participants' sexual orientation (based on self-report or on clinical file). This ROC was based on a very small number of gay participants, however. The difference score performed even better when sexual interest was deduced from whether participants had stated any degree of sexual interest in males or had reported any sexual contact with male adults or children. Though it is clearly problematic to classify people in this way it is also problematic to rely solely on self-report. This was illustrated by the fact that several of the offenders for which there was file information available indicated on their questionnaires that they were heterosexual, while the clinical file mentioned that they were, in fact, gay. In addition, volunteer bias in sexual research can be particularly problematic (Trivedi & Sabini, 1998; Wiederman, 1999). It is likely that such research may attract people questioning or struggling with their sexuality. In addition it is hypothesised that the pictorial modified Stroop task is a measure of attention related to the arousal process. It is likely that this process alone is not sufficient to capture the complexity of human sexual orientation.

Though the results of the pictorial modified Stroop indicated that the task is measuring attentional processes related to sexual interest, these results were not as clear-cut as had been expected. A clustered approach was taken instead of the blocked one used in study 2 (chapter 3) in order to reduce the potential for an effect of order. However a blocked

design may have the capacity to induce a stronger effect and thereby potentially better discriminate between those with a sexual interest in children and those without. Study 5 (Chapter 5) in this thesis explores whether, using a between subjects experimental design, blocked presentation of the pictorial modified Stroop task for sexual interest significantly outperforms clustered and random presentation with non-offending participants. Future studies should consider dividing the child image blocks or clusters into younger and older children (i.e. Tanner stage 1&2 images as one block/cluster and Tanner stage 3&4 as another). This would allow an exploration of both paedophilic and ephebophillic sexual interest.

Taken together the results of the Implicit Association Tests used in this study are unclear. Both seem to be influenced by participants' sexual associations, though it is difficult to state this confidently given the apparent differences in cognitive speed between the offending and non-offending samples and also due to unequal numbers of gay participants. In both tasks, there was an apparent slowdown of response time as the participant switched from carrying out the first concept pairing (e.g. when *child* and *sex* were paired in the Age-Sex IAT) to the second concept pairing (e.g. when *adult* and *sex* were paired). The influence of associations surrounding sexuality seemed to moderate or exaggerate that slowdown in line with the participants' hypothesised sexual interest. While this effect was observable on a group level it was difficult to determine whether it also constituted a reliable individual effect. A difference score based on the Child–Sex IAT did discriminate between offenders and non-offenders using ROC analysis at a level that was better than chance. However the actual AUC was quite low. While the IAT difference score used was based on Greenwald et al.'s (2003) *D* algorithm and therefore designed to reduce the influence of individual differences participants' speed of response, there is a chance that groups differences in cognitive speed impacted on this result. In addition since Greenwald et al.'s (2003) *D* score is designed to for use with an IAT that has an opposite concept category it may be that the inclusion of a neutral (furniture) category impacted on the size of an IAT effect that the task could produce. The presence of a control category means that the *D* score now contains data from half the responses where there isn't hypothesised to be either a strong or a weak association. This would potentially dilute the IAT effect by including what is, essentially, random data with data that may contain an effect.

Study 2 (chapter 3) found evidence for a relationship between the results if the Gender-Sex IAT and the pictorial modified Stroop task (blocked version) in non-offending controls. While the results of the current study supported the hypothesis that the pictorial modified Stroop task would tap into sexual interest and was able to give tenuous support to the hypothesis that the IAT would measure the strength of associations related to sexual cognitions, there was no correlation between the two measures. In the case of the Gender-Sex IAT, the lack of agreement between the two measures may have been due to sampling size issues. It is interesting, though, that in the case of the Age-Sex IAT and the pictorial modified Stroop task that the Stroop task found differences between those most likely to demonstrate deviant sexual interest (i.e. extrafamilial offenders and those with an admitted sexual interest in children) and everyone else, while the IAT found differences between offenders in general and non-offenders. This is consistent with the idea that the pictorial modified Stroop task would measure attention indicative of sexual interest, while the Age-Sex IAT would measure associations indicative of distorted schema or implicit theories. However to draw this conclusion is beyond the scope of the data presented in the current study. Discrepancies between the results of the two types of task may, instead, be as a result of a lack of gay participants (in the case of the Gender-Sex IAT), as a result of the use of a clustered Stroop paradigm or maybe any relationship was obscured by the heterogeneity of the sample used. Alternatively, the original correlation between the Gender-Sex IAT and the blocked pictorial Stroop task from study 2 may have simply been due to chance. Further studies should seek to address these issues in addressing the lack of a correlation between the pictorial modified Stroop and the IATs. A study using a sample of offenders identified as having a sexual interest in children by PPG along with a sample of offenders without deviant PPG scores would allow a clearer comparison of the two methods. If the pictorial modified Stroop is a measure of arousal-related attention and the IAT is a measure of schema-related associations as hypothesised here, these two groups should perform differently on the Stroop task but not necessarily on the IAT. While this may have been the case in this study, the lack of a comparison measure of sexual interest to provide criterion-related validity made a more definitive conclusion impossible.

ROC values calculated in studies such as this one should be interpreted cautiously. Without an objective measure of deviant sexual interest or distorted cognition relating

to children and sex it is impossible to tell what the prevalence of these are among offending or control samples. Therefore the ROC analyses are attempting to quantify the correct classification of participants in deviant interest/association categories based on best-guesses of who should be in those categories. The use of scores that utilise participants' standard deviations to minimise the influence of individual differences is also potentially problematic. Both the ipsative z-scores and the D algorithm use this technique. While justified in their use here, since there was a large discrepancy in cognitive speed across participants, the use of standard deviations in their calculations introduces the potential to inflate the relative differences between trial types of individuals who have low variability across trials. The opposite is true of participants with very large standard deviations.

In conclusion, the present study found evidence for the utility of the pictorial modified Stroop in the assessment of sexual interest in offenders. The evidence for the utility of IATs as designed for this particular study was less strong. However both IATs did seem to measure to some degree concepts or associations related to sexual orientations/interests. Both types of tasks would benefit from their administration to a larger sample of participants, with modifications similar to those just discussed. A comparison of PPG results and the pictorial Stroop and the IATs would further investigate the validity and reliability of these tasks and potentially improve our understanding of the cognitive mechanisms that underpin them. It would be beneficial to researchers and clinicians to have distinct cognitive tasks that were capable of measuring distorted cognition as well as deviant sexual interest.

4.2 Study 4: Comparing the results of a pictorial modified Stroop task and PPG results produced by child molesters and rapists from the Sand Ridge Secure Treatment Center, Wisconsin.

4.2.1 Introduction

A study in progress at Sand Ridge Secure Treatment Center (SRSTC) in Wisconsin afforded the opportunity to test some of the questions raised by Study 3. Researchers there were administering a version of the pictorial Stroop task as one of many different

experimental tasks being used with inmates of the centre. The version of the task used was developed in collaboration with the author of this thesis and used the stimuli that had been developed by the author for the pictorial Stroop task used in studies 2 and 3. Inmates in the centre were also tested using PPG as part of their standard assessment procedure. This allowed a direct comparison between the results of the PPG and the Stroop task.

The institutional procedure at the SRSTC, as with other situations, is to use the PPG results to calculate indices. These indices attempt to measure three dimensions: gender preference, coercion preference and age preference. The PPG procedure involves the measurement of arousal while images depicting individuals of different ages are depicted: infant, preschool age, grammar school age, teens and adults. The preschool age category corresponds to Tanner 1, grammar school to Tanners 2 and 3 and the teen category corresponds to Tanner 4 (P.M. Byrne, personal communication, September 16, 2009). Images are accompanied by audio descriptions of sexual activity with the individuals depicted involving coercion or persuasion (except for infants). Ipsative z-scores are calculated from the maximum recorded arousal response for every stimulus category (the mean of all the maximum responses across categories is subtracted from the responses to individual categories and divided by the standard deviation of all responses). The average z-score for all female categories, regardless of age and level of coercion, is subtracted for the equivalent male category mean to compute a gender preference index. The means for coercion scenarios and persuasion scenarios are used to calculate the coercion index and the age preference index is calculated by computing the difference between arousal to scenarios involving children (not including the teen category) and adults. These indices are used for clinical assessment. For comparison with the pictorial Stroop task, arousal to individual stimulus categories could be used.

As the SRSTC study is still underway, some important information was not available to the author, such as the self-reported sexual orientation of the participants and the breakdown of male and female victims where the offender had victims of both genders. This data was to be collected at a later date, when it would be possible to confirm the data using polygraph. However it was still possible to explore the relationship between PPG, the pictorial Stroop and other offence characteristics. It was hypothesised that the pictorial Stroop task would show group differences between child molesters and rapists.

128

It was further hypothesised that these results could discriminate between the two groups with a reasonable trade-off of sensitivity and specificity. Since all child molesters in the sample are offenders with a clinical diagnosis of paedophilia, have all molested children who are not their own and have all offended against multiple victims, it was predicted that the majority would show some sexual interest in children and therefore were treated as a homogenous group for analysis purposes. It was hypothesised that the pictorial Stroop task would demonstrate differences between offenders with male or female only victims. It was hypothesised that the PPG would also be able to identify the group differences stated above and would be able to accurately discriminate between group memberships. It was also expected that there would be a relationship between arousal measured by the PPG to content-induced delay measured by the pictorial modified Stroop task to corresponding stimuli.

4.2.2 Method

4.2.2.1 Design

A mixed factorial design was used with participants convicted of sexual offences involving children and offences involving adults, referred to as child molesters and rapists respectively in this study[28], completing a battery of experimental tasks, which included a version of the traditional Stroop task, the pictorial modified Stroop task and penile plethysmograph (PPG) testing.

4.2.2.2 Participants

Participants were selected from a pool of 165 civilly committed patients at the Sand Ridge Secure Treatment Center (SRSTC) who had completed PPG testing. Participants were selected for the 'child molester category' if they had contact offences where they had more child victims (less than or equal to 12 years old) than teen victims and no adult victims. Participants in the 'rapist' category were selected if they had more adult contact victims than teen victims and no child victims. All the patients volunteered but were also told that the procedures being tested may become part of the standard SRSTC assessment procedure and were therefore "getting the testing out of the way" by

---

[28] Terms used for brevity.

volunteering then. The tasks and their general purposes (e.g., to measure sexual interests etc) were explained, but no in-depth explanation was give of the implicit methods.

The current study used PPG and pictorial Stroop task data from 40 participants tested to date at the SRSTC, 20 child molesters and 20 rapists. Rapists had a mean age of 50.35 years (SD = 7.47). Child molesters also had a mean age of 50.35 (SD = 8.67). Mean time from last offence was 21.05 years (SD = 4.91) for rapists and 19.1 years (SD = 3.77) for child molesters. Rapists had a mean full scale IQ of 84.5 (SD = 15.08) and child molesters had an IQ of 86.05 (SD = 16.22). PPG results were missing for one participant. Patients in the centre participate in group cognitive-behavioural treatment set in a therapeutic regime.

4.2.2.3 *Apparatus/Materials*

The two Stroop tasks were two of a battery of tasks administered to patients of the SRSTC which included several IATs, an emotional Stroop task, explicit ratings of certain concepts, a self esteem index and a deception scale. The analysis of these measures is being conducted separately and is beyond the scope of the PhD. The battery of tasks was incorporated into the usual assessment procedure carried out at SRSTC. PPG is part of this usual procedure as are the Hare Psychopathy Checklist- revised (PCL-R), the Static 99, and the WAIS. Demographic and offence details are also recorded.

The traditional Stroop task differed from the traditional Stroop tasks in studies 2 (chapter 3) and 3 (this chapter) in that there were only two trial types. One consisted of neutral words while the other consisted of colour names that were incongruous with the font in which the words were presented. There was no congruent condition. The incongruous colour names included many different colour names (e.g. crimson, grey, purple, brown, topaz, mauve, maroon etc.). As with the Stroop tasks used in the other studies in this thesis participants had to identify the colour of the presented words (and later, images) as being coloured in one of four colours; red, green blue and yellow. The pictorial modified Stroop task contained the same images as has been used in studies 2, 3 and 4.

Stimuli used in the PPG procedure consisted of a standard set of non-nude images of children and adults along with audiotaped descriptions of coercive and consensual sexual encounters developed by Behavioral Technology Inc, commonly referred to as the MONARCH stimuli. The MONARCH 21 assessment system was used to measure circumferential penile arousal.

4.2.2.4 Procedure

The Stroop tasks were administered using E-Prime stimulus presentation software. Participants first completed the traditional Stroop task. They then completed an emotional Stroop task and the pictorial modified Stroop task in random order. Trials in the traditional Stroop task were presented in two blocks (neutral and incongruous), the order of which was randomised across participants. Trials were randomised within blocks. There were seven different image types in the pictorial modified Stroop task: adult males, adult females, male children (Tanner stages 1 & 2), female children (Tanner stages 1 & 2), male children (Tanner stages 3 & 4), female children (Tanner stages 3 & 4) and large cats. Images were presented in clusters of 12 images and there were two clusters of each image type.

PPG measurement was conducted separately using the Monarch 21 assessment system. Arousal was measured as the highest cm change in penile circumference during the presentation of images from a given stimulus category segment in conjunction with audio descriptions of sexual activity. Responses were recorded for segments involving infant, pre-school, grammar school, teen and adult stimuli/scenarios. Each of these scenarios had a version involving male and female victims/partners. With the exception of the infant category there were scenarios involving a coercive and a persuasive sexual encounter for all age groups. Therefore there were 18 raw responses for each participant.

4.2.3 Results

4.2.3.1 Characteristics of the rapists and child molesters

There were no significant differences between the two groups in terms of Full Scale IQ, age or number of years since last offence.

4.2.3.2 Results of the traditional Stroop task

A 2x2 mixed factorial ANOVA was carried out comparing the results of the neutral and incongruous trials in the traditional Stroop task for both rapists and child molesters. Mean reaction times were used with outliers that were more than three times the interquartile range beyond the 25th and 75th percentiles removed. The ANOVA produced a significant main effect of trial type; $F(1, 38) = 6.232$, $p = .017$, partial $\eta^2 = .141$. There was no significant interaction. However there was a significant between groups effect demonstrating that rapists were significantly slower overall in their response times, $F(1, 38) = 192.994$, $p < .001$, partial $\eta^2 = .835$. Ipsative z-scores were then calculated for each trial type by subtracting the overall mean response time to trials in the traditional Stroop task from the mean response time for each trial type and dividing by the overall standard deviation[29]. Repeating the above ANOVA using these z-scores instead of means, removed this between groups effect entirely along with making the interaction even smaller and the main effect of trial type more significant.

4.2.3.3 Results of the pictorial modified Stroop task

The data set did not contain information regarding the participants' adult sexual preference or a breakdown of the number of male and female victims in the case where the individual had offended against both males and females[30]. This limited some of the analyses that were possible. The best approximation of sexual preference available was that offenders with only male victims were likely to be homosexually orientated and those with only female victims were likely to be heterosexually orientated. However, it was possible the some of the offenders would not be orientated towards adults at all. Therefore, and to mirror the calculation of the PPG gender preference index, average response times to male and female images, regardless of age, were calculated (using the ipsative z-score method) and a 2x2 ANOVA was carried out comparing these for offenders with either male only or female only victims. Those with male only victims numbered only five and were therefore potentially problematic for inclusion in the

---

[29] Means and standard deviations calculated after removing outliers based on 3 times the interquartile range
[30] This information will be collected and verified by polygraph at a much later date as part of a larger research project

ANOVA (there were 20 participants who offended exclusively against females). Despite this, there was no significant main effect of gender of image but there was a significant disordinal interaction of gender of image and gender of victim, $F(1, 23) =$ 6.162, $p = .021$, partial $\eta^2 = .211$. Offenders against males had longer reaction times to male stimuli than female stimuli and the reverse pattern was true for offenders against females. Using ROC analysis, a difference score based on these mean values for males and females were able to significantly discriminate participants who had offended against males from those who had offended against females, AUC = .81, $p= .035$, *SE* =.102.

Similarly to the method adopted above, mean scores for young children (Tanner 1& 2), older/teenaged children (Tanner 3&4) and adults were calculated. A 3x2 mixed factorial ANOVA looking at the three ages of stimulus and the two offender types (child molesters and rapists) revealed a main effect of age depicted, $F(2, 76) = 4.324$, $p = .017$, partial $\eta^2 = .102$, but no interaction of offence type and age depicted. However, in this case getting an average of all child times etc. may not make sense, since a high (and potentially deviant) *z*-score towards child females may be nullified by a low score towards child males and vice versa. This approach was based on the method of calculating the PPG child preference index, which is therefore open to the same criticism. As an alternative, instead of averaging within age groups, a new variable was calculated that recorded the highest mean reaction time (regardless of gender) to each of the three age groups for each participant. The ANOVA carried out on these values again showed a main effect of age depicted, $F(2, 76) = 4.385$, $p = .016$, partial $\eta^2 = .103$, but no interaction of age depicted and offender classification, $F(2, 76) = .311$, $p = .734$, partial $\eta^2 = .008$. Figure 4.5 plots this result.
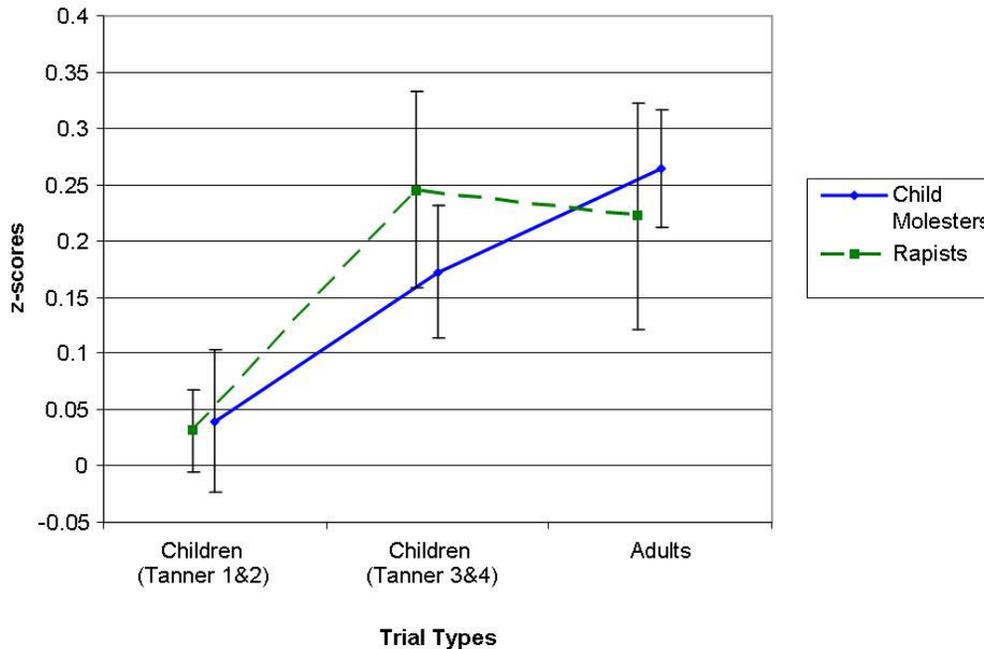
Figure 4.5 Mean reaction times, converted to *z*-scores, of the slowest response form each participant to child and adult images in the pictorial modified Stroop task (error bars indicate one standard error).

4.2.3.4 Results of the PPG

As is normal practice, the results of the PPG recording were converted to ipsative *z*-scores by subtracting the overall mean arousal value from the value for each stimulus category and dividing by the standard deviation. This was then, initially, used to calculate three indices based on these values: the difference between mean responses to male and female stimuli, regardless of age or level of coercion (referred to here as the gender preference index); the difference between mean responses to coercive and persuasive stimuli, regardless of gender and age (coercion index) and the difference between mean responses to adult and child (excluding teen) stimuli, regardless of level of coercion and gender (deviance index).

A 2x2 mixed factorial ANOVA investigated mean arousal to male and female stimuli and gender of victim, comparing participants with male only and female only victims. There was a significant main effect of stimulus gender, $F (1, 23) = 8.727$, $p = .007$, partial $\eta^2 = .275$, and also an interaction of stimulus gender and victim gender, $F (1, 23) = 8.147$, $p = .009$, partial $\eta^2 = .262$. As would be expected offenders with female

134

victims demonstrated higher mean arousal to female stimuli than to males and the opposite was true for offenders with male victims. As mentioned with the pictorial modified Stroop task, the number of participants with male only victims was low, so the result should be interpreted with caution. However the difference score calculated based on this mean response to all female and male (i.e. the gender preference index) scenarios yielded an ROC that demonstrated a good ability to discriminate between offenders with male and female victims; AUC = .82, p= .03, SE =.118. This difference score is the same as the gender preference index that is used clinically in the SRSTC.

A 2x2 mixed factorial ANOVA was carried out where mean arousal to child (excluding teen) and adult stimuli were compared for rapists and child molesters. There was no main effect of stimulus type and the interaction between stimulus type and offender type was non-significant. Additionally, the difference score/index calculated from these two scores (referred to as the deviance index) did not significantly discriminate between child molesters and rapists using ROC analysis, AUC = .618, p= .206, SE =.091. However as pointed out with the Stroop task an approach that averages arousal to scenarios depicting males and females, and across age categories, runs the risk of nullifying a deviant result by averaging with a non-deviant result. Figure 4.6 shows the highest arousal result in each age category, averaged across participants for rapists and child molesters. Both sets of participants show a similar pattern of responding across categories.

A 2x5 mixed factorial ANOVA comparing offender type across each of the five stimulus age categories confirmed that there was a main effect of age category, F (3.054, 112.998[31]) = 14.088, $p < .001$, partial η^2 = .276, but no interaction of category and offender type. The clear difference that can be seen on the graph between the average arousal of the two groups towards the preschool category is significant, t (37) = 2.275, $p = .029$, $r = .35$, two-tailed.

---

[31] Degrees of freedom were adjusted using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .764$) since Mauchly's test of sphericity indicated that sphericity could not be assumed, χ^2 (9) = 20.669, $p = .014$.
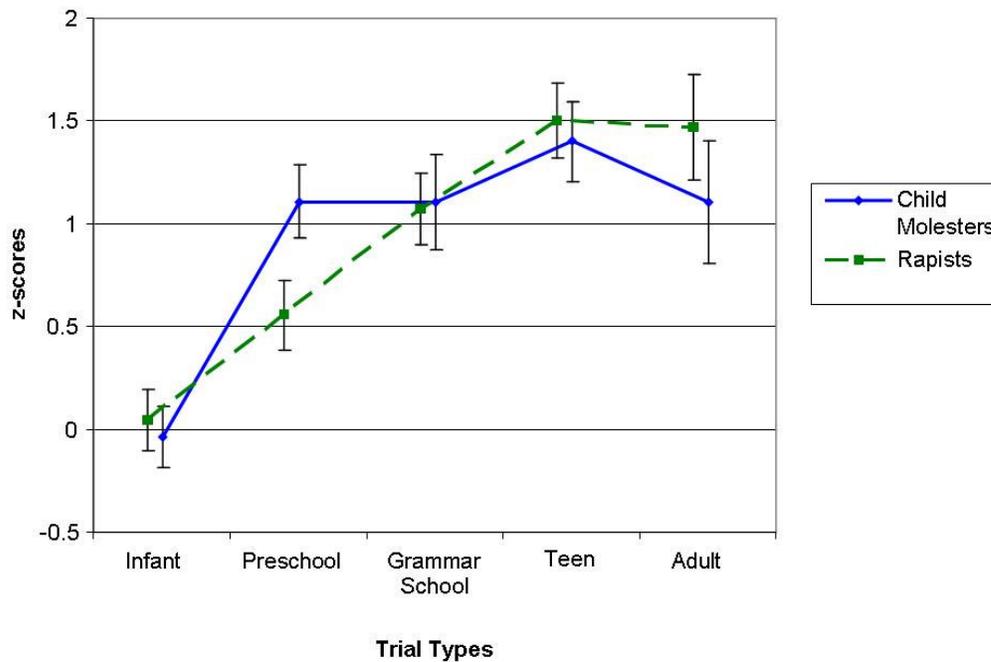
Figure 4.6: The highest recorded arousal for all scenario types and both genders in each of the age categories averaged across rapists and child molesters (error bars indicate one standard error).

A between groups *t*-test was performed comparing the mean score for child molesters and rapists on the coercion index. There was no significant difference between the mean values for both groups, $t(37) = .737$, $p = .466$, $r = .12$, two-tailed.

Taken together the results above suggest that the PPG does not yield particularly useful information about the sample of offenders studied. PPG results do, like the pictorial Stroop task, seem to indicate gender preference of participants, but do not seem to identify group differences in age preference or arousal to coercive scenarios. However, PPG results, especially those obtained using circumferential devices, like that used here, have been found to be unreliable at the early stages of the erectile response (Kalmus & Beech, 2005). Therefore, a separate 2x5 mixed factorial ANOVA was carried out again comparing offender type across each of the five stimulus age categories but only for participants whose raw circumferential change for the their three highest arousal responses had a mean of over .675cm change to penis circumference. This cut-off point is considered appropriate by the manufacturers of the PPG equipment to be considered genuine or 'clinically significant' arousal (P. Byrne, personal communication, September 16, 2009). The new analysis included 15 child molesters and 14 rapists.

This ANOVA yielded a main effect of stimulus age category; F (2.991, 80.762) = 13.293, p < .001, partial η^2 = .33 and also an interaction of stimulus age category and offender type that was approaching significance; F (2.991, 80.762) = 2.663, p = .054[32], partial η^2 = .09. The ANOVA is plotted in Figure 4.7. The graph clearly indicates that rapists and molesters have almost identical mean responses for the infant, grammar school and teen age categories. Independent samples t-tests reveal that the difference between arousal to adult stimuli is significantly greater among rapists; t (27) = -2.128, p = .043, r = .37. Molesters have higher mean arousal to preschool stimuli than rapists, but this difference is approaching significance only; t (27) = 1.899, p = .068, r = .34.
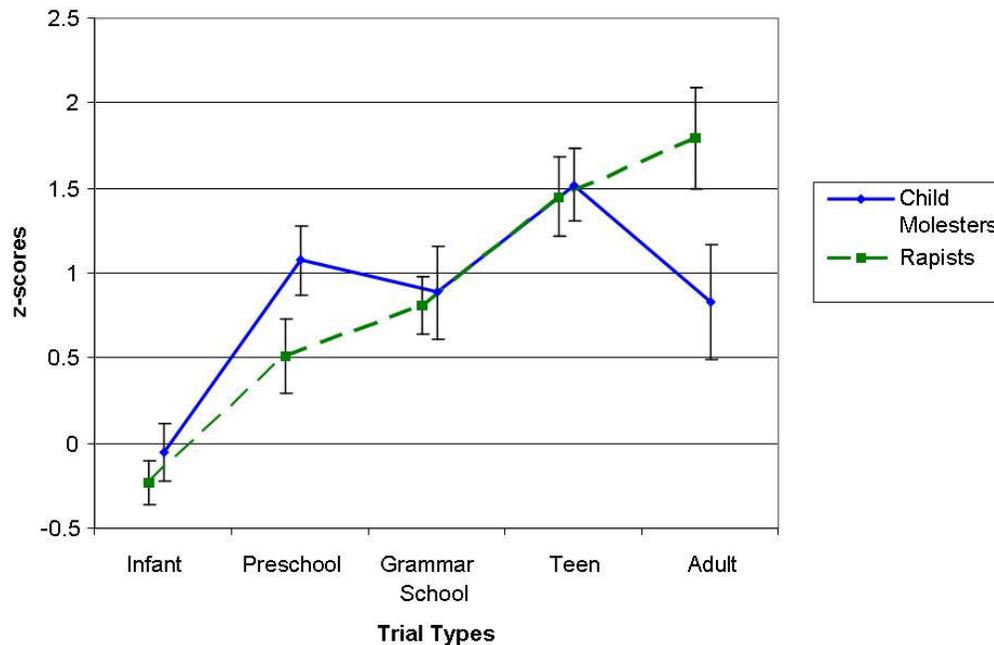
Figure 4.7: The highest recorded arousal for all scenario types and both genders in each of the age categories averaged across rapists and child molesters for offenders demonstrating clinically significant arousal (error bars indicate one standard error).

Calculating an ROC curve to explore how well the PPG results as measured by the deviance index for those with clinically significant arousal differentiates between child molesters and rapists yielded an AUC of .71 which was approaching significance; p =

---

[32] As with all other ANOVAs in this thesis where the assumption of sphericity was unsafe, the degrees of freedom for the main effect and interaction were adjusted using the Greenhouse-Geisser estimates of sphericity. In all other cases this did not effect whether a result was found to be significant or not. Using the Greenhouse-Geisser correction (ε = .748) in this instance did result in a non-significant interaction, where the uncorrected results had indicated a significant result (p = .036). Field (2009) argues that the Greenhouse-Geisser correction is too conservative, whereas the alternative Huynh-Feldt correction (giving p = .044) is too liberal. An aggregate of the two values would yield a p value below .05. The interaction was therefore treated as significant.

.055, *SE* = .097. This represented a better classification than that reported earlier (AUC = .618, *p*= .206, *SE* =.091) using all participants but still represented a discrimination that Tape (1999) would classify as fair. As already mentioned the deviance index is calculated by averaging responses to child stimuli and subtracting the average arousal to adult stimuli from that value. The difference in arousal patterns identified by the above interaction between stimulus age and offence type (graphed in Figure 4.7) would suggest that a significantly lower arousal to adults contributes in a large way to the observed differences between child molesters and rapists. Put another way there is not clear evidence in this sample of offenders for deviant arousal among child molesters. Rather there may be an absence of appropriate arousal.

4.2.3.5   Relationship between PPG and the pictorial modified Stroop task

As can be seen from Figure 4.5 and Figure 4.6 both the pictorial modified Stroop task results and the PPG results show similar patterns across age categories. They also show that there is not much difference between rapists and child molesters in their patterns of responding. On the other hand, Figure 4.7 includes only participants where there was a change in penile circumference greater that .675cm on average across their three strongest responses and indicates group differences between child molesters and rapists in responses to adult stimuli and potentially to preschool aged children.

When the difference between average response to males and females is calculated for both pictorial Stroop and PPG measures, they show a similar ability to classify offender by gender of victim (for offenders with victims of only one gender). These two difference scores are significantly correlated with each other, *r* = .359, *p* = .025, two-tailed. A logistic regression revealed that the combination of both the Stroop gender difference score and the corresponding PPG difference score yielded very little improvement in classification relative to the classification ability of either one on their own. Indeed the logistic regression indicated both together explained 52.4% of the variance in the gender of victim variable (in cases where there were victims of only one gender) but that 35.2% of this was overlapping between the two tasks. PPG accounted for 13.4% unique variance, while Stroop accounted for 14.8%.

While the above results are quite positive and suggest that the Stroop is tapping into sexual interest as measured by PPG to an extent, the two measures did not significantly correlate when the age categories of stimulus in the PPG and in the Stroop task were compared with one another. Neither was there a correlation between the deviance index calculated from the PPG and a similar index from the pictorial Stroop results calculated by subtracting the mean reaction time to adult stimuli (converted to z-score) from the mean response to child stimuli. Using all PPG responses or just those that had exceeded the 'clinically significant' threshold produced insignificant correlations with the pictorial Stroop task. Neither the pictorial Stroop task nor the PPG task reliably discriminated between rapists and child molesters for this sample of participants. The PPG did however seem to have more potential than the pictorial Stroop task given that (again for participants exceeding the clinically significant threshold for arousal) an ROC curve using deviance index scores had a fair (though non-significant) ability to classify participants as child molesters or rapists and there was a significant interaction of age of stimulus and offender type. However taken as a whole, the results of both measures suggest that both offender groups had similar arousal or attention to the different age categories of child (with the possible exception of pre-school children in the PPG measure). The PPG results did indicate group differences in responses towards adults; a pattern which was not identified by the pictorial modified Stroop task.

4.2.4   Discussion

Results indicated that both PPG and the pictorial modified Stroop task were able to discriminate between offenders based on their gender of victim with a balance between sensitivity and specificity that could be considered good. In addition both measures correlated moderately when it came to the measurement of response to either gender. These results were in line with the study's hypotheses. Both measures also showed responses indicative of a higher arousal to stimuli depicting older individuals compared with younger individuals. Neither measure seemed to indicate a significant difference between child molesters and rapists in their arousal to individuals of different ages. When participants with low arousal levels were removed from the PPG analysis the rapists and child molesters differed in their responses to adult stimuli and possibly to preschool aged stimuli. Neither task was able to significantly predict group membership (rapist versus child molester), though the PPG deviance index had a 'fair', but non-

significant, ability to discriminate between the groups. Therefore, the hypotheses predicting both measures' ability to identify group differences between child molesters and rapists received only partial support. The pictorial Stroop task and the PPG results did not correlate when deviance indices were calculated for both or when responses within individual age categories were compared.

The results of both the PPG and the pictorial modified Stroop seemed to indicate that the sample of rapist and child molesters were, on average, quite similar in terms of their sexual interests despite their disparate offences. Both seemed to indicate that mean responses were consistent with individuals with a sexual interest in adults or teenage children across both groups. On an individual level, responses to both measures varied quite considerably, but did not seem to coincide with group membership. This was unexpected, not only because of the nature and severity of the offences involved but also as all molesters had been given a diagnosis of paedophilia according to the DSM criteria (American Psychiatric Association, 2000). Duration since last offence, which was approximately 20 years on average (and no less than nine years) may have had an impact on the true extent to which deviant interest still existed within this sample. PPG results for those with a 'clinically significant' level of arousal demonstrated that child molesters had lower arousal to adult stimuli but had similar responses to rapists to all other age of stimuli with the possible exception of preschool aged children.

Regardless of the efficacy of the PPG measure in this study in determining group differences, it is clear that there was some relationship between the pictorial Stroop and the PPG results in the case of gender preference but very little identifiable relationship between the measures in terms of sexual interest in children. This could be explained in several ways. First, the PPG was unable to accurately measure sexual interest and therefore was unlikely to correlate strongly with the pictorial Stroop task. Second, the Stroop task may function quite well in determining sexual orientation towards adults, as demonstrated by the previous studies, but that sexual interest in children is a different phenomenon that is not as discernible through a paradigm such as the pictorial Stroop task. Third, it may be that the methodology of the Stroop task is not sufficiently refined to accurately maximise the distraction caused by the child stimuli that is indicative of a sexual interest, while minimising the distraction caused for other reasons. Fourth, the PPG stimuli separated images into age categories that corresponded to infant, Tanner

140

stage 1, Tanner stage 4 and adult while combining Tanner stages 2 and 3. On the other hand the Stroop Design combined Tanner stages 1 and 2 and stages 3 and 4. This lack of a coherent approach between the two tasks could also have impacted on the relationship between the measures. Finally, the absence of data regarding the participants' sexual orientation and predominant gender of victim where both have been abused impacted on the sophistication of the analyses conducted and hypotheses tested. Further testing and refinement of the pictorial modified Stroop task is therefore required before its clinical utility, or lack thereof, can be conclusively demonstrated.

4.3   Chapter Conclusions

Taken together the results of studies 3 and 4, looking at the clinical utility of the implicit tasks, are mixed. All seem to be tapping into sexual interests and/or associations approximately as hypothesised, but are confounded by other factors to the point that they would not be clinically useable in their present forms. Changes in methodology in the case of the Stroop task may have weakened its ability to find an effect. The IATs may have been more influenced by order than in study 2, possibly due to the impact of cognitive speed/ability in the offending sample and in the more heterogeneous control sample used for study 3. The limited ability of the PPG measure to demonstrate deviant sexual interest among child molesters in study 4 made it difficult to draw any firm conclusions about the potential use of the Stroop task to do likewise. The remaining two chapters of this thesis will focus on exploring methods of both maximising the effectiveness of the IAT and pictorial Stroop paradigms and exploring further what exactly both types of task are measuring.

Chapter 5: A comparison of three different pictorial modified Stroop task methodologies

The results of study 2 (chapter 3) indicated the blocked pictorial modified Stroop task used in that study was capable of accessing sexual interest. In addition it demonstrated a clear ability to discriminate between participants based on their sexual orientation. However, the results also indicated that the results of the blocked pictorial Stroop could potentially be confounded by order of block presentation. A primary goal of this dissertation is to explore the utility of tasks such as the pictorial Stroop task in the individual assessment of men who have sexually offended against children. To that end it was decided to modify the design of the pictorial Stroop task slightly when preparing the task for use with offenders against children. It was hypothesised that a design based on multiple clusters of similar trials as opposed to blocks containing all similar trials might be less influenced by order but at the same time able to recreate the same 'slow' Stroop effects found in the blocked design. Unfortunately the decision to adopt this approach had to be a pragmatic one rather than one based on experimental findings due to time constraints involved in devising a finished protocol for use with the clinical sample. As apparent from Chapter 4 the results of study 3 were difficult to interpret for many reasons, mostly involving the heterogeneity of the sample. However a lack of confidence in the clustered design's ability to tap into sexual interest as well as the blocked design had, was also an issue in the study.

5.1 Study 5: a comparison of blocked, clustered and random stimuli presentation in the pictorial modified Stroop task for sexually salient stimuli.

5.1.1 Introduction

While unfortunately conducted 'after the fact', the current study was designed to directly compare the blocked and clustered designed with non-offending controls. In addition, the study afforded the opportunity to further explore what was being measured by the pictorial modified Stroop task for sexually salient stimuli. The inclusion of a third pictorial Stroop design, where stimuli were randomly presented instead of in

blocks or clusters allowed the investigation of the presence of 'fast' effects as part of this particular Stroop phenomenon. As discussed in Chapters 1, 2 and 3, fast effects in a pictorial Stroop task refer to the immediate impact of the presentation of a stimulus on the participant's ability to name the target colour. Slow effects can refer to carry-over effects as a result of this fast effect or to the influence of higher order rumination and the activating of stimulus-related cognitions. If the pictorial Stroop effect found in study 2 is due to fast effects only then there will be no difference between difference scores (i.e. the difference between response times for male and female images) across all three pictorial Stroop designs. If carry-over effects or rumination are driving the effect then both the clustered and blocked designs should elicit more of an effect than the random design. Finally, if rumination is the main component of the effect, the blocked design should demonstrate a larger effect than either of the other two designs. It has been suggested by the author in Chapter 1 that the pictorial Stroop taps into two components of Spiering et al's (2004) information processing model of sexual arousal, that of the pre-attentive encoding stage and also the later attentive stage. It is therefore hypothesised that the pictorial Stroop for sexually salient stimuli effect includes fast effects corresponding to the pre-attentive stage of arousal and slow effects in the form of rumination corresponding to the attentive stage. Furthermore, it is possible that additional carry-over effects are a necessary bye-product of fast effects. Thus it is hypothesised that all three pictorial Stroop designs will show increased reaction times to female images relative to male images (since all participants in the current study are straight) but with the random pictorial Stroop showing less of an effect than the other two designs. The difference between the clustered and blocked designs in terms is hypothesised to be a product of the contribution of rumination effects to the overall effect.

5.1.2   Method

5.1.2.1   Design

A mixed factorial design was adopted where three groups of participants each completed a different version of the pictorial modified Stroop task. There were three trial types contained in each of the pictorial Stroop versions, images of female, images of males and images of large cats. In the first version of the task, the blocked pictorial

Stroop design, all images belonging to one trial type were presented together. The second version, the clustered design, had images of each trial type clustered together in smaller blocks of 16 images, with three such clusters of each trial type. The third version of the task, the random design, presented all trial types randomly.

5.1.2.2 Participants

Participants were recruited through university notice-boards along with university emailing. Recruitment materials called for heterosexual males only to take part. The majority of participants were students recruited through the School of Psychology and participated in the study for course credit. Some participants, not taking part for course credit, were given €10 to compensate them for their time. One participant, in addition to stating a strong sexual interest in females, indicated some sexual interest in males. His results were excluded from the final analysis. Participants were randomly allocated to the three experimental conditions, with 11 participants in each condition. All participants had an education level of leaving cert or above. The mean age of participants for the total sample was 25.61 ($SD = 6.874$). The mean age for the blocked condition was 23.64 ($SD = 3.295$), for the clustered condition was 30 ($SD = 9.317$) and for the random condition was 23.18 ($SD = 4.644$).

5.1.2.3 Apparatus/Materials

Computerised tasks were presented on a Gateway Pentium III computer with a Gateway EV9108 19" CRT monitor. Participant responses were made via a Cedrus response pad (model RB-620) with four coloured buttons (red, green blue and yellow). Tasks were run using the SuperLab4© stimulus presentation software. A 4-item questionnaire (see appendix 7) was also administered asking the participants about age, education, colour blindness and sexual orientation. Stimuli included images of male and female adults in bathing suits along with control images of large cats. As already mentioned in chapter 2 some of the adults were taken from the NRP image set (Laws & Gress, 2004) with the balance being created by the author.

5.1.2.4 Procedure

Responses were given via the response pad. All instructions for the task were presented on the computer screen. First, participants completed a practice trial where words were presented in the different colours. Feedback was given in the form of a red X appearing if the wrong colour was selected. Participants could not continue to the next word until they gave the correct answer (this was to ensure that participants properly understood the procedure; no feedback was given on the subsequent sections and participants had no opportunity to correct mistakes). Second, participants completed a traditional Stroop task with congruent, control and incongruent font-colour to word pairings. Procedure up to this point was identical to that in studies 2 (chapter 3) and 3 (chapter 4). Third, participants were presented with the modified pictorial Stroop condition. The design of this condition varied for the three groups of participant. Stimulus types were either presented in one large block (containing 48 images) in three smaller clusters (each containing 16 images) or were presented randomly. Images were of adult males, adult females, and large cats (controls). Chapter 2 describes the creation of these stimuli. As with the traditional Stroop task, participants had to identify, using a button box, the colour with which each image had been tinted. For the blocked design block order was randomised across participants and trials were randomised within blocks. For the clustered design order of clusters was randomised with the constraint that a cluster from each trial type had to be presented before the second cluster of any type was presented and so on. Response time was recorded for every trial along with whether that response was correct or not. When they had finished with the computerised tasks participants filled in the questionnaire.

5.1.3 Results

5.1.3.1 Results of the traditional Stroop task and a possible difference in ages of participants

A 3x3 mixed factorial ANOVA was carried out comparing the response times to the control, congruous and incongruous traditional Stroop trial types across each of the

three experimental groups (blocked, random and clustered[33]). Outliers over 3 times the interquartile range beyond the 25th and 75th percentiles were removed. A significant main effect of trial type was found as expected, F (1.542, 46.254[34]) = 20.442, $p < .001$, partial η² = .405. Importantly, there was no interaction effect indicating that participants in the three experimental groups did not differ in their patterns of responding.

A one-way ANOVA comparing ages across the three experimental groups indicated that there was a significant difference of age, F (2, 30) = 4.022, $p = .028$, partial $\eta^2 =$ .211 with participants in the clustered design group being significantly older than those in the random group ($p = .043$) and with the difference in age between participants in the clustered and blocked conditions approaching significance ($p = .062$)[35]. It is worth re-stating that there was no significant difference between the three groups on their performance on the traditional Stroop task.

5.1.3.2 Results of the pictorial modified Stroop tasks

A 3x2 mixed factorial ANOVA was carried out with response times concerted to z-scores for male and female stimuli as the repeated-measures factor and pictorial Stroop presentation group (blocked, clustered or random) as the between groups factor. The z-score method was used since it was suggested in study 2 (chapter 3) that it could moderate the effects of cognitive speed. There was a significant main effect of gender of stimulus (trial type), F (1, 30) = 32.971, $p < .001$, partial η² = .524. There was also a significant interaction of presentation group and trial type, F (2, 30) = 7.455, $p = .002$, partial η² = .332. This interaction is graphed in Figure 5.1 below. In order to explore the source of this interaction, three repeated-measures t-tests were carried out comparing male and female stimuli separately for each of the three pictorial Stroop presentation types. There was a significant difference (critical α set at .0166 using a Bonferroni conversion) between male and female stimuli when participants carried out the blocked

---

[33] It should be noted that all participants were presented with an identical version of the traditional Stroop task where all stimuli were randomly presented. The three experimental groups only differed in the design of the pictorial modified Stroop tasks that they subsequently carried out.
[34] Degrees of freedom were adjusted using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .771$) since Mauchly's test of sphericity indicated that sphericity could not be assumed, χ^2 (2) = 10.227, $p = .006$.

[35] Post-hoc testing carried out using Tukey's HSD method.

pictorial Stroop, $t(10) = 5.117$, $p < .001$, $r = .85$, two-tailed. The difference between the stimuli was also significant with the clustered pictorial Stroop design, $t(10) = 2.952$, $p = .014$, $r = .68$, two-tailed. There was no significant difference between male and female stimuli when participants carried out the random pictorial Stroop design, $t(10) = 1.009$, $p = .337$, $r = .3$, two-tailed. In addition, the two more standard measures of calculating a difference score were also calculated. The same pattern of results (as the 3x2 mixed factorial ANOVA) was found whether the reaction times were measured using the z-score method or by using means with outliers removed that were either 1.5 times (conservative method) or 3 times (less conservative method) the interquartile range beyond the mean.
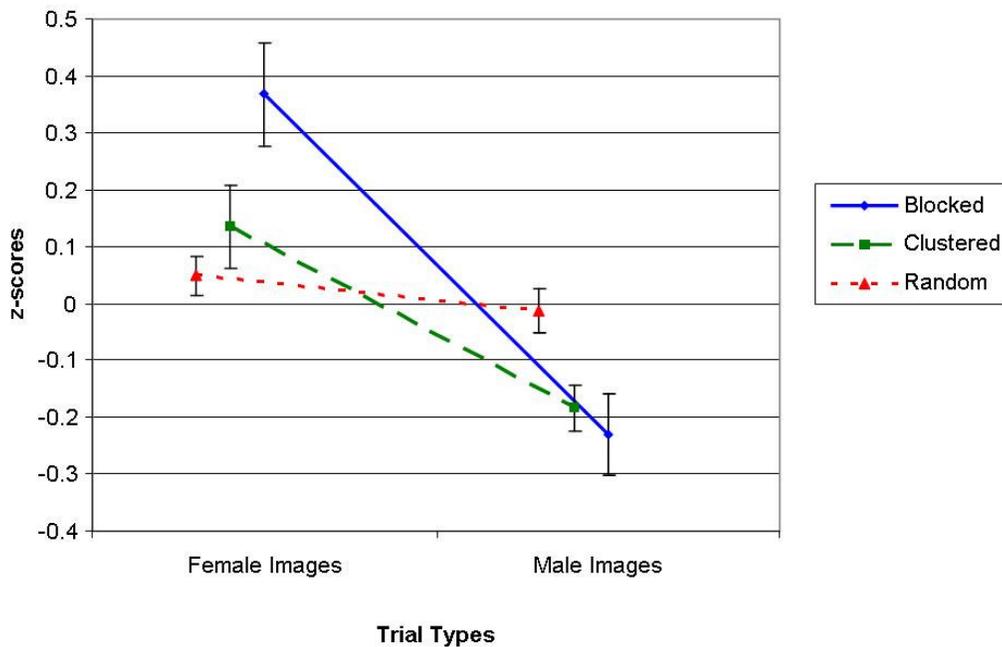
Figure 5.1: Reaction times (as ipsative z-scores) across trial types for three presentation groups (blocked, clustered and random). Error bars indicate one standard error.

As with the previous studies in this thesis, a difference score was calculated based on subtracting response times to female images from response times to male images. Response times converted to ipsative z-scores was initially used but ones based on means were also compared. A one-way ANOVA was carried out comparing the difference scores between male and female stimuli across each of the three pictorial Stroop presentation types (blocked, clustered, random). In line with the result of the factorial ANOVA conducted above, the result indicated a significant difference between

the groups, F (2, 30) = 7.455, p = .002, partial η^2 = .332. Post hoc tests revealed a significant difference between the difference scores for the blocked design and the random design groups only. Using the Bonferroni method of reducing familywise error yielded a significant p-value for this comparison of .002. If using mean reaction times instead of ipsative z-scores to calculate difference scores the ANOVA was still significant. However the difference between the blocked design and clustered design was now also significant (p = .039, less conservative method of outlier removal) or approaching significance (p = .056 conservative method). Table 6 shows the mean difference scores for the three different methods and for the three pictorial Stroop designs.

Table 6: Means of difference scores across pictorial Stroop presentation types (standard deviations in brackets)

Difference between male and female stimuli based on:	Type of pictorial Stroop Presentation		
	Blocked	**Clustered**	**Random**
Mean (outliers removed based on 1.5x IQR)	122.5ms (105.6)	42.9ms (44.8)	20.2ms (69.1)
Mean (outliers removed based on 3x IQR)	127.8ms (94.5)	45.4ms (52.9)	24.5ms (70.3)
z-scores (i.e. number of ipsative standard deviations from an ipsative mean)	.6 SDs (.39)	.32 SDs (.36)	.06 SDs (.2)

A one-way ANOVA, again using the ipsative z-score method of scoring, found no difference between response times to large cats across all three pictorial Stroop designs, F (2, 30) = 2.095, p = .141, partial η^2 = .122. Both alternative methods of calculation showed the same pattern of result, with smaller effect sizes in both cases.

Study 2 (chapter 3) in this dissertation indicated that a difference score of zero between mean response times for male and female images (using the less conservative method of outlier removal) was the best discriminator between gay and straight participants. When this cut point was applied to the current data, all participants in the blocked pictorial Stroop design were classified as straight, all but two of the clustered design participants

are classified as straight and all but four of the random participants were classified as straight. Again using a cut-off point of zero but using the difference scores calculated from means using the more conservative method of outlier removal, three random, three clustered and one blocked design participants were classified as gay. Given that all participants had identified themselves as straight, these results indicated that the blocked design was the most accurate at discriminating sexual orientation. When data from the one participant that had a negative difference score in the blocked design using the conservative method of outlier removal (which was also the lowest score for this group using the less conservative score) it was found that this participant was the only one who had been presented with trials in the order most likely to confound the results: male images first, followed by cat with females at the end. Interestingly the blocked design participant with the highest difference score had the opposite order of presentation, which may have therefore exaggerated any differences. The two clustered participants who demonstrated longer reaction times to males compared with females both showed this pattern pretty consistently (one participant had the same pattern if the difference score was computed separately for each block of clusters while the second had the pattern for two of the three blocks of clusters).

5.1.3.3   Graphical representations of the pictorial Stroop data: frequency distributions and order of presentation plots
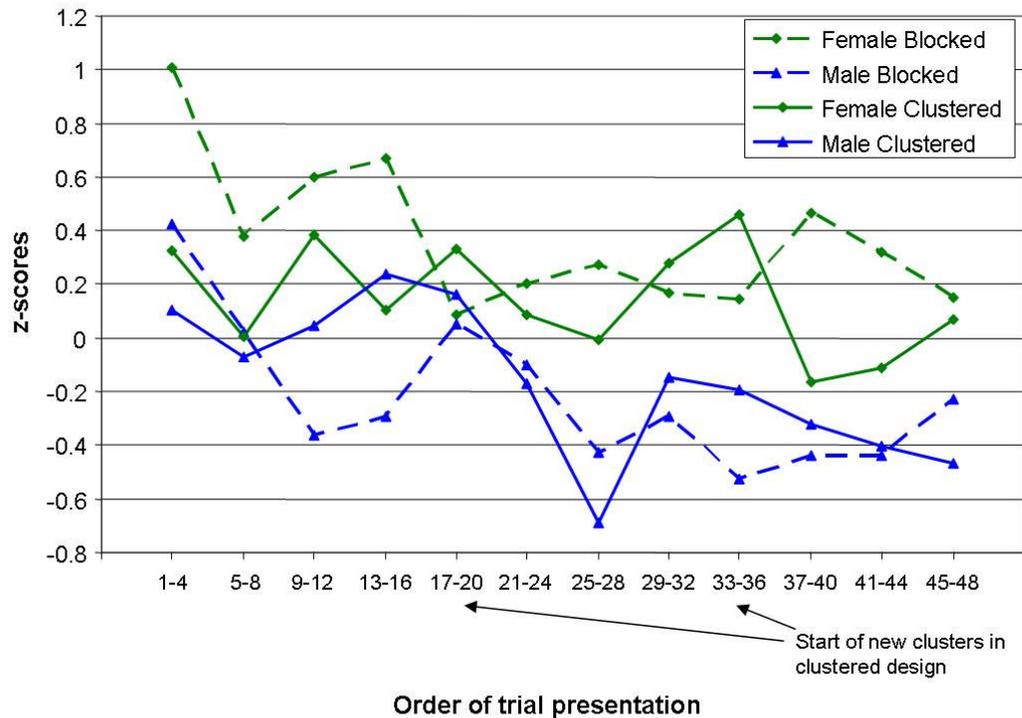
Figure 5.2: Plot depicting average reaction times by order of trial for Male and Female stimuli and for blocked and clustered stimulus presentation. Individual reaction times were converted to *z*-scores and then averaged across participants. Means for four trials at a time are presented

Responses to female and male images were converted to ipsative z-scores and recorded in order of presentation within blocks. These responses were then averaged across participants and further averaged over four positions in the order of progression. These were plotted as can be seen in Figure 5.2. Subjectively, the blocked design seems to maintain a clearer average difference between male and female responses than the clustered design. The clustered designs may show peaks corresponding to the start of the new clusters (more apparent at the start of the third cluster). The peak at trial 37-40 for female stimuli in the blocked design could indicate a rumination effect but is difficult to interpret given the variability that exists across the graph.

150
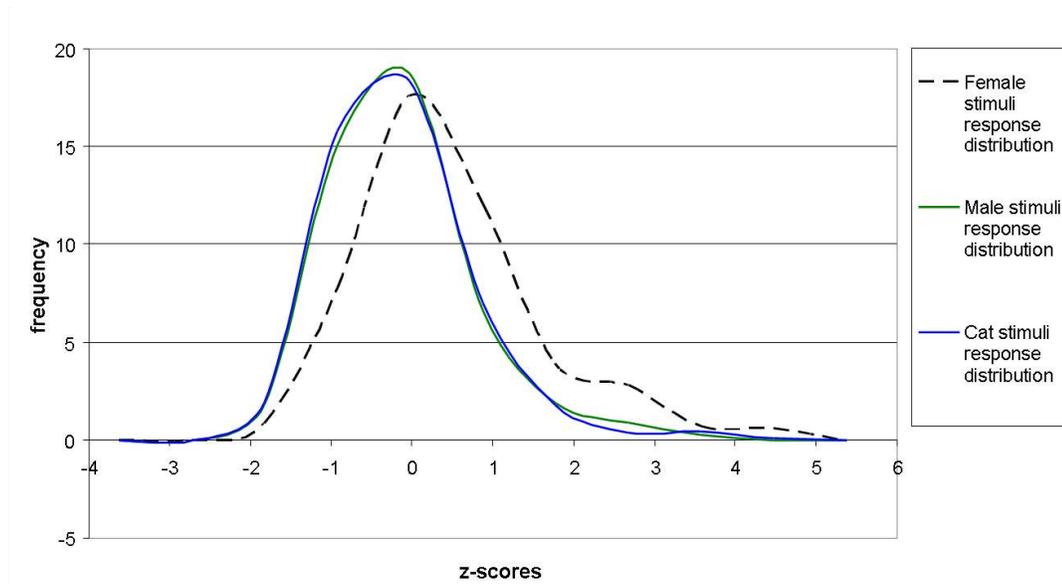
Figure 5.3: Averaged frequency distributions of blocked pictorial Stroop presentation z-scores for different trial types

Figure 5.3 plots the distributions for the three trial types (female, male, cat) produced by the blocked pictorial Stroop design. Ipsative *z*-score were calculated for responses to each stimulus and then averaged across participants. Figure 5.4 was calculated in the same way and shows the distributions of the three trial types for the clustered pictorial Stroop design. The distribution of female images in the blocked pictorial Stroop design (Figure 5.3) represents a noticeable shift from the cat and male images, which map onto each other almost exactly. As well as a shift there seems to be some changes in the shape of the distribution as well with potentially more observations in the tau (τ) region of the distribution.
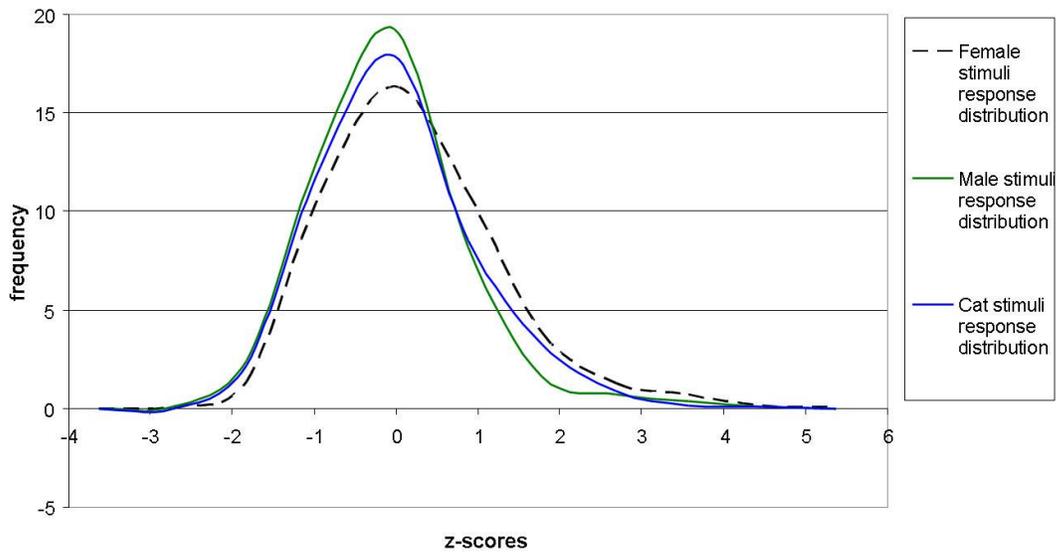
Figure 5.4: Averaged frequency distributions of clustered pictorial Stroop presentation z-scores for different trial types

While potentially different, the distribution of female images for the clustered pictorial Stroop (Figure 5.4) does not represent as clear a difference from the clustered male and cat images as is found in the blocked design. The distributions of the three trial types for the random pictorial Stroop design (Figure 5.5) does not appear very different from that of the clustered design. Again, the response times for the male and cat images are very similarly distributed (even more-so than in the clustered design) and the female images seem to follow a slightly differently shaped distribution.

152
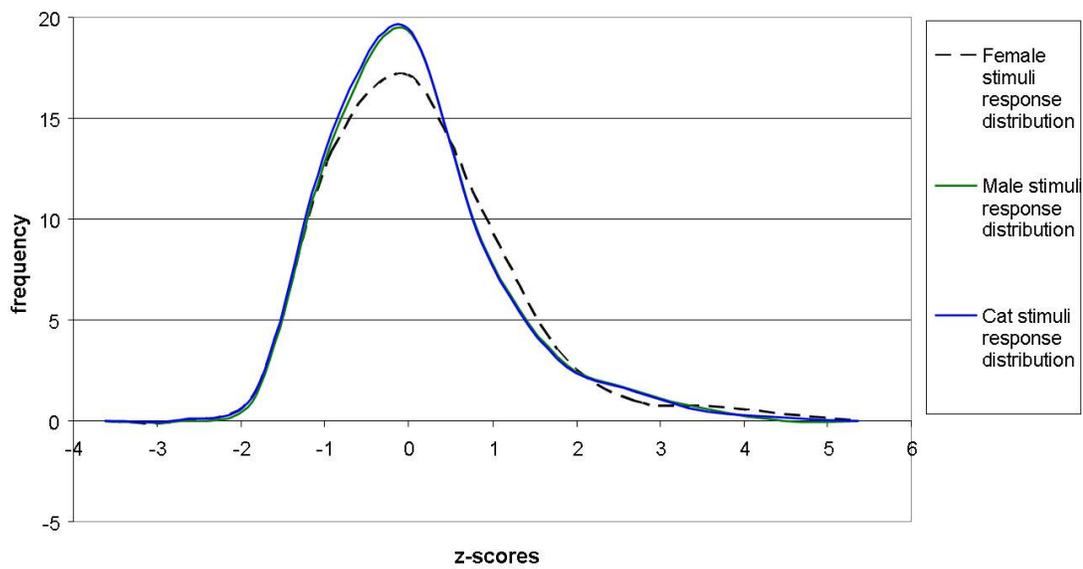
Figure 5.5: Averaged frequency distributions of random pictorial Stroop presentation z-scores for different trial types

5.1.4 Discussion

Results clearly indicated the blocked pictorial Stroop design outperformed both other designs in terms of the magnitude of the effect and the categorisation[36] of participants. The clustered pictorial Stroop design also demonstrated a significant difference between reaction times to male and to female images. However this difference was not as large as that of the blocked pictorial Stroop design. In addition, several of the clustered results yielded difference scores that were negative, indicating a bias towards males, despite the fact that all participants had identified their sexual interest as exclusively heterosexual. Random stimulus presentation faired the worst of the three approaches; though even with this approach the majority of difference scores were positive (suggesting a bias towards female images). The hypothesis that the three designs would demonstrate significant differences between response times to male and female images was therefore supported only in the case of the blocked and clustered designs.

Taken as a whole the results of the study indicates that the size of the clusters used in the clustered design were not sufficient to produce as large an effect as the blocked

---

[36] Though the study does not categorise participants so much as indicate whether their differences in response times between male and female stimuli are consistent with their sexual orientation.

design. While there was some indication that the blocked design had the potential to exaggerate or minimise the sexual salience effect as a result of order it seems that attempting to reduce the effect of order (i.e. using the clustered design) also reduced the strength of the salience effect. Attempting to eradicate the order effect entirely (i.e. the random pictorial Stroop) further reduced the salience effect to the point that the difference between male and female images no longer reached statistical significance. While it cannot be discounted that some of the participants for whom the clustered (and also the random) pictorial Stroop design indicated a bias towards male images may have a sexual interest in males[37], it is more likely that the results were as a result of a lack of precision in the task.

The fact that the blocked design so clearly outperforms the other two designs has interesting implications for the understanding of what the pictorial modified Stroop task is actually measuring. It is likely that any differences in reaction times between image sets is a product of several factors. For the sake of clarity this discussion groups these factors together into two categories. The first category is variability of response time due to sexual salience, which is the variability that we are interested in measuring. The second category is the remaining factors that might influence variability of response. These nuisance factors could include attention, motivation and the salience of individual images not related to sexual interest among other things. Theoretically this variability of responses should not be systematically related to the gender of the image so would therefore simply add noise to the data. A real effect of variability of response time due to sexual salience would need to exceed the nuisance variability in order to produce a noticeable difference between response times for males versus females or vice versa. The lack of a significant difference between reaction times to males and females in the random pictorial Stroop task indicates that either there was no effect of sexual salience variability or that this effect was too small to be distinguishable from the effect of nuisance variability. However, given that most of the differences for the random pictorial Stroop were in the direction expected it seems likely that there was some, albeit small, effect of sexual salience. Given that any carryover effect or rumination effect is randomly distributed across trial types in random presentation an effect of sexual salience that may exist must be as a result of fast effects. By extension the ability

---

[37] See section on volunteer bias in chapter 2

of both the clustered and the blocked design to identify significant differences between male and female trials indicates, not only that both were able to create a sexual salience effect that was strong enough to overpower any nuisance effects, but that there must be carry-over or rumination effects involved in causing the sexual salience effect. If the effect was due to carry over effects only (i.e. an attentional drain due to the salience of a stimuli that spills onto subsequent stimuli) there should not be a great difference between the effect observed in a blocked versus a clustered design. However, the blocked design yielded an effect that explained at least 25% more of the variance in the difference between male and female reaction times in that sample than the clustered design did in the other sample of participants. This sizeable difference suggests that carry-over effects cannot account alone for the improvement of blocked/clustered designs over random design. A rumination effect would explain the difference between blocked and clustered designs since a greater effect of rumination would be more likely to occur over a longer presentation of one trial type (i.e. a 48 image block compared with a 16 image cluster). It is therefore sensible to conclude that a modified pictorial Stroop effect for sexually salient material consists of a higher order-rumination effect (slow) and quite possibly both effects due to an immediate capture of attention by the stimulus presented (fast) and also carry-over effects due to that capture (slow). This is in line with the theory proposed in Chapter 1 that the pictorial Stroop task is operating on two levels of Spiering et al.'s (2004) information processing model of sexual arousal where fast effects would relate to encoding (as would slow, carry-over effects) and rumination effects would be related to the attentive stage of the process.

While any interpretation of the graphical representations of data presented in the above results section must be made cautiously it seems that the distribution and order plots both support and conflict with the conclusions made above. As with the histogram that compared reaction times (expressed as ipsative z-scores) to orientation-consistent, age-appropriate stimuli with other stimuli in Study 2, the histogram comparing reaction times to male, female and cat stimuli for the blocked pictorial Stroop suggested a difference in distribution shape, and not just means, between responses to females and the other two trial types. While conclusions are subjective since, there may (again, as with Study 2) be an increase in the tau (τ) region of the distribution, which was tenuously linked to rumination effects in the discussion of Chapter 3. A comparison between the graphs of the three trial type distributions for the blocked pictorial Stroop

(Figure 5.3) and the clustered pictorial Stroop (Figure 5.4) reveals that, while the distribution of the female images may be different to the cat and male images, that difference is not as stark as that found for the blocked pictorial Stroop (in terms of mean or shape). The distribution of the three trial types in the random design (Figure 5.5) is almost identical to the clustered design but again suggests a slightly different distribution for female images compared with the other two image types. It must be noted however that a subjective inspection of the plot that graphs averaged reaction times (again represented as ipsative z-score) by trial types across order of presentation shows no steady increase of response times as order increases for female trials in the blocked condition. One might expect this if a rumination effect took the form of a cumulative interference with the colour-naming task. An increase in response times towards the end of presentation of blocked female stimuli could be indicative of a rumination effect. However, given that clustered and blocked designs should perform similarly over the first 16 trials but that this is not indicated by the graph it is difficult to interpret whether fluctuations on the graph are indicative of differences due to task design or due to individual differences. A similar graph using many more participants could shed light on this issue.

As already mentioned, the findings of this study suggest that a large part of the pictorial Stroop effect for sexually salient stimuli is due to slow effects. The discussion also concludes, very cautiously, that fast effects are also likely to be present. The presence of fast and carry over effects might be further clarified if future studies included a version of the pictorial Stroop task where stimuli were presented randomly but followed by one or several 'filler' stimuli to identify carry over effects and to prevent those carry over effects from confounding the results on different trial types. More participants would have increased the power of this study to potentially find significant effects in the random pictorial Stroop design. However using the current design, and if the sexual salience is as small as was found in the current study (i.e. an effect size of $r = .3$) a follow up study would need 45 participants in the random pictorial Stroop condition alone to find a significant effect. Not including gay participants in this particular study was a pragmatic decision taken in order to ensure that available gay participants could be recruited to studies where sexual orientation was more centrally related to the research question. However, this decision came at the expense of being able to use ROC curves or Logistic Regression to quantify the ability of the three tasks to classify

participants correctly. Future studies should consider including participants of different sexual orientations.

Chapter 6: Comparing a pictorial modified Stroop task, a pictorial Implicit Association Test and a Choice Reaction Time task

Studies 2-5 have consistently found that the pictorial Stroop task has the ability to tap into sexual interest. Studies 2 and 5 indicate that a blocked design yields the largest effect of Stroop and is taken as evidence that the pictorial modified Stroop effect for sexually salient stimuli is, in the main, due to higher order rumination of the stimulus category and possibly the activation of associated concepts. It is hypothesised in this thesis that this higher order rumination and activation is linked with the attentive stage of the arousal process described by Spiering et al. The evidence for the effectiveness of the Implicit Association Task in tapping into sexual interest/associations is a little more mixed from the previous studies. While it is clear that, at least in some cases, it is indirectly measuring sexual orientation, it is not wholly clear whether the lack of a consistently strong overlap between this measure and the pictorial Stroop task indicates a degree of task error or the fact that the underlying sexual construct that is being measured is different to that measured by the Stroop task. It is quite likely that both explanations are valid. It is suggested here that the pictorial Stroop task indirectly measures sexual orientation or sexual interest by measuring arousal related attention while the IAT measures schema related to sexuality that in most cases is an indirect measure of orientation or interest. Therefore, the results of a Gender-Sex IAT in a population where there should be no discrepancy between a participant's sexual arousal and their sexual schema should match closely the results of a pictorial modified Stroop task. It seems safe to assume that much of a lack of agreement between the measures, especially when using a non-offending population, is due to task error in one or both tasks.

The final study in this thesis set out to see whether changing the design of the IAT could bring the pictorial Stroop results and the IAT more in line with each other. Exploring this has two purposes. First, establishing the best method of carrying out these tasks is a primary goal of this thesis. Second, developing the methodology for both so that they do co-vary strongly in situations where, on the face of it, they should

(i.e. when participants' sexual arousal and schema should be in line with each other and their reported sexual orientation) will allow researchers in future studies to be confident that a divergence of IAT and pictorial Stroop results indicates a divergence of the underlying constructs being measured and not task error. In addition the final study set out to explore the relationship between the choice reaction time (CRT) paradigm, used by several previous studies, and the pictorial modified Stroop task and the IAT. Before carrying out the final study, it was necessary to create some new stimuli for the tasks.

6.1 The development of new stimuli for the new tasks

6.1.1 Rationale

Each of the pictorial Stroop tasks outlined in the previous three chapters of this thesis used identical stimuli (though study 5 only used the adult images from the set). The process of creating these stimuli is outlined in Chapter 2. Chapter 2 also explains how some of the images were created by the author and some were taken from the Not Real People (NRP) stimulus set created by Laws and Gress (2004). The first experimental study (study 1, chapter 3) found that all five image types yielded very similar response times when they were scrambled so that their content was no longer recognisable. However, that study did not address whether the content of these images varied between stimulus types in terms of attractiveness which may have impacted on their ability to elicit a content induced delay. For example, the both studies using offenders (chapter 4) might have identified clearer differences between offenders and controls if stimuli were more potent in their ability to induce an attentional-based reaction delay. To explore this issue at the time would have been problematic since the child stimuli would have also needed to be rated for attractiveness (most likely by sexual offenders with an admitted sexual interest in children). Given the difficulties in recruiting offenders that were faced, it was not practical to do this. While Laws and Gress (2004) did have their images evaluated in order to accurately categorise them into Tanner stages they did not have them rated for attractiveness. Given that all the images used in the previous studies were morphs it was also deemed important to establish how realistic the images were, since a lack or 'realness' could potentially contribute to a delay in responding.

For the final experiment in the thesis it was decided to compare a pictorial Stroop task, a Choice Reaction Time Task and a pictorial IAT. These tasks would use adult stimuli only and would require the creation of new morphed images of adults. It was decided to create a new set of images to combine with the existing set and to have the entire series of adult images rated by participants on three dimensions:

1) attractiveness of the image
2) realness of the image
3) age of the person depicted

From this rating process the 21 most attractive male and female images were chosen for use in the final study. Six images of each gender would be used for the pictorial IAT and the remaining images would be used for both the modified Stroop task and the CRT task.

6.1.2   Method

6.1.2.1   Design

The evaluation of images was carried out using an exploratory survey where participants rated images based on attractiveness, realness and age.

6.1.2.2   Participants

Potential participants were informed of the survey through the use of college emails, college electronic notice boards, posting on online forums and by snowballing methods using both emails and social networking websites. Participation in the survey was voluntary and participants received no payment/reward for completing it. There were 522 completed or incomplete responses to the online study. Since order of images presented for rating was random, data from both completed and partial responses were used. However, given that some participants may have started the study a second time after having only partially completed a previous effort, partial responses were removed if they originated from the same internet protocol (IP) address as a complete response and both had the same age, gender and sexual orientation indicated. Some responses were also removed where the participant had answered demographic questions but had not completed any of the image ratings. In total 482 valid responses were retained. Those 482 participants had an age range of 17-63 years, with a mean age 28.17 years

(*SD* = 7 years). The gender breakdown of the sample was 50.8% female and 49.2% male. The largest proportion of the sample was made up of straight women (43.4%), followed by straight men (24.5%) and gay men (22.2%). Lesbians, bisexual women and bisexual men made up 4.8%, 2.7% and 2.5% of the sample respectively.

6.1.2.3 Apparatus/Materials

There were 24 images of males and 24 images of females used in the online survey. Half of the images were those that had been used in the original series of studies. As has already been mentioned in chapter 2, six of these original male images and three of the female images came from the NRP image set. The remainder were created by the author. The 24 new images were also created by the author in the same way as the original set had been (See chapter 2, section 2.2). The images were posted on an online survey website called www.surveygizmo.com.

6.1.2.4 Procedure

On accessing the online survey, participants were first presented with the following instructions:

"In this survey you will be asked to rate images of people in three ways. First, depending on your stated sexual preference, you may be asked how attractive you find the image. Second, you'll be asked what age you think the person in the image is and third, how real you think the image is. Some of the images have been created by morphing together images of different people so I would like you to rate how realistic you think the images are[38]. The survey has less than 50 images to rate and takes about 15 minutes to complete. Images are of male and female adults in bathing suits."

Participants were then asked to state their age, gender and sexual orientation. The options given for sexual orientation were Gay/Lesbian, Bisexual or Heterosexual. The survey was set up in such a way that participants were only asked to rate the attractiveness of images that were consistent with their sexual orientation. Participants

---

[38] In fact, all the images were created by morphing but it was felt that participants would be more

accurate in rating the realness of the images if they felt that at least some of the pictures could be real.

rated the realness and age of all images. Images were presented one at a time with two or three (attractiveness only being asked when image was consistent with orientation interest) questions presented below the image. When attractiveness was being rated, participants were asked the following question:

"On a scale of 1 to 10 (**1** being **very unattractive** and **10** being **very attractive**) how sexually attractive do you find the person in this image?"

Participants then selected the appropriate response. Participants rated the age of the person depicted in the image by answering the following question:

"What age would you say the person depicted is?"

From a drop down menu, respondents selected the age range the image belonged to. Options ranged from 14 years to 49 with each option spanning three years, e.g. the youngest range was 14-17[39] years and so on. Participants were also asked the following question:

"How much does the image look like a real person?"

Participants were presented with Likert-type responses with which to answer: definitely not a real person, probably not a real person, not sure, probably a real person, definitely a real person. Participants had to answer all questions before moving on to the next image. Order of images presentation was randomised.

6.1.3   Ratings produced

The average response of participants to each image was calculated for each sexual orientation group separately and in total. Given that the final experiment, as with the previous studies, will have male participants only the responses of gay men and straight were used to determine the 21 most attractive male and female images respectively. As mentioned, different numbers of participants rated each image but there was an average of 53.87 (SD = 2.05) gay raters for each male image and 71.42 (SD = 8.6) straight raters for each female image. Table 7 compares the attractiveness ratings of these 21 male and 21 female images. None of the 42 images selected for use in the final study had a mean evaluated age range that indicated participants typically saw them as less than 18 years

---

[39] All original materials used with which to create the images were of models over the age of 18.

However the younger category was included to see whether some of the completed morphs might

resemble people younger than that.

of age. In addition, while some of the images in the 42 were rated quite low on realness most had an average rating of either "not sure" or "probably a real person". The three images from each gender that were not included in the final 21 were all images that had been used in the original studies in this thesis. Additionally, All 8 adult images (3 female 5 male) that were taken from the NRP set were rated in the bottom five images for each gender. Only three NRP images made it into the final study.

Table 7: How gay men evaluated the attractiveness of the male images and straight men evaluated the attractiveness of the female images (values out of a maximum of 10 and scores are calculated for the 42 images to be used in the final study only)

	Males	Females
Mean attractiveness	6.14	7.03
Mean of most attractive image	7.98 (SD = 1.66)	8.35 (SD = 1.59)
Mean of least attractive image	3 (SD = 1.73)	5.36 (SD = 2.04)

6.2 Study 7: Comparison of pictorial modified Stroop, Pictorial IAT and choice reaction time tasks

6.2.1 Introduction

Snowden, Wichter and Gray (2008) found that a pictorial IAT design yielded promising results in discriminating between participants based on sexual orientation. Using images as stimuli to represent the categories of male and female isn't problematic when creating an IAT simply to discriminate between gay and straight participants. A pictorial design becomes somewhat problematic if you want to have an IAT that looks at the association of male or female and sex, regardless of the age of the males and females depicted or if you want to explore associations of children versus adults and sex but do not want to specify gender. However since the current experiment was testing non-offending participants only and attempting to maximise the agreement between the Stroop task and the IAT, it was decided to adopt a pictorial IAT design. In addition, the IATs used thus far in this thesis departed from the conventional IAT design by not having an opposite to sex category. It was decided to explore this issue by designing the current category IAT with *sexual* and *nonsexual* as the target categories. Creating an opposite to sex category is potentially problematic and *nonsexual* as a category label is

not ideal given that it refers to an absence rather than an opposite of sexual attributes. However, it was felt that such a label would still offer a greater contrast than the neutral *furniture* label had in the previous studies and, in addition, the words chosen for this category attempted to capture a certain amount of the opposite to sexual nuance (see section 6.2.2.3).

In addition to further exploring the relationship between the Stroop and IAT tasks, the current study set out to compare the results of both with a version of the Choice Reaction Time (CRT) task. The CRT task is an experimental design whereby the participant needs to identify as quickly as possible the location of a dot superimposed on an image (in this case of male, female or control images). This task has been used several times to attempt to measure sexual interest generally (Santtila et al., 2009; Wright & Adams, 1994, 1999) and also to attempt to measure sexual interest in offenders (Giotakos, 2005; Gress, 2008). Wright and Adams (1994) demonstrated that the CRT paradigm could be used to correctly classify gay and straight men and women. The vast majority of their participants demonstrated longer reaction times to nude images of their preferred gender. Wright and Adams (1994) replicated these findings, demonstrating that, using nude images, 85% of participants performed as expected. The 15% that did not perform as expected were all straight women. In the same study non-nude stimuli resulted in 75% of participants performing as expected. Santtila et al. (2009) also replicated the Wright and Adams studies but used sexually explicit and non-explicit versions of the same stimuli to demonstrate their effect. They found that participants typically took longer to identify the location of the dot on explicit versions of images of their preferred gender. This effect habituated over the course of the experiment but in the first phase of the experiment a difference score calculated by subtracting reaction times to sexually explicit male images from explicit female images yielded and AUC of .82 using ROC analysis. Stimulus types were presented randomly and not in blocks. Both studies by Wright and Adams (1994, 1999) also randomly presented their stimuli.

Gress (2008) did not find clear results that indicated her CRT was tapping into sexual interest. Methodological issues may have contributed to the inability to find a clear effect of sexual interest. Giotakos (2005) on the other hand found that extrafamilial offenders produced longer reaction times to images of female children than to female

adolescents and adults. Intrafamilial offenders viewed images of adolescent images longest and rapists had longer reaction times to adult females. Giotakos used a two dot only version of the CRT and had the images appear on the screen for a period of two seconds before the dot appeared on the image. Again, both the studies looking at offenders used a random presentation of images. Given that previous studies have used the CRT task to demonstrate a sexual content induced delay it was appropriate to compare this task with the modified pictorial Stroop task. Since the results of study 5 (chapter 5) demonstrated that a blocked Stroop design yielded the largest effect size and best classified participants in terms of their sexual orientation this approach was adopted with both the pictorial Stroop and the CRT used in this study. This departs from the previous CRT studies as all adopted a randomised design. The pictorial Stroop design used was similar to the previous experiments in the series, except that the newly rated images were included and that there were an additional three pictures of both genders included. The traditional Stroop task that was included before every other pictorial Stroop was excluded in this study as it was desirable to have the methodologies for the Stroop and CRT tasks as similar as possible.

It was hypothesised that the pictorial modified Stroop task and the Choice Reaction Time task would both show group differences in response times for gay and straight participants. It was further hypothesised that a difference score based on the male and female response times for both tasks would significantly predict sexual orientation and that results of both tasks would be correlated. It was hypothesised that the pictorial IAT would match the results of the pictorial Stroop task (and the CRT task) more closely than the previous IATs had.

6.2.2   Method

6.2.2.1   Design

A mixed factorial design was used with gay and straight participants completing all experimental tasks; the pictorial modified Stroop task, the pictorial IAT and the choice reaction time task.

6.2.2.2 Participants

Fourteen straight and eleven gay male participants took part in the study. Participants were recruited through a combination of college emails, college notice-boards, internet forums and snowballing techniques. Undergraduate students recruited through the School of Psychology, participated in the study for course credit. The remaining participants, not taking part for course credit, were given €10 to compensate them for their time. Straight participants had an average age of 21.57 (SD = 3.18) while for gay participants the average age was 28.27 (SD = 8.23).

6.2.2.3 Apparatus/Materials

Computerised tasks were presented on a Gateway Solo 9300 Laptop. Participant responses were made using a Cedrus response pad (model RB-620) and an Ergodex DX1 input system. The Cedrus response pad used four coloured buttons (red, green blue and yellow) and the Ergodex virtual keyboard had 5 buttons arranged in the shape of a quincunx (like the 5 on a die). Tasks were run using the SuperLab4© stimulus presentation software. A questionnaire was also administered to participants asking about sexual interest (see appendix 8).

Images for all tasks were of male and female adults and of large cats. All images of humans were morphed and were from the set of images created by the author or from the NRP image set (Laws & Gress, 2004). Following the results of the online survey outlined in the previous section only the 21 most attractive images form each gender were used. Six male and six female images were selected for use in the IAT. The remaining 15 male and 15 female images were used in both the pictorial modified Stroop and CRT tasks. To select the images for the IAT all male and female images were ranked by their attractiveness ratings. The 1st, 5th, 9th, 13th, 17th and 21st ranked image in each gender was selected for use in the IAT and the remainder were used in the pictorial Stroop and CRT tasks. For the pictorial modified Stroop task images (including large cat images) were coloured using the same method outlined in previous studies yielding four versions of each image, each in one of four colours: blue, green, red and yellow.

New IAT words were generated by compiling a list of previously used words (e.g. Banse et al., 2009; Gray et al., 2005; Mihailides et al., 2004; Nunes et al., 2007; Smith

166

& Waterman, 2004) and expanding on those lists using the authors own judgement and collaboration with other researchers (including Banse, Schmidt and Nunes) carrying out similar research. Words were then excluded from that list on the grounds of word length, frequency and also on *a priori* grounds. The sexual words used were: lust, lick, kiss, naked, orgasm, arouse and attractive. The nonsexual words were: ugly, cold, dull, avoid, bland, boring and unattractive.

For the CRT task a white dot was superimposed in each of five positions on the images, yielding five versions of each image; where the white dot was located either in the top right, top left, bottom right, bottom left or middle of the image. The white dot was 8 pixels (or approximately 6.5mm) in diameter which at a distance of 3 feet would yield a visual angle of approx .48° which would approximate the 1.5 inch dot used at a distance of 15 feet by Wright and Adams (1999). The outer 4 dots typically fell on the gray background and not on the body of the person/animal depicted. The central dot was always located on the body of the person/cat. In cases where the dot fell on a light-coloured area the shading of the image was sometimes altered slightly to make the dot more noticeable.

6.2.2.4   Procedure

Participants were allowed to seat themselves a comfortable distance from the monitor (approximately three feet) and responded during the task via the response pad or virtual keyboard which was placed on the desk a comfortable distance in front of them. Participants each completed three tasks, a pictorial modified Stroop task, a pictorial IAT and a CRT task. The IAT was always completed second with the order of the pictorial modified Stroop task and the CRT (i.e. being presented either first or last) randomised across participants. Participants used the Cedrus response pad for both the IAT and the pictorial modified Stroop Task. For the CRT task participants responded using the Ergodex virtual keyboard. Participants were instructed to use their dominant hand to press the five buttons on the virtual keyboard as required. For all tasks, participants were asked to try and respond as quickly as possible while trying to be as accurate as possible. They were told that it was normal to make a few mistakes since when performing a task quickly errors are likely. The instructions for all tasks were presented on the screen.

The procedure for the pictorial modified Stroop task was similar to that used in the previous Stroop tasks in this study. However, in order to be as methodologically similar to the CRT task, the traditional Stroop task was removed from the beginning of the task. As with the previous Stroop tasks participants had to identify the colour in which an image was presented by responding using one of the four coloured buttons on the response pad. Trails were presented in three blocks (cat, male and female) with order of blocks randomised across participants and order of trial randomised within each block. There were 15 images in each block presented in each of four colours, yielding a total of 60 images in each block. Before starting the blocks, participants completed a set of practice trials which consisted of one of each of the three image types presented in each of the four colours, yielding a total of 12 images. Order of these practice trials was also randomised. If participants responded incorrectly to a trial a red X appeared on the screen and the participant could not continue until they had corrected their mistake. The actual experimental trials did not include such error feedback.

In the CRT task participants had to identify as quickly as possible, using the Ergodex virtual keyboard, the location of a dot superimposed on each image. The same images were included in the practice trials and in the three blocks of trials (cat, male female) as had been used in the pictorial Stroop task. However, since there were now 5 versions of each image, corresponding to the five dot locations, instead of the four versions of the Stroop task, the practice trials now contained 15 trials and each block had 75 trails. The task was randomised in exactly the same way as the pictorial Stroop task.

For the pictorial IAT participants were instructed to use the two outer buttons (left and right) on the Cedrus response pad. Over seven blocks, participants first categorised images of man and women that appeared on the screen one-by-one into as *male* or *female* using the left and right buttons. Participants then classified words as being *sexual* or *nonsexual* words in the same way. Once this had been practiced, categories were combined so that categories were paired with each other. For example the participant had to push the left button if the word that appeared on the screen was either a sexual word or a female image and push the right button if the word was either a nonsexual word or a male image. In later trials this order was reversed, thus creating new pairings (females paired with nonsexual words, while sexual words and males were

168

paired). If participants answered incorrectly a red X appeared and they could only move on to the next word once they corrected their response. Time to initial response was recorded as well as time for each subsequent response until the correct answer was given.

6.2.3 Results

The difference in age between gay (Mean = 28.27, *SD* = 8.23) and straight (Mean = 21.57, *SD* = 3.18) participants was significant with gay participants being significantly older than the straight ones, *t* (23) = -2.807, *p* = .01, *r* = .51, two-tailed. While gay participants did also seem to have many more sexual partners on average than the straight participants (16.6 partners versus 7.6), this difference turned out to be non-significant, *p* = .138. However, a boxplot indicated that the number of sexual partners reported by two participants (one gay and one straight) were identified as extreme outliers. When these values were removed from the analysis the mean number of sexual partners for straight participants (M = 3.92, *SD* = 4.75) was significantly fewer than that of gay participants (M = 12.3, *SD* = 4.37), *t* (21) = -4.336, *p* < .001, *r* = .69, two-tailed.

6.2.3.1 *Results of the Choice Reaction Time task*

As with previous studies in this thesis using the pictorial modified Stroop task several methods were used to analyse the results of the CRT task. Four measures of central tendency were calculated: Mean reaction times, mean reaction times where outliers 1.5 times the interquartile range beyond the 25th and 75th percentiles were removed, mean reaction times where outliers three times the interquartile range beyond the 25th and 75th percentiles were removed, and medians. In addition ipsative *z*-scores were calculated, where mean responses to trial types are represented in number of standard deviations away from the overall mean for all trial types. Outliers were removed using three times the interquartile range when calculating means and standard deviations for these *z*-scores. Since the ipsative *z*-score method proved useful in previous studies it was used as the dependent variable for the analyses conducted here. Other methods are only reported if they deviate from findings based on the ipsative *z*-scores.

A 2x3 mixed factorial ANOVA was carried out on the results of the CRT task with sexual orientation as the between groups variable and the three trial types (male, female, cat) as the within-groups factor. A main effect of trial type was found, F (1.396, 32.112[40]) = 4.88, p = .024, partial η^2 = .175. Post-hoc tests using a Bonferroni correction demonstrated a significant difference between reaction times to male images and cat images, $p < .001$. There was no significant interaction between trial type and sexual orientation. This pattern of results was found regardless of which measure was used to analyse CRT response times. Figure 6.1 graphically depicts the pattern of results across trials and between sexual orientations. It confirms what was found by the statistical analysis: that there seemed to be differences in reaction times for different trials, but that these differences were unrelated to sexual orientation. It should be noted that, from the graph and as indicated by the non-significant interaction, the pattern of response to male and female stimuli by gay and straight participants seemed to match quite closely.
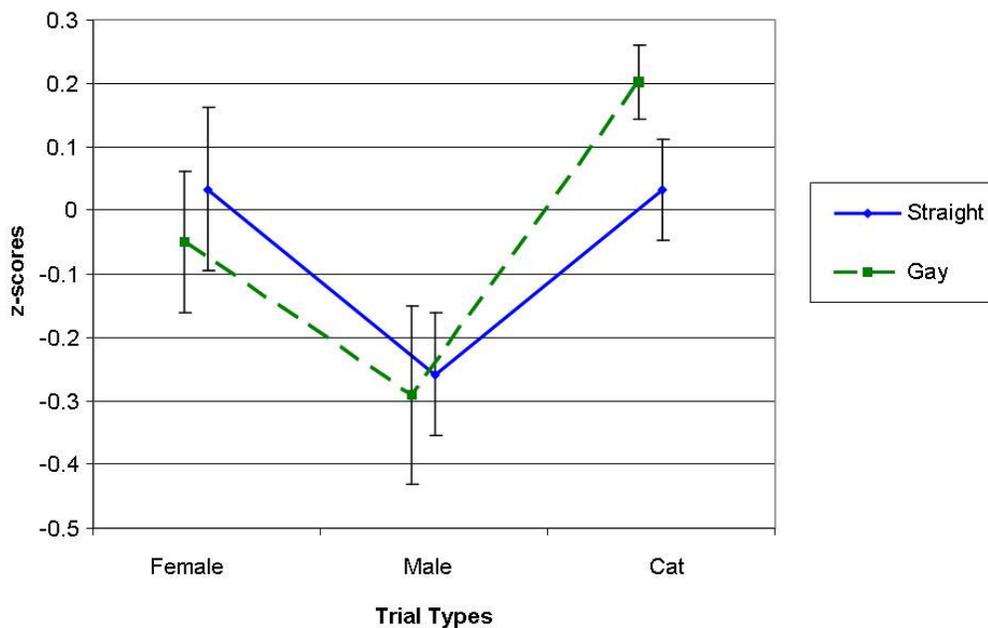
Figure 6.1: Ipsative *z*-scores of mean reaction times to trial types for gay and straight participants

Looking separately at CRT results depending on order of task presentation (i.e. whether the Stroop task of the CRT task was presented first) did not yield any interaction

---

[40] Degrees of freedom were adjusted using Greenhouse-Geisser estimates of sphericity (ε = .698) since Mauchly's test of sphericity indicated that sphericity could not be assumed, χ^2 (2) = 12.464, p = .002

between orientation and gender of trial. However, there was evidence for an influence of order of block presentation within the task. A one-way repeated measures ANOVA found a significant effect of block order, $F(2, 48) = 9.463$, $p < .001$, partial $\eta^2 = .283$. Reaction times were fastest for the first presented block (Mean z-score = .2073, $SD = .247$), then the second (Mean z-score = -.0754, $SD = .357$) and then the third (Mean z-score = -.2993, $SD = .419$). When post-hoc tests were carried out using the Bonferroni method, the response times to block 1 images were found to be significantly slower than both the block 2 ($p = .037$) and the block 3 ($p = .001$) images. Blocks 2 and 3 did not differ significantly ($p = .272$).

Given a large order effect it would be difficult to identify if sexual interest was also having an impact on response times. To attempt to test this, data for male and female images only were coded into a new within-subjects variable with the reaction time to whichever of those two trial types was presented first as one level and reaction time to the second presented trial as the second level. Participants were then divided into those that received orientation-consistent trials first and those who received them second. A 2x2 mixed factorial ANOVA was carried out to see whether response times across the first- and second presented non-control trials were moderated by whether participants were presented with orientation consistent or inconsistent trials first. Consistent with the previous finding of an order effect, there was a main effect of order. However there was also an interaction that was approaching significance, $F(1, 23) = 3.722$, $p = .066$, partial $\eta^2 = .139$. The direction of this interaction indicated that the order effect in the CRT task may be somewhat moderated by sexual interest. In other words, when participants responded to images in line with their sexual orientation first the response times were longer than participants who were responding to orientation inconsistent trials first. The opposite pattern was found for the second presented trials: while responses were quicker overall (indicated by the main effect of order) those who responded to orientation-consistent trials second had slower reaction times than those who were given orientation consistent trials second.

For the sake of completeness a difference score was calculated based on the ipsative z-scores for female and male images. An ROC curve was plotted which yielded an area under the curve (AUC) of .539 which represents an ability that is not significantly greater than chance to discriminate between gay and straight participants.

6.2.3.2 Results of the pictorial modified Stroop task

The same measures of central tendency and ipsative *z*-scores were calculated for the pictorial modified Stroop task as was for the CRT task. Again, the analyses reported use ipsative *z*-scores unless there was a discrepancy between the different measures. A 2x3 mixed factorial ANOVA was carried out where the results of gay and straight participants were compared across all three trial types (male, female and cat). There was no main effect of trial type. The interaction between trial type and orientation of participant was approaching significance, F (2, 46) = 2.956, p = .062, partial η^2 = .114. However, the same analysis using each of the other measures of central tendency found a statistically significant interaction, so it seems safe to assume that this lack of significance using the *z*-score method represented a type II error. This interaction was in the direction expected with longer response times among gay participants to male images and the opposite pattern among straight participants.

If order of task is taken into account, by carrying out two separate two-way ANOVAs, this interaction is only found when the pictorial Stroop task was presented last. Order of block presentation within the pictorial Stroop task does not have significant impact on response times, though response to the last presented block are typically lower than the other two. An ROC analysis revealed that a difference score calculated from the response times to male and female images from the pictorial modified Stroop task was able to discriminate between gay and straight participants at a level greater than chance, AUC = .779, p=.019, *SE*=.093. This AUC value is considerably lower than the AUC of .926 found in study 2 (chapter 3). However, when order of task is taken into account, the difference score when the pictorial Stroop task is completed first yields an ROC with poor discriminant ability, AUC = .571, p=.685, *SE*=.179, while the difference score when pictorial Stroop is completed last has an excellent discriminant ability, AUC = .905, p=.015, *SE*=.086.

6.2.3.3 Results of the Implicit Association Test

For the initial analysis, the same five approaches as for the pictorial Stroop and the CRT tasks were taken to treating the data from the IAT. The ipsative *z*-scores were calculated

based on the overall mean and standard deviation of the 'experimental' blocks of the IAT, i.e. blocks where two concepts were associated with each button (Blocks 3, 4, 6 and 7). Using the *z*-scores, a 2x2 mixed factorial ANOVA comparing combined reaction times for blocks 3 and 4 and blocks 6 and 7 across gay and straight participants found no significant main effect of trial type but a significant interaction of trial type and orientation, $F(1, 23) = 40.987$, $p < .001$, partial $\eta^2 = .641$. In the above analysis the first trial type consists of trials that are consistent with a heterosexual sexual orientation, i.e. trials where a word or image is presented that the participant must classify using either the *female/sexual* button or the *male/nonsexual* button. The second trial type consists of trials consistent with a homosexual sexual orientation, where the participant must classify words or images using the *male/sexual* or the *female/nonsexual* buttons. The interaction between these trial types is plotted in Figure 6.2. The pattern of no main effect and significant interaction was also found if responses to pictorial and word stimuli were analysed separately. This indicates that the IAT effect was present for both pictorial and word stimuli.
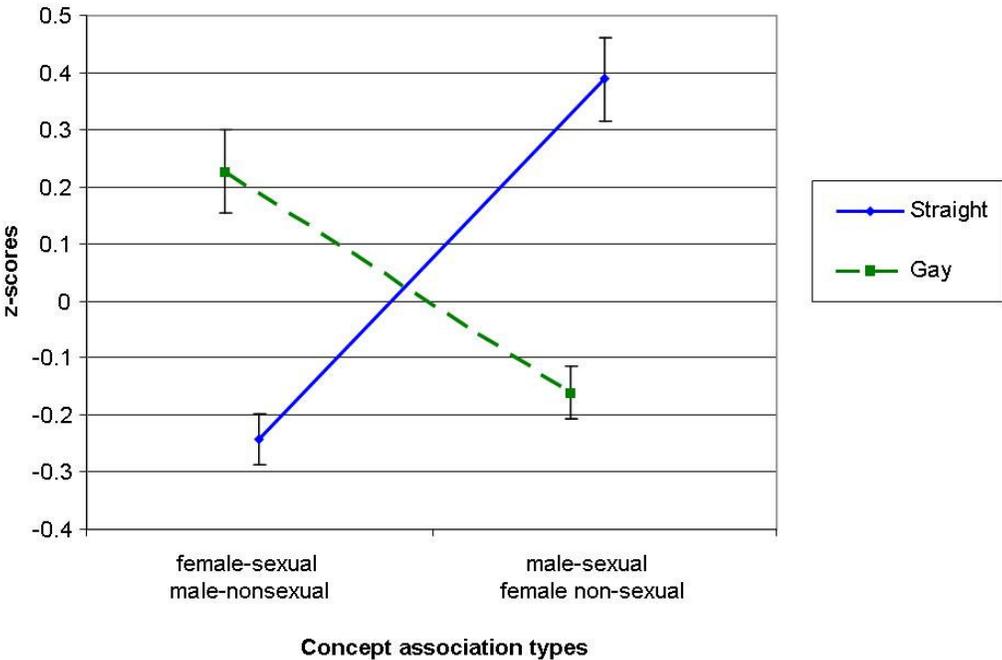

Figure 6.2 Means of reaction times converted to ipsative *z*-scores for the two concept pairing configurations for gay and straight participants in the pictorial Gender-Sex IAT (error bars indicate one standard error).

There were 14 possible methods available to calculate a difference score between the two trial types. Three were based on Greenwald et al's (2003) recommendations, six were versions of those algorithm's but with outliers removed based on 3 or 1.5 times the interquartile range, four were based on the measures of central tendency mentioned above and the remaining one was based on the ipsative z-score method of data treatment. Unlike previous studies in this thesis, each of the methods performed similarly when calculating a ROC curve of the accuracy of those difference scores in discriminating between gay and straight participants. Each difference score had an excellent ability to discriminate between the groups, with AUCs ranging from .916 to .994. All AUCs were significant at the .001 level and had standard errors ranging from .056 to .011. The best performing difference score was that which simply used the raw mean reaction times for both values, AUC = .994, $p < .001$, $SE = .011$. What this meant in terms of classification was that when the difference score was calculated based on the raw mean, yielding negative values for results consistent with a gay orientation and positive values for results consistent with a gay orientation, there was one straight participant (out of 14) with a negative value, and one gay participant (out of 11) with a positive value. Both these individuals produced the data points that were closest to zero on this difference score.

6.2.3.4   Relationship between the three tasks

There was no correlation between difference scores on the pictorial modified Stroop task and the CRT task. Further there was no relationship between response times to individual stimulus types from both tasks, i.e. there was no correlation between participants responses to male, female and cat images across both tasks. Stroop response times overall (Mean = 727.63, SD = 192.46) were slower than overall response times[41] for the CRT task (Mean = 599.39, SD = 215.6), t (24) = -5.986, $p < .001$, $r = .77$, two-tailed. Additionally the CRT difference score did not correlate with any of the IAT difference score measures. There was a significant medium-sized correlation between the difference scores based on z-scores for the Stroop task and IAT, $r = .448$, $p = .025$, two-tailed. The correlation between Stroop difference score and the IAT was significant

---

[41] Overall response times were the average mean reactions times to each of the three stimulus types with outliers removed based on 3 times the interquartile range. Method of outlier removal did not impact on significance of this statistic.

for 6 of the 14 IAT difference scores and over .3 in all but one case (in that case r = .292). Given that the results of the Stroop task had a better ability to discriminate between gay and straight participants when the Stroop task was carried out last, correlations between Stroop and IAT were compared separately for when Stroop was completed first or last. Correlations between IAT and Stroop were much higher, when Stroop was completed last (r between .348 and .597 depending on the IAT score used).

Two forced entry binary logistic regressions were carried out to assess whether a combination of both the IAT and Stroop scores yielded a better categorisation of participants as either gay or straight than either did individually. The z-score method was used for both[42]. For the first regression, the IAT difference score served as the predictor variable in the first block and the Stroop difference score was the predictor variable in the second block. The order was reversed in the second regression, in order to determine each variable's unique contribution to classification as well as the combined classification. Taken together both scores correctly classified the sexual orientation of all participants. Individually the IAT score outperformed the Stroop score in terms of categorisation (84% compared with 68%). Both scores explained large amounts of the variance. However the IAT model (R^2 = .83) explained considerably more variance then the Stroop (R^2 = .36). The combination of both predictor variables yielded a 17% increase in total variance explained to explain the total variance in the orientation variable (R^2 = 1). The IAT score accounted for 64% of unique variance while the Stroop accounted for 17%. Nineteen percent of the variance in categorisation was attributable to overlapping variance among the IAT and Stroop. The addition of the Stroop score to the regression model including the IAT score increased the classification rate by 16% which corresponded to the correct classification of two additional gay and two additional straight participants.

6.2.4 Discussion

Results indicated that orientation did not seem to have much of an impact of response times in the CRT task. There was no interaction between trial type and orientation and a

---

[42] While the difference score for the IAT that had the highest ROC value was the one using the raw mean scores it was decided not to use this value for this analysis as the possible presence of outliers in the raw mean results may have incorrectly inflated its discriminant ability due to chance.

difference score based on male and female images did not significantly predict sexual orientation. There was a large effect of order, which may have masked any impact of sexual interest. Indeed results suggested that sexual orientation may have moderated the impact of order. Sexual orientation/interest had a clearer effect on the pictorial modified Stroop task. There was an interaction between trial type and orientation and a difference score based on the Stroop results for male and female images demonstrated a fair ability to discriminate between gay and straight participants. However, this discriminant ability was not as accurate as previous Stroop studies in this thesis had been. Order of task presentation seemed to affect the strength of the modified Stroop effect, with the task being presented last seeming to produce more robust effects. The IAT used in this study demonstrated a clear interaction between trial type and sexual orientation. In addition the IAT difference score(s) had an excellent ability to discriminate straight and gay participants. The results of the IAT and Stroop tasks both correlated significantly and combining the two measures resulted in the correct classification of all participants in terms of sexual orientation. Neither the pictorial modified Stroop task nor the IAT correlated significantly with the CRT task.

Overall the results of the IAT used in this task were encouraging. The results of the Stroop task were also positive, though it was disappointing that the procedure did not have better discriminant ability on its own. It is very positive that both tasks combined were able to accurately classify all participants. The fact that the Stroop effect was strongest when Stroop was the last task presented is interesting. For the participants carrying out the task it is a little challenging at first to learn the location of each of the four coloured buttons so that they do not have to look down when responding. Given that the version of the task used in the current study did not contain a traditional Stroop task at the start meant that participants had less time than usual to learn the location of the buttons before starting on the trials of interest. It may be that after completing the CRT task, using the virtual keyboard, and the IAT, using the same button box as for the Stroop, that the participant presented with the Stroop last were quicker to master this and therefore had less nuisance variance in their data associated with the novelty of the task. However, there was no significant difference in overall response time between participants where the Stroop task was first versus last, so an alternative explanation may be necessary. An alternative may be that the preceding exposure to the CRT and IAT had 'primed' their sexual interest and that the process of sexual arousal was

therefore more readily accessible. The fact that a robust effect was found in study 2 when Stroop was presented first, albeit with a traditional Stroop task before it, seems to go against this theory. The phenomenon known as 'ego depletion' may offer a third explanation of the effect of task order (Baumeister, Bratslavsky, Muraven, & Tice, 1998). It may be that tasks demanding the use of cognitive control, such as the traditional Stroop and the IAT used here, may impact on the individual's ability to disinhibit the sexual response to the images, thereby maximising the Stroop effect when the participant is 'depleted' rather than fresh. It should be noted that order of task presentation did not improve performance on the CRT task when presented last.

There was no clear evidence that the CRT task was able to tap into sexual interest. A significant effect of order of block presentation may have masked any potential influence of sexual orientation. There was limited evidence that the size of the order effect may have been moderated somewhat by sexual interest. Given that past studies have found a CRT effect (Santtila et al., 2009; Wright & Adams, 1994, 1999), the results in this case were surprising. The design of the current study was unlike previous CRTs in that the trial types were presented randomly in the past. Therefore, order effects should not have caused any problems in those tasks. The rationale for blocking the stimuli in the current study was to test if the CRT task was measuring an analogous phenomenon to the pictorial modified Stroop task. The Stroop task produced significantly longer reaction times across trials than did the CRT task. This indicates that the Stroop is a cognitively more demanding task. This may reflect that the buttons on the pad for the CRT task correspond to the location of the dots on the screen and, therefore, it may be easier for the participant to respond. The Stroop task on the other hand has colours that are matched to the buttons in a pattern which the participant must learn. It may also be that the Stroop task itself is more difficult and therefore requires more attention, thus making it easier for those attentional resources to be captured by sexually salient images. A possible reason for this potential increased level of difficulty may be that the target stimulus (i.e. colour) in the Stroop task is fully integrated with the to-be-ignored stimulus (i.e. image of a human or cat) while, in the CRT task the white dots were, with the exception of the middle dot, located on the gray background surrounding the target image. Future studies could manipulate some of the variables mentioned above to explore further the link, or lack thereof, between the pictorial modified Stroop task and the CRT task for sexually salient stimuli.

As mentioned, previous randomised CRT tasks did find what the authors termed a sexual content induced delay. However, the same was not found in study 5 for a randomised pictorial modified Stroop task. In addition, in the current study the Stroop task was able to identify group differences based on sexual orientation but the CRT wasn't. These results taken together suggest that the pictorial modified Stroop Task and the CRT are not measuring the same processes. The distinction between fast and slow effects in the Stroop task has been made elsewhere in this thesis. It seems that the slow effects, that the author has argued drives the Stroop effect are not present (or are not as pronounced) in the CRT task. While a randomised CRT task and a blocked pictorial modified Stroop task are likely to be both tapping into cognitive processes surrounding sexual arousal, the actual processes they are measuring seem to be different. Alternatively they may both be tapping into several processes but that these processes are weighted differently in how the tasks are measuring them. It may be that the CRT design is more adept at measuring so-called 'fast' effects related to the process of sexual arousal and that the pictorial Stroop is better suited to capturing 'slow' effects. It should be noted that the CRT design adopted by Giotakos (2005), whereby stimuli were presented for two seconds before a dot appeared, is likely to be measuring slow effects as well.

The IAT in the current study performed better than the IATs used in the previous studies in this thesis. There were two differences between this IAT and the previous ones. First, the gender categories used images instead of names as the stimuli belonging to those categories. Second, there was an attempt at an opposite target category to the sexual category in the study, not just a neutral one. Both of these changes may have contributed to the improved classification of participants. Despite, this considerable improvement in classification, the task did not correctly classify participants as gay or straight in every case. Similarly the Stroop task had limited accuracy in classification, even in those cases where Stroop accuracy was improved by being presented after the CRT and IAT. Both measures were significantly correlated, the strength of this correlation improving when only the later-presented Stroops were included. A combination of both scores, using logistic regression, did correctly classify all participants. It is difficult to say whether both tasks simply contain a degree of error that is balanced by the inclusion of a second task as a 'double check'. Alternatively, both

178

measures could be accurately measuring distinct but related constructs or processes that both feed into reported sexual orientation. It is possible that both tasks are measuring distinct but related constructs/processes in that the IAT is a measure of sex-related schema and that the Stroop task is a measure of sexual arousal-related attention. In a non-offending sample of adult males, who have a clear understanding of their own sexual orientation, it is unlikely that there would be any disagreement between these constructs. Therefore in the current sample, assuming that none of the participants were confused or dishonest regarding their sexual orientation, both tasks should have performed as accurate indirect measures of sexual orientation. It is likely then, that nuisance variability arising from the task is sufficient in some cases to mask the variability due to sexual interest or associations.

In summary, the study found that the CRT task did not clearly tap into sexual interest using the blocked design adopted. This finding was in contrast with most of the previous studies that used the CRT. However, all previous studies had adopted randomised trial presentation. It is likely therefore that the influence of order found masked any potential effect of trial type. The CRT did not correlate with either the IAT or the pictorial modified Stroop task. The inclusion of pictorial stimuli and a somewhat opposite to *sexual* category in the Gender-Sex IAT improved that tasks performance relative to the previous Gender-Sex IATs in this series of studies. The exclusion of the traditional Stroop task at the start of the pictorial modified Stroop task seemed to weaken its ability to tap into sexual interest. This weakening was ameliorated by the presentation of the pictorial Stroop task after the IAT and CRT tasks. Both the IAT and the pictorial Stroop task were significantly correlated, with that correlation being stronger when the pictorial Stroop was presented last. Taken in combination the IAT and the pictorial Stroop tasks were able to correctly categorise participants by their sexual orientations 100% of the time. Future studies should attempt to compare the results of the pictorial Gender-Sex IAT with the pictorial modified Stroop task and a CRT task that adopts a random design.

Chapter 7: General Discussion

This thesis set out to explore the clinical utility of several implicit tasks in the assessment of sexual interest. In addition the research was concerned with the theoretical implications of the presence or absence of agreement between different implicit tasks. The thesis sought to address issues with regard to the pictorial modified Stroop task and the Implicit Association Test (IAT) under three broad and overlapping headings. First, the thesis sought to explore, from a methodological point of view if the Stroop and IAT tasks could tap into sexual interest in non-offending controls and in an offending population. Second, the thesis sought to address what the results of these tasks meant for a theoretical understanding of what each task was measuring and also whether the tasks might give some theoretical insight into offence-related sexual interest and cognition. Third, the thesis sought to comment on the practical utility of these tasks in the assessment of sexual offenders and in future research with that population. This chapter will first summarise the results of the separate studies conducted and then discuss those findings in terms of the three goals of the research, outlined above.

7.1 Summary of Findings

Study 1 (chapter 3) compared responses to scrambled versions of all the images used for the Stroop tasks in studies 2-5 (and some of those used in study 6). The results indicated that there was no significant differences between the scrambled versions of the images suggesting that any differences observed in later Stroop tasks were likely due to the influence of the content of those images.

Study 2 (chapter 3) piloted the pictorial modified Stroop procedure and two Implicit Association Tests, the Gender-Sex IAT and the Age-Sex IAT. The study found that both the pictorial Stroop and the Gender-Sex IAT significantly discriminated between gay and straight non-offending participants. Both also correlated with each other and produced greater accuracy in classifying participants by sexual orientation when used together. Results of the Child-Sex IAT, suggested that there was no difference between the strength of participants' child-sex and adult-sex associations. This may have indicated that the task was not performing as expected. Alternatively, the results may

have also indicated that individuals for whom sexuality and children are not problematic may simply not have strong schema relating to age and sex.

Study 3 (chapter 4) used a clustered version of the pictorial modified Stroop task and the two IATs from study 2 with a sample of offenders against children from the Granada Institute and Arbour Hill prison. The results of the study indicated that the pictorial modified Stroop task identified differences on a group level between the patterns of responding of those participants most likely to demonstrate a sexual interest in children (extrafamilial offenders and those admitting a sexual interest in children) and those least likely (intrafamilial offenders and control participants). While the results of the Child-Sex IAT tasks were interpreted as potentially suggestive of an influence of group (i.e. offender vs. control), the results were inconclusive, partially due to an influence of order of trial type presentation. This order effect, which was found for both IATs, had not been found in study 2.

Study 4 (chapter 4) explored the utility of a pictorial modified Stroop task and its relationship to penile plethysmography (PPG) results using a sample of child molesters and rapists from the Sand Ridge Secure Treatment Center (SRSTC). Results of the PPG measure in this study suggested that the child molesters and the rapists who participated in this study did not differ greatly in terms of their overall arousal response to underage stimuli (with the possible exception of preschool aged children). Among offenders who reached a clinically significant level of arousal, child molesters did differ from rapists in their arousal responses but this difference was mostly explained by lower arousal to adult stimuli among molesters rather than a greater arousal to underage stimuli. The pictorial modified Stroop task also showed a degree of similarity between both offender types, with a matching pattern of longer reaction times to older stimuli. Measures of gender preference for both tasks correlated with one another but other indices or responses to individual stimulus types did not.

Study 5 (chapter 5) compared the results of three versions (blocked, clustered and random) of the pictorial modified Stroop task using a sample of straight participants. Results indicated that a blocked design for the pictorial Stroop task produced larger effects than both the random and clustered designs. The risk of order effects in the blocked Stroop was still a cause for concern.

Study 6 (chapter 6) explored the relationship between a blocked pictorial Stroop design and a blocked CRT task. It also compared both these tasks with a pictorial Gender-Sex IAT that used a non-sexual category instead of the neutral category adopted in studies 2 and 3. The study found that a blocked version of the CRT task did not clearly tap into sexual orientation. There was very little evidence for an effect of sexual interest on the task. The study found that an IAT that used an opposite to sex category and pictures of males and female as the gender category stimuli was able to predict sexual orientation with a high degree of accuracy. On the other hand a pictorial Stroop design that was not preceded by a traditional Stroop task did not seem to perform as well as the one that was used in previous tasks. The task performed better when it came last in the battery of tasks. The IAT and pictorial Stroop tasks were significantly correlated.

7.2   Methodological findings and implications

The results of the six studies in this thesis have clear implications for the design of future indirect tasks relating to the measurement of sexual interest and associations. Taken as a whole it is clear that both the pictorial Stroop task and the Gender-Sex IAT (word-word or picture-word versions) tap into sexual interest. What exactly each is measuring and whether they are useful in measuring both deviant and non-deviant sexual interest will be discussed in sections 6.5 and 6.6.

7.2.1   The pictorial modified Stroop task

The Stroop task demonstrated an excellent ability in the first study to correctly classify participants as straight or gay. With the exception of study 6 (chapter 6) the pictorial Stroop task had an ability between good and excellent (Tape, 1999) in discriminating between gay and straight participants. However in study 6 (chapter 6), the discriminant ability of the pictorial Stroop was again in the excellent range when only participants who completed the Stroop task last were included. Study 5 (chapter 5) indicated that a blocked design outperformed both randomised and clustered pictorial Stroop designs. Despite this, both studies 3 and 4 (chapter 4), utilising a clustered design, demonstrated a good ability in discriminating participants based on sexual orientation. Results indicate that while a clustered design has a certain amount of utility, the optimum

design is one that uses more stimuli in each cluster than was used in studies 3, 4 and 5 or one that uses blocks instead of clusters. As mentioned, study 6 found that a blocked design, using extra (more attractive) stimuli, but that lacked a traditional Stroop task at the beginning of the procedure in study 6, underperformed relative to the Stroop task in study 2 when presented first but not when presented last in a series of tasks. This indicates that practice trials and/or cognitive demand may have an impact on the size of the pictorial modified Stroop effect observed.

Images of large cats were included as control images in each of the pictorial Stroop tasks. Large cats were chosen because it was felt that the control images should be animate. Cats were chosen somewhat arbitrarily but were preferred over images of apes, for example, since the amount of lighter shades on the cat images allowed for a more successful colouring of the images. When group comparisons were carried out the cat controls typically functioned adequately, i.e. mean response times differed very little between groups. However individual mean response times to cat stimuli differed quite significantly and individual participants didn't typically have consistent response times to cats relative to their response times to male and female images. This may indicate that cat images held some salience for some participants, or at least an ability to distract them from the task. Anecdotally, one offending participant remarked to the researcher after completing the tasks that he found cats fascinating and that he'd enjoyed that portion of the task. Since results were recorded in such a way that individual participants were not identifiable it was not possible to check whether this interest in cats had resulted in an increased response time to large cats. Adopting an alternative control image type may have yielded a more dependable baseline response time to pictorial stimuli; however it is likely that most categories of stimulus might have the potential to be salient to some participants.

The CRT task and the pictorial modified Stroop task are similar tasks, where the participant has to ignore irrelevant stimuli and focus on target stimuli in order to complete a goal. In the CRT task that goal is to identify the location of a dot and in the Stroop task the goal is to name a colour. In study 6, both tasks were set up almost identically and contained the same number of blocks and the same images. However, the tasks yielded very different results despite being hypothesised as measuring roughly equivalent phenomena given that several studies had successfully used the CRT task to

demonstrate a sexual content induced delay (Giotakos, 2005; Santtila et al., 2009; Wright & Adams, 1994, 1999). The CRT tasks in the previously published literature were all random designs so it is likely that the blocked design in the current CRT was responsible for the null finding in this case. However the fact that a blocked CRT design produced a null effect while a blocked pictorial Stroop did not is an interesting finding and has implication for the understanding of what both tasks are measuring (will be discussed later).

While similar, the pictorial Stroop and CRT tasks did differ from each other in several ways, however. First, since the Stroop task had only four target colours and the CRT had five dot locations, the CRT had an extra version of each image and was therefore slightly longer. Second, study 1 demonstrated that images used for a very similar Stroop task did not differ in the speed with which participants could categorise images by colour when the images were scrambled. No such check was carried out on the CRT task used in study 6. Indeed it is difficult to imagine an appropriate paradigm by which one could check whether it was inherently more difficult to locate the dot in any of the three stimulus types independent of the content of those stimuli. Third, responding to both tasks was slightly different. The CRT task involved pressing buttons that were arranged in the same spatial pattern as the target dots. The Stroop task, on the other hand, involved learning the arbitrary layout of the four coloured response buttons. Finally, and perhaps most importantly, the target stimuli in the pictorial Stroop task is more integrated with the to-be-ignored stimuli than in the CRT task. In the Stroop task the participant may look at the to-be-ignored stimulus as a whole or at any particular pat of the image they want and still be able to find the correct response. In the CRT task, the participant must scan the image looking for a small target and thus may not be as distracted by the to-be-ignored stimuli. With the exception of times where the dot was located in the centre of the screen, the target dot in many of the CRT stimuli presented in study 6 was not located on the body of the person or animal that was to-be-ignored (since the to-be-ignored stimuli filled the central portion of the images, not the corners). It is likely that this lack of integration of the target and to-be-ignored was a major component of the lack of agreement between the Stroop and CRT tasks. The implications for a theoretical understanding of the two tasks are discussed later (Section 6.5.1).

184

The studies in this thesis were devised under the assumption that a deviant sexual interest in children would look similar to adult sexual interest as measured by the pictorial modified Stroop task. Indeed, using a clustered pictorial Stroop design, study 3 yielded a pattern of results across groups that made intuitive sense. Control participants had, on average, the longest response time to images consistent with their sexual orientation that were age appropriate. The next longest (though significantly slower) response time was to child images of the gender that was consistent with their sexual orientation. Response times to orientation-inconsistent stimuli were quickest. The pattern of responses were similar for offenders hypothesised to be the most likely to have a sexual interest in children except that response times to adult orientation-consistent images were not significantly different to reaction times to orientation-consistent child images. The pattern found in study 2, using a blocked Stroop design, differed from that found in study 3 in that male child image response times were typically the second slowest time after adult images, regardless of the orientation of the participant. The results of study 2, therefore, go against the results of study 3 which seemed to indicate that child images elicit responses consistent with sexual orientation, though still significantly slower than adult images except for offenders (with a hypothesised sexual interest in children). The lack of a similar finding in study 2 may have something to do with the design used in that study, blocked instead of clustered, or be due to chance. Furthermore there could be some additional confound affecting the degree to which child images in the pictorial modified Stroop design taps into deviant sexual interest. It has already been pointed out in Chapter 2 that the male images from the NRP set, especially from the Tanner 4 age category, do not seem to be as realistic as other images. This or some other characteristic of the male images could have caused increased rumination in the blocked study. In addition, it is possible that male participants in these studies process child images in a different way to adult images, perhaps mindful of being considered 'to have deviant sexual interest', despite a guarantee of anonymity. As a result, it is possible that, for the child images, any slowdown of response times may not be as clear a measure of arousal-related attention as for the adult images.

The above explanation may go someway towards explaining the unexpectedly high response times to child images for some control participants. The presence of false positive responses due to factors other than sexual interest would have necessarily

impacted on the ability of ROC analysis etc. to demonstrate that a difference score between a response times to children and adults differentiates between those with a sexual interest in children and those without. It would have also impacted on the strength of any relationship between the Child-Sex IAT and the pictorial modified Stroop task. The full exploration of this question would require several steps. These would include the validation of the child images used in the Stroop task in terms of both attractiveness (to offenders with a sexual interest in children) and realness, the establishing of an optimum number of stimuli per cluster/block and the identification of an appropriate sample of participants of which half have a clear demonstrable sexual interest in children and half do not.

7.2.2   The Implicit Association Test (IAT)

Study 2 demonstrated that a Gender-Sex IAT using all-word stimuli and a neutral target category instead of an opposite to the *sex* category was able to tap into the sexual interests of a relatively homogenous sample of non-offending control participants. An almost identical Age-Sex IAT indicated no clear pattern towards stronger adult-sex than child-sex associations. This lack of an effect was taken, albeit cautiously, to indicate that this group of participants simply did not hold a clear schema relating to age and sex. However in study 3 neither IAT seemed to tap into cognitive processes as effectively as the Gender-Sex IAT had in study 2. A small sample of gay participants in the study (resulting from the small number of gay offenders) may have impacted on the ability of the Gender-Sex IAT to find an effect. However, there were a sufficient number of individuals hypothesised to have an association of children with sex to have reasonably expected an Age-Sex IAT effect if the task was functioning as predicted. Order of trial type presentation within both IATs was the same across all participants, since study 2 had indicated that order did not impact on the results of the tasks. However, it is possible that the greater heterogeneity, in terms of age and cognitive ability in the 3rd study resulted in an impact of order. While both IATs in study 3 yielded results that seemed driven by order, both demonstrated some indication that response times were being moderated in line with hypotheses.

From a methodological point of view, the above findings from studies 2 and 3, suggest that, while the IAT is capable of tapping into cognitive processes surrounding sexual

interest, the design used in those studies were not capable of eliciting a sufficiently large effect to demonstrate clear evidence of group differences. While the scope of the current thesis prohibited further testing of offenders, it was possible to attempt to develop a Gender-Sex IAT that was able to produce a clearer effect in a non-offending sample and therefore potentially indicate the best methodology to use in future studies. The Gender-Sex IAT used in study 6 adopted a pictorial design, and used the target category of *sexual* instead of *sex* in order to be able to use the opposite category of *nonsexual* (though the label of the category suggests an absence of sexual attributes rather than a true opposite, the category words themselves attempted to capture an 'unarousing' nuance of non-sexual). The change from a word-word to a pictorial-word IAT and the inclusion of an opposite target category produced results that had an excellent (Tape, 1999) ability to discriminate between gay and straight participants.

7.2.3 The Choice Reaction Time (CRT) task

Given that the development of a CRT task was not a primary goal of this thesis, the contribution that the results of study 6 can make to the methodology of future studies is somewhat limited, especially given that the version of the task used was considerably less successful than other versions (e.g. Santtila et al., 2009; Wright & Adams, 1994, 1999) in tapping into sexual interest. A null finding in study 6 indicates that a blocked design is not the optimum design to produce a CRT effect. However, methodological issues could not be ruled out as having contributed to this lack of effect.

7.3 Theoretical findings and implications

7.3.1 What are the tasks measuring?

This thesis set out the position in Chapters 1 and 2 that the pictorial modified Stroop task measures attentional components of the arousal process. It was also suggested that the CRT task taps into this process. On the other hand it was suggested that the Gender-Sex and Age-Sex IATs measure schema that may be related to sexual interest. Therefore, while both the pictorial modified Stroop task and the IATs (and later the CRT task) were hypothesised to be indirect measures of sexual orientation or sexual

interest, it was suggested that both are measuring separate processes/structures that typically correspond to orientation/interest.

The results of the thesis, taken as a whole, did not disprove this theory. However, the results provided only limited support for the theory. Evidence in favour of the suggestion that both are measuring different structures or processes would come from a clear agreement between results of the different tasks in cases where, theoretically, there should be no conflict between schema, arousal and self-reported sexual interest, i.e. control participants. There should also be a disagreement between tasks in some cases where there is a potential disagreement between schema and arousal and, possibly, self reported sexual interest. An example of this would be cases where offenders hold the "children as sexual beings" implicit theory (Ward & Keenan, 1999) but do not have a sexual interest in children or vice versa. In the absence of a validated objective measure to indicate the presence of cognitive distortions surrounding sex with children, an exploration of this question was relying on both the Stroop and the IATs to demonstrate clear group differences between non-offenders and offenders. Having a large number of offenders demonstrating a Stroop effect towards children, consistent with offence characteristics and admitted sexual interest in children, and a large number of offenders demonstrating a child-sex association in the Age-Sex IAT would facilitate a discussion on what degree of overlap there was between the two measures. A large amount of overlap would indicate that both were in fact measuring the same thing or that schema and arousal were closely matched in this sample. Less overlap would suggest that both were measuring separate things. Unfortunately, the IAT results in study 3 were not clear enough for such a comparison. In addition, if either the Age-Sex IAT or the pictorial modified Stroop scores were taken as a 'pure' measure of sexual interest/associations towards children, the current studies' findings would be indicating that many of the control participants have such 'deviant' interests or associations. While, it is not unlikely that some control participants would have deviant sexual interests or distorted cognitions relating to children, it is more likely, given the numbers involved, that the tasks used in study 3 did not have sufficient sensitivity or specificity to adequately discriminate those who did from those who didn't.

Despite the difficulties highlighted above in determining the relationship between the Stroop task and the IATs, some of the other results of the various studies do shed light

188

on the relationship. Both studies 2 and 6 found a correlation between the results of the Stroop task (for adult stimuli responses) and the Gender-Sex IAT. The correlation was between medium and large, depending on the method of calculation (and on task order in the 6th study), but no greater than $r = .6$. Both the Stroop and the IAT significantly predicted sexual orientation in both studies and in both studies a logistic regression indicated that both scores combined had a better ability to classify participants by their sexual orientation. These findings are consistent with the theory that both measures are measuring distinct constructs that are likely to largely agree in the case of control participants.

However, the findings are also consistent with the alternative explanation, that both the IAT and Stroop task are measuring the same construct but contain a certain amount of error. This explanation is unlikely given the difference between the two tasks, i.e. that one measures a slowing down in response to certain stimuli, while the other measures a speeding up. There is no clear evidence however, from the current findings that discounts this alternative totally. The complete lack of an Age-Sex IAT effect in study 2, despite the fact that the task was almost methodologically identical to the Gender-Sex IAT and despite the fact that the Stroop task measured a bias towards adult images in that study suggests a certain divergence between the underlying cognitive constructs/processes being measured by both the IAT and the Stroop paradigms.

Chapter 1 suggests that the pictorial modified Stroop task is measuring an attentional component of the arousal process. The study uses the model of arousal hypothesised by Spiering et al. (2004) to theorise how this attentional component might relate to other components of the arousal process. It was expected that there would be a clear relationship between the results of the pictorial modified Stroop task and sexual arousal as measured by the penile plethysmograph (PPG). Study 4 found little by way of a clear relationship between the measures. A gender preference index calculated for both tasks did correlate and both indicated a similar pattern of responding for both rapists and child molesters. Neither was able to significantly predict group membership (i.e. rapists or child molesters) and both were almost equivalently successful in discriminating offenders against males only from offenders against females. Despite this, when the two measures were broken down into response times and arousal towards different

categories of sexual partner/victim there was no clear pattern of correlation between the tasks.

While this result, runs counter to the hypothesis that both tasks would be clearly related, further studies involving the two measures would need to be conducted before a more solid conclusion could be drawn. Although this Stroop task used the same stimuli as has been used previously, it was programmed in a different institution, using different stimulus presentation software and included some design changes. One of these design changes involved the method of responding to the task. In the SRSTC Stroop participants responded to the images by selecting one of four coloured buttons on the screen using the mouse. Similarly to the argument made earlier regarding the effect of integration between to-be-ignored and to-be-attended-to stimuli, selecting on-screen responses via mouse may influence the manner in which the participant attends to the stimuli. While the pictorial modified Stroop task used in the study was clearly able to tap into sexual interest, methodological issues *may* have impacted on the strength of that effect and thus the strength of the relationship between that effect and the PPG results. In addition, the fact that the PPG was not able to identify group differences between the arousal towards underage stimuli between a sample of rapists and child molesters may indicate that there could be certain concerns about the reliability of those PPG results in this case.

As mentioned above, it was suggested in the introductory chapters of this thesis that the pictorial modified Stroop task was tapping into the attentional components of the arousal process[43]. It was suggested that this attentional component could influence the pictorial modified Stroop effect in three ways. First, a so-called 'fast' effect could occur when the participant's attention is grabbed during a pre-attentive stimulus encoding stage. Second, there could be so called 'slow' effects involving a slowing down of responses to subsequent tasks as result of the initial attention grabbing 'fast' effect. The likely cause of this would be that the demands of processing the original image would impact on the processing of any subsequent stimuli. Third, there could be additional slow effects as a result of higher order rumination or the activation of associated concepts surrounding the type/category of stimuli that are being presented at the time.

---

[43] Clearly any assumed link between the pictorial modified Stroop task and arousal must take into account the limited relationship between the task and PPG results in study 4.

The results of study 5 which compared a random, clustered and blocked design, taken with the results of studies 2, 3 and 7 indicate clearly that a blocked pictorial modified Stroop yielded the most robust effect. Taken together, these indicate that it is the higher order rumination or activation of associated concepts that drives the pictorial Stroop effect more so than fast or carry over effects. If fast and carry-over effects were sufficient to explain the phenomenon there would be very little difference between blocked and clustered performance. If fast effects alone were sufficient there would be very little difference between all three designs. It is not clear from the current series of studies what role, if any fast and carry over effects play or whether rumination etc. is sufficient to explain the entire phenomenon. One might expect rumination to produce a cumulative effect across a cluster or block with increasing reaction times over the course of the block or cluster. The graphs plotting response times over the course of the presentation of blocks and clusters produced in study 2 and 5 suggest that this may not be the case.

As suggested above (section 6.4.1) there may be a difference between what is being measured by the pictorial modified Stroop task when adult images are presented versus when child images are presented. While the results of studies 2-5 and 7 demonstrate quite clearly that an individual's pattern of responding is typically quite indicative of their sexual interest, individual responses to child stimuli can vary quite considerably, even among control participants. While, it is possible that deviant sexual interest is quite widespread and may, indeed, explain some of the responses observed, an alternative or additional explanation would seem necessary to explain the phenomenon more fully. As mentioned, it is possible that individual responses to child stimuli are contributed to by factors other than arousal-related attention. Indeed, it is likely that this is also the case for adult stimuli; however other variables may have more of an impact when child stimuli are presented. Increased reaction times may reflect a certain discomfort with the stimuli or perhaps a fear of having responses that seem "deviant". This fear of performing badly on the task may encourage the participant to be more meticulous in responding, and thus have a slowed reaction time, or simply to be distracted by the thought of doing badly on this particular set of images. Further alternative explanations for some of the unexpected results among the child images in the Stroop task have been suggested in the pictorial Stroop methodology section (6.4.1)

that includes questions about the quality/realness of some of the images used from the NRP image set (Laws & Gress, 2004). Additionally, since this thesis is contending that the pictorial modified Stroop effect for sexually salient images is driven primarily by a rumination effect and possibly by the activation of associated concepts, it is possible that such rumination does not always relate directly to the stimulus category. It could be that some features, or perhaps particular images from the set presented, can have a greater influence than others on what is being ruminated on. It is possible that images of older children had a greater influence on the rumination of some participants, thus making their responses seem potentially more deviant. This explanation is not sufficient to account for all the unexpected responses to child image categories, since in some cases the image category with the slowed reaction time was not the category which would have been predicted by the participant's self-reported sexual interest or by their adult stimuli Stroop results. While the 4th study did separate out older and younger children in their image clusters, the lack of full information about sexual preference and gender of victims made an investigation of this question difficult. However both the results of the pictorial modified Stroop task and the PPG measure in study 4 do indicate a large degree of ephebophilic sexual interest across the entire sample. This will be discussed in section 6.5.2. Clearly and as mentioned in section 6.4.1 further studies should attempt to explore the degree to which sexual interest in children is driving any slowdown of response times to child images in the pictorial modified Stroop task, and the degree to which confounding variables can be controlled for.

The thesis did not focus in sufficient detail on the CRT task to explore exactly what is being measured by that task. However, based on the findings that a blocked CRT design did not produce a sufficiently strong effect to overcome the influence of order and (potentially) any differences in the inherent difficulty in locating the dot across image types, it seems that rumination may not factor in the CRT task to the same degree as in the Stroop task. Given that previous studies did find a CRT effect using randomised presentation of trials (Santtila et al., 2009; Wright & Adams, 1994, 1999), it may be that a CRT effect for sexually salient stimuli is driven by fast effects, i.e. by attention capture during the pre-attentive stimulus encoding stage of arousal. If it is the case that a pictorial modified Stroop task is driven by slow effects and a CRT task is driven by fast effects, it may be the degree of integration between the to-be-ignored stimuli and the to-be-attended-to stimuli that creates the distinction. As previously mentioned, a

participant carrying out a pictorial modified Stroop task can focus his attention anywhere on the animal or person depicted and be able to respond quickly. This may lull the participant into focusing on those features of a given stimulus type that they find pleasing, thus allowing rumination and the activation of associations while maintaining an acceptable performance on the task. Acceptable performance is referring to the level of speed and accuracy on the task that the participant feels is sufficient to be considered to be doing the task correctly by the researcher. This expectation would be self-imposed by the participant, since the researcher only tells the participant to go as fast as they can while making as few mistakes as possible. In the CRT, in order to maintain an acceptable performance, the participant may have to work a little harder to ignore the to-be-ignored stimuli, since the stimuli and target are less integrated than in the Stroop task. In this way, a slowing down of responses on the CRT task would be where the participant has had their attention 'grabbed' by the to-be-ignored stimuli. The CRT design by Giotakos (2005) may have produced a slightly different effect since stimuli were presented first without the dot for two seconds, at which point the dot was superimposed on the image. Therefore, the task measures the degree to which the participant was able to refocus their attention after looking at the image for two seconds. It is possible that this method would capture more than one process.

7.3.2 What do the results tell us about offending/offenders and non-offending control participants?

Given that the results of studies 3 and 4 were not as clear-cut as had been hoped, there is a limit to the conclusions that can be drawn from this thesis regarding cognitive distortions and deviant sexual interest in offenders. Future pictorial modified Stroop tasks and IATs that follow on from the methodological recommendations made in this chapter may well shed more light on the cognitive process involved in deviant arousal and offence-related schema. The combining of these tasks with other implicit measures along with clinical investigations and inventories of distorted cognition/ implicit theories may well prove a fruitful avenue of research.

Interestingly, in study 4, neither the pictorial modified Stroop task nor the PPG measure, demonstrated a clear difference between the child molesters and the rapists regarding arousal and response times to underage stimuli. Among offenders who had

clinically significant PPG arousal levels rapists did typically have higher arousal to adult stimuli than child molesters. Again among these participants with higher arousal there was a difference in responses to preschool aged children between rapists and child molesters that was approaching significance (this was significant if all offenders were included regardless of the arousal level). However, for infant, grammar school and teen stimuli, both groups performed similarly. Neither task was able to discriminate significantly between rapists and child molesters. However a gender preference index for both indicated that both tasks were, indeed tapping into sexual interest. In addition the pattern of responding/arousal to both tasks indicated, typically, lower sexual interest in younger children. Both groups seemed to demonstrate a certain amount of ephebophillic sexual interest. Mean highest arousal, as measured by the PPG, using all participants, was highest for 'teen' images (Tanner stage 4) for both rapists and child molesters. Importantly, the calculation of a deviance index from the PPG results does not include responses to the teen category of images, presumably because arousal to teenage scenarios is not especially indicative of deviant behaviour. This suggests that the inclusion of Tanner 4 images in the pictorial modified Stroop task in study 3 may have weakened that tasks ability to discriminate between those with a sexual interest in younger children or those with a sexual interest in adults or late teens. While Tanner 1 and 2 images were clustered together separately from Tanner 3 and 4 images in study 4, certain differences between the methodology adopted in this study and those adopted in the studies conducted through Trinity College Dublin (see section 6.4.1), in addition to the fact that the PPG measure did not discriminate rapists from child molesters, make it difficult to determine why the Stroop task was unable to discriminate between the two groups in this particular sample.

It is contended that the pictorial modified Stroop tasks used in the thesis are indirectly measuring arousal and, thus, sexual interest and that the IATs are measuring sexual interest related schema. Clearly this contention is accompanied by the many caveats pointed out already. The results of the various studies indicate quite well that there is a relationship between the two types of measure and thus between schema and arousal. However at the same time, the results would suggest, albeit tenuously, that schema and arousal are distinct constructs, indicated by the lack of a clear convergence of IAT and pictorial Stroop. Spiering et al.'s (2004) model of arousal may offer a framework by which the overlap and divergence of these constructs/processes occurs. They state that

194

presentation of sexual stimuli can activate 'sexual memory' which includes, implicit sexual memory, sexual schemata and scripts. The presentation of sexual stimuli also brings about the regulation of information processing, which involves increased attention. Spiering et al. (2004) seem to favour the role of a regulatory process as the primary explanation of sexual content induced delay (SCID). However, it seems likely that the activation of sexual memory, both implicit and explicit, and the processing of these in working memory would also contribute to a slowing down of responses to another task. The current study, indeed suggests that not all content-induced delays are analogous, i.e. that a 'fast' Stroop or CRT effect is fundamentally different to a 'slow' effect. Findings in studies 2 and 5 that seem to indicate differences in the tau (τ) regions of the response distributions when sexually salient stimuli were presented using a blocked design highlight the point that it is unsafe to assume any content-induced delay is indicative of any one process. Regardless of the exact reason for a slow-down towards sexual stimuli, it appears safe to assume that the effect is not solely driven by the activation of sexual schema. On the other hand any increase in cognitive demand as a result of the arousal process would likely reduce an IAT effect, not exaggerate it. Therefore, a reduced reaction time when *male* and *sexual*, for example, are paired relative to *female* and *nonsexual* indicate a clear schematic association of these two concepts.

In the current series of studies, convergence between schema (as measured by the IAT) and arousal (as measured by the pictorial modified Stroop task) is only found in 'normal processes', i.e. between response times to adult stimuli and gender-sex associations. Also, this convergence was found only in studies 2 and 7 and not in study 3. The lack of convergence in study 3, again for adult Stroop images and the Gender-Sex IAT could be due to the methodology of the Stroop used or to the impact of cognitive ability or speed on results of the IAT. These may also explain the lack of a relationship between the pictorial modified Stroop task and the Age-Sex IAT in study 3. A relationship between the pictorial Stroop task and the Age-Sex IAT may not have been expected in study 2 as it has already been hypothesised that non-offenders may simply not have a schematic association of age and sex in the same way that an offender might. Despite this there is not enough evidence for distorted sexual schema in the current series of studies to extrapolate the role of a process equivalent to that proposed by Spiering et al. (2004) in offence-related cognition and arousal.

The thesis set out to potentially explore the relationship between cognitive distortion and deviant sexual arousal in offenders, since both are typically treated as distinct but potentially related state factors in offending, for example in the Integrated Theory of Sexual Offending (ITSO, Ward & Beech, 2006, 2008). Clearer findings in study 3 would have permitted a detailed discussion on this matter. Results with non offending participants were taken to indicate an overlap between sexual interest and associations for adult stimuli. Such a pattern found among offenders towards child stimuli by future lesser would suggest a clear link between deviant sexual interest and schema that contribute to some of the implicit theories that may be held by offenders.

As mentioned, the thesis sought to establish reliable measures (IATs) capable of measuring schema related to sexual cognitions and thus provide a potential method of measuring the sexual implicit theories of offenders. Again the inconclusive findings for the IAT tasks in study 3 prevented the exploration of this issue. Developing such measures are key in the further refinement of the judgement model of cognitive distortions (Ward, Gannon & Keown, 2006; Ward et al., 2007) and in teasing out the relationship between cognitive structures, processes and products for both cognitive distortions and sexual interest. Perhaps crucially, this thesis demonstrates that a clear theory of the processes underlying different implicit or indirect tasks is required before these interrelationships between the cognitive architecture of cognitive distortion and arousal can be unravelled. Treating attentional bias measures, such as the pictorial Stroop task and the CRT task as indirect measures of sexual interest and treating the facilitative influence of associations in the IAT as indicative of schemas is a useful starting point for such a theory. While a Gender-Sex IAT and a pictorial modified Stroop task for sexually salient stimuli may indirectly measure sexual orientation, treating the processes captured by these measures as equivalent does little to advance the understanding of the true underlying processes. The thesis results, taken with previous findings in the area, suggest that implicit or indirect measures that are carefully constructed, and that build on the methodological recommendations here (and elaborated on in section 6.6) may begin to provide an arsenal of measures by which the cognitive processes, structures etc. underpinning cognitive distortions and deviant sexual interest can be explored.

7.4 Practical issues in clinical and future research use stemming from the current findings

The current series of studies have demonstrated that the Implicit Association Test (IAT) and the pictorial modified Stroop task have potential for use in the assessment of sexual offenders. However, both are clearly in need of methodological refinement before their reliability and validity can be firmly established.

The pictorial modified Stroop task has been shown to be able to indirectly measure sexual interest towards adults. While promising, it's efficacy at reliably discriminating those with a sexual interest in children from those without is still inconclusive. It has also been shown by study 6 that the blocked pictorial modified Stroop task is likely to be measuring a slightly different aspect of arousal to the Choice Reaction Time (CRT) task. The Implicit Association Tests used in this study are hypothesised to tap into sexual interest-related schema. The utility of the Gender-Sex IAT as an indirect measure of sexual interest was demonstrated in both studies 2 and 7. It was not the intention of the current thesis to demonstrate that any one measure should be used as a complete measure of sexual interest and associated concepts; instead, the intention was to explore the potential for the inclusion of Implicit Association Tests and pictorial modified Stroop tasks in a battery of measures which combined could produce a fuller picture of sexual interest and offence-related cognition. As such there are drawbacks to both the IAT and the Stroop task that should be taken account of in the design of any test battery.

Since the pictorial modified Stroop task adopts a blocked or clustered design there is a risk that order may affect the results. A sufficient amount of practice trials, using neutral or word stimuli should minimise the risk of this somewhat. Given that the target colour in the pictorial design needs to be integrated with the to-be-ignored stimuli, there may be some difficulty in insuring that images depicting individuals with large variations in skin colour are matched in terms of the amount of target colour present. Similarly, the paradigm may not suit images that are complex or involve fine but essential detail since the act of colouring over the image with the target colour could obscure the important details. In addition, previous research on PPG have found that incest offenders may be more aroused by audio descriptions of abuse than by slide stimuli (Abel et al., 1981;

Lang et al., 1988; Murphy et al., 1986), perhaps suggesting that when the offender is allowed to imagine their own victim(s) in the abuse scenario, they demonstrate deviant arousal (Murphy et al., 1986). Indeed, one of the participants in the current study remarked to the researcher when filling in the question about sexual interest in children on the questionnaire that he did have a sexual interest in children but in "only one child"[44]. The importance of stimulus modality may result in a pictorial Stroop design not being the optimum design for the assessment of incest offenders. This further highlights the need for a battery of assessment tasks.

The IATs may also suffer from an impact of order as was seen in study 3. The impact of order may be reduced by adopting a design of IAT similar to that used in study 6. That study used pictures of males and females as the stimuli for the gender concepts and also used an approximately 'opposite to sex' category by using both *sexual* and *nonsexual category* labels. The use of pictorial stimuli in an Age-Sex IAT could prove problematic since both adult and child images would contain images of both males and females. Therefore there could be a conflict between a male/female-sex association and a child/adult-sex association. However if it is the category label and not the stimulus exemplars that drives the effect (De Houwer, 2001) there should be no significant impact of gender of stimulus. Furthermore, if the stimulus exemplars are not driving the effect, a word version of the Age-Sex IAT *should* be just as effective. In addition, it is possible to record, and therefore analyse, values for when male children and female children are presented separately. Finally, there is no indication in Ward and Keenan's (1999) Implicit Theories theory whether offenders would likely hold 'the children as sexual beings' implicit theory for all children regardless of gender or not. If they do, and if the IAT is truly measuring the 'children as sexual beings' implicit theory, an Age-Sex IAT containing images of male and female children should still yield an effect among offenders who hold that implicit theory.

---

[44] The researcher was available to the participants to answer questions as they were filling in their questionnaires.

7.5 Conclusions

The results of the thesis demonstrated that the pictorial modified Stroop task and the Implicit Association Test can tap into sexual interests and associations. They also demonstrated that in most cases the two measures were related to one another. The results were less clear regarding attention and associations to child stimuli among both offending and control participants. It was not clear whether this lack of clarity was wholly as a result of a lack of sensitivity within the tasks or an indicator that sexual associations and interest towards children were phenomenologically different to those towards adults. A lack of clear results from the IAT task in the study comparing IAT and pictorial Stroop results of sexual offenders against children impacted on the thesis' ability to address whether there was a divergence of responses to the different tasks for offenders. Such a divergence would have added more weight to the suggestion that the pictorial Stroop task measures sexual arousal-related attention and that the IAT measures schematic sexual associations possibly relating to implicit theories. Despite this, the clear but imperfect correlations found between the pictorial Stroop and IAT tasks, despite the fact that one is measuring a slowing down of responses to a sexually salient stimuli, while the other is measuring a speeding up towards a concept pair, strongly indicates that both are indirectly measuring related cognitive processes or structures through different routes. The suggestion that these different routes are related to schema and arousal needs be further explored by future research. The thesis concludes that while the Implicit Association Test and the pictorial modified Stroop task are both paradigms that offer clear potential for use in the assessment of sexual offenders, both need to be further tested and validated. Importantly the methodological recommendations made in this thesis should be tested and a clear understanding of the processes measured by each task should be established so that the emerging field involving cognitive or indirect approaches to forensic assessment can establish best practices in experimental design and a framework by which to explore the cognitive processes of sexual offenders.

References

Abel, G. G. (1981). Two methods of measuring penile response. *Behavior Therapy, 12*(3), 320-328.

Abel, G. G., Becker, J. V., & Cunningham-Rathner, J. (1984). Complications, consent, and cognitions in sex between children and adults. *International Journal of Law and Psychiatry, 7*(1), 89-103.

Abel, G. G., Becker, J. V., Murphy, W. D., & Flanagan, B. (1981). Identifying dangerous child molesters. In J. R. Stuart (Ed.), *Violent behavior: social learning approaches to prediction, management and treatment* (pp. 116-137). New York: Brunner/Mazel.

Abel, G. G., Gore, D. K., Holland, C. L., & Camp, N. (1989). The measurement of the cognitive distortions of child molesters. *Annals of Sex Research, 2*(2), 135-152.

Abel, G. G., Huffman, J., Warberg, B., & Holland, C. L. (1998). Visual reaction time and plethysmography as measures of sexual interest in child molesters. *Sexual Abuse: Journal of Research and Treatment, 10*(2), 81-95.

Abel, G. G., Jordan, A., Hand, C. G., Holland, L. A., & Phipps, A. (2001). Classification models of child molesters utilizing the Abel Assessment for child sexual abuse interest. *Child Abuse & Neglect, 25*(5), 703-718.

Abel, G. G., Jordan, A., Rouleau, J. L., Emerick, R., Barboza-Whitehead, S., & Osborn, C. (2004). Use of visual reaction time to assess male adolescents who molest children. *Sexual Abuse: Journal of Research & Treatment, 16*(3), 255-265.

Algom, D., Chajut, E., & Lev, S. (2004). A Rational Look at the Emotional Stroop Phenomenon: A Generic Slowdown, Not a Stroop Effect. *Journal of Experimental Psychology: General, 133*(3), 323-338.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders : DSM-IV-TR* (4th ed.). Washington: American Psychiatric Association.

Ashwin, C., Wheelwright, S., & Baron-Cohen, S. (2006). Attention bias to faces in Asperger Syndrome: A pictorial emotion Stroop study. *Psychological Medicine, 36*(6), 835-843.

Ballesteros, S., Reales, J. M., & Manga, D. (2000). Effects of type of design (blocked vs. randomized) on Stroop and emotional Stroop tasks. *Psicothema, 12*, 60-63.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General, 128*(1), 32-55.

Banse, R., Schmidt, A., & Clarbour, J. (2009). *Indirect Measures of Sexual Interest in Child Sex Offenders: A Multi-Method Approach.* Manuscript submitted for publication.

Barbaree, H. E., & Mewhort, D. J. K. (1994). The effects of the z-score transformation on measures of relative erectile response strength - a reappraisal. *Behaviour Research and Therapy, 32*(5), 547-558.

Barboza-Whitehead, S. E. (2001). *Discriminant validity of the Abel Assessment for Sexual Interest with juveniles who admit versus deny their sexual offenses.* Univ Microfilms International.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 74*(5), 1252-1265.

Becker, E. S., Strohbach, D., & Rinck, M. (1999). A specific attentional bias in suicide attempters. *Journal of Nervous and Mental Disease, 187*(12), 730-735.

Beech, A. R., Fisher, D. D., & Thornton, D. (2003). Risk assessment of sex offenders. *Professional Psychology: Research and Practice, 34*(4), 339-352.

Beech, A. R., Kalmus, E., Tipper, S. P., Baudouin, J. Y., Flak, V., & Humphreys, G. W. (2008). Children Induce an Enhanced Attentional Blink in Child Molesters. *Psychological Assessment, 20*(4), 397-402.

Blake, E., & Gannon, T. A. (2008). Social perception deficits, cognitive distortions, and empathy deficits in sex offenders: A brief review. *Trauma, Violence, and Abuse, 9*(1), 34-55.

Blanchard, R., Kuban, M. E., Blak, T., Cantor, J. M., Klassen, P., & Dickey, R. (2006). Phallometric Comparison of Pedophilic Interest in Nonadmitting Sexual Offenders Against Stepdaughters, Biological Daughters, Other Biologically Related Girls, and Unrelated Girls. *Sexual Abuse: Journal of Research and Treatment, 18*(1), 1-14.

Blanton, H., & Jaccard, J. (2006). Arbitrary Metrics in Psychology. *American Psychologist, 61*(1), 27-41.

Boyer, M., & Dickerson, M. (2003). Attentional bias and addictive behaviour: automaticity in a gambling-specific modified Stroop task. *Addiction, 98*(1), 61.

Bradley, B. P., Mogg, K., Millar, N., & White, J. (1995). Selective processing of negative information: Effects of clinical anxiety, concurrent depression, and awareness. *Journal of Abnormal Psychology, 104*(3), 532-536.

Braet, C., & Crombez, G. (2003). Cognitive interference due to food cues in childhood obesity. *Journal of Clinical Child and Adolescent Psychology, 32*(1), 32-39.

Brown, A. S., Gray, N. S., & Snowden, R. J. (2009). Implicit Measurement of Sexual Associations in Child Sex Abusers: Role of Victim Type and Denial. *Sexual Abuse: A Journal of Research and Treatment, 21*(2), 166-180.

Bumby, K. M. (1996). Assessing the cognitive distortions of child molesters and rapists: Developments and validation of the MOLEST and RAPE scales. *Sexual Abuse: A Journal of Research and Treatment, 8*, 37–54.

Cannon, B. J. (2003). An emotional Stroop effect to malingering-related words. *Perceptual and Motor Skills, 96*(3), 827-834.

Carpenter, K. M., Schreiber, E., Church, S., & McDowell, D. (2006). Drug Stroop performance: Relationships with primary substance of use and treatment outcome in a drug-dependent outpatient sample. *Addictive Behaviors, 31*(1), 174-181.

Carrigan, M. H., Drobes, D. J., & Randall, C. L. (2004). Attentional Bias and Drinking to Cope With Social Anxiety. *Psychology of Addictive Behaviors, 18*(4), 374-380.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J.: L. Erlbaum Associates.

Constantine, R., McNally, R. J., & Hornig, C. D. (2001). Snake Fear and the Pictorial Emotional Stroop Paradigm. *Cognitive Therapy & Research, 25*(6), 757.

Côté, S., & Bouchard, S. (2005). Documenting the Efficacy of Virtual Reality Exposure with Psychophysiological and Information Processing Measures. *Applied Psychophysiology and Biofeedback, 30*(3), 217-232.

Côté, S., & Bouchard, S. (2009). Cognitive Mechanisms Underlying Virtual Reality Exposure. *CyberPsychology & Behavior, 12*(2), 121-129.

Cox, W. M., Fadardi, J. S., & Klinger, E. (2006). Motivational Processes Underlying Implicit Cognition in Addiction. In R. W. Wiers & A. W. Stacy (Eds.), *The handbook of implicit cognition and addiction* (pp. 253-266). Thousand Oaks, CA: Sage Publishers.

Cox, W. M., Fadardi, J. S., & Pothos, E. M. (2006). The Addiction-Stroop Test: Theoretical Considerations and Procedural Recommendations. *Psychological Bulletin, 132*(3), 443-476.

Cox, W. M., Yeates, G. N., & Regan, C. N. (1999). Effects of alcohol cues on cognitive processing in heavy and light drinkers. *Drug and Alcohol Dependence, 55*(1), 85-89.

Craig, L. A., Browne, K. D., & Beech, A. (2008). *Assessing Risk in Sex Offenders: A Practitioner's Guide*. Sussex, UK: Wiley.

Dalgleish, T., Taghavi, R., Neshat-Doost, H., Moradi, A., Canterbury, R., & Yule, W. (2003). Patterns of processing bias for emotional information across clinical disorders: A comparison of attention, memory, and prospective cognition in children and adolescents with depression, generalized anxiety, and posttraumatic stress disorder. *Journal of Clinical Child and Adolescent Psychology, 32*(1), 10-21.

De Houwer, J. (2001). A Structural and Process Analysis of the Implicit Association Test. *Journal of Experimental Social Psychology, 37*(6), 443-451.

Dubner, A. E., & Motta, R. W. (1999). Sexually and Physically Abused Foster Care Children and Posttraumatic Stress Disorder. *Journal of Consulting and Clinical Psychology, 67*(3), 367-373.

Egloff, B., & Schmukle, S. C. (2004). Gender differences in implicit and explicit anxiety measures. *Personality and Individual Differences, 36*(8), 1807-1815.

Elsesser, K., Heuschen, I., Pundt, I., & Sartory, G. (2006). Attentional bias and evoked heart-rate response in specific phobia. *Cognition & Emotion, 20*(8), 1092-1107.

Erblich, J., Montgomery, G. H., Valdimarsdottir, H. B., Cloitre, M., & Bovbjerg, D. H. (2003). Biased cognitive processing of cancer-related information among women with family histories of breast cancer: Evidence from a cancer Stroop task. *Health Psychology, 22*(3), 235-244.

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin, 125*(6), 777-799.

Field, A. P. (2009). *Discovering statistics using SPSS : (and sex and drugs and rock 'n' roll)* (3rd ed.). Los Angeles ; London: SAGE.

Field, M. (2005). Cannabis 'dependence' and attentional bias for cannabis-related words. *Behavioural Pharmacology, 16*(5), 473-476.

Field, M., Duka, T., Tyler, E., & Schoenmakers, T. (2009). Attentional bias modification in tobacco smokers. *Nicotine Tob Res, 11*(7), 812-822.

Finkelhor, D. (1984). *Child Sexual Abuse*. New York: Free Press.

Firestone, P., Kingston, D. A., Wexler, A., & Bradford, J. M. (2006). Long-term follow-up of exhibitionists: Psychological, phallometric, and offense characteristics. *Journal of the American Academy of Psychiatry and the Law, 34*(3), 349-359.

Flak, V., Beech, A., & Fisher, D. (2007). Forensic assessment of deviant sexual interests: The current position. *Issues in Forensic Psychology, 6*, 70-83.

Foa, E. B., Feske, U., Murdock, T. B., Kozak, M. J., & McCarthy, P. R. (1991). Processing of Threat-Related Information in Rape Victims. *Journal of Abnormal Psychology, 100*(2), 156-162.

Formea, G. M., & Burns, G. L. (1996). Selective processing of food, weight, and body-shape words in nonpatient women with bulimia nervosa: Interference on the Stroop task. *Journal of Psychopathology and Behavioral Assessment, 18*(2), 105-118.

Fortune, D. G., Richards, H. L., Corrin, A., Taylor, R. J., Griffiths, C. E. M., & Main, C. J. (2003). Attentional Bias for Psoriasis-Specific and Psychosocial Threat in Patients with Psoriasis. *Journal of Behavioral Medicine, 26*(3), 211.

Freeman, J. B., & Beck, J. G. (2000). Cognitive Interference for Trauma Cues in Sexually Abused Adolescent Girls With Posttraumatic Stress Disorder. *Journal of Clinical Child Psychology, 29*(2), 245-256.

French, C. C., Richards, A., & Scholfield, E. J. C. (1996). Hypomania, anxiety and the emotional Stroop. *British Journal of Clinical Psychology, 35*(4), 617-626.

Frenzel, R. R., & Lang, R. A. (1989). Identifying sexual preferences in intrafamilial and extrafamilial child sexual abusers. *Annals of Sex Research, 2*(3), 255-275.

Freund, K., & Watson, R. (1991). Assessment of the sensitivity and specificity of a phallometric test: An update of "Phallometric Diagnosis of Pedophilia". *Psychological Assessment, 3*, 254-260.

Freund, K., Watson, R., & Dickey, R. (1991). Sex offences against female-children perpetrated by men who are not pedophiles. *Journal of Sex Research, 28*(3), 409-423.

Gannon, T. A. (2006). Increasing Honest Responding on Cognitive Distortions in Child Molesters: The Bogus Pipeline Procedure. *Journal of Interpersonal Violence, 21*(3), 358-375.

Gannon, T. A., Keown, K., & Polaschek, D. (2007). Increasing Honest Responding on Cognitive Distortions in Child Molesters: The Bogus Pipeline Revisited. *Sexual Abuse: A Journal of Research and Treatment, 19*(1), 5-22.

Gannon, T. A., & Polaschek, D. L. L. (2005). Do Child Molesters Deliberately Fake Good on Cognitive Distortion Questionnaires? An Information Processing-Based Investigation. *Sexual Abuse: Journal of Research and Treatment, 17*(2), 183-200.

Gannon, T. A., & Polaschek, D. L. L. (2006). Cognitive distortions in child molesters: A re-examination of key theories and research. *Clinical Psychology Review, 26*(8), 1000-1019.

Gannon, T. A., Ward, T., & Collie, R. (2007). Cognitive distortions in child molesters: Theoretical and research developments over the past two decades. *Aggression and Violent Behavior, 12*(4), 402-416.

Gannon, T. A., Wright, D. B., Beech, A. R., & Williams, S. (2006). Do child molesters hold distorted beliefs? What does their memory recall tell us?
. *Journal of Sexual Aggression, 12*(1), 5-18.

Geer, J. H., & Bellard, H. S. (1996). Sexual content induced delays in unprimed lexical decisions: Gender and context effects. *Archives of Sexual Behavior, 25*(4), 379-395.

Geer, J. H., & Melton, J. S. (1997). Sexual content-induced delay with double-entendre words. *Archives of Sexual Behavior, 26*(3), 295-316.

Giotakos, O. (2005). A combination of viewing reaction time and incidental learning task in child molesters, rapists, and control males and females. *European Journal of Sexology; Sexologies, 54*, 13-22.

Glasgow, D. V., Osborne, A., & Croxen, J. (2003). An assessment tool for investigating paedophile sexual interest using viewing time: an application of single case methodology. *British Journal of Learning Disabilities, 31*(2), 96-102.

Gotlib, I. H., & Cane, D. B. (1987). Construct accessibility and clinical depression: A longitudinal investigation. *Journal of Abnormal Psychology, 96*(3), 199-204.

Gray, N. S., Brown, A. S., MacCulloch, M. J., Smith, J., & Snowden, R. J. (2005). An Implicit Test of the Associations Between Children and Sex in Pedophiles. *Journal of Abnormal Psychology, 114*(2), 304-308.

Gray, N. S., & Snowden, R. J. (in press). The Implicit Association Test as a measure of sexual interest. In D. Thornton & D. R. Laws (Eds.), *Cognitive Approaches to the Assessment of Sexual Interest in Sexual Offenders*.

Green, M. W., Elliman, N. A., & Rogers, P. J. (1996). Hunger, caloric preloading and the selective processing of food and body shape words. *British Journal of Clinical Psychology, 35*(1), 143-151.

Greenberg, D. M., Firestone, P., Nunes, K. L., Bradford, J. M., & Curry, S. (2005). Biological fathers and stepfathers who molest their daughters: Psychological, phallometric, and criminal features. *Sexual Abuse-a Journal of Research and Treatment, 17*(1), 39-46.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464-1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*(2), 197-216.

Gress, C. L. Z. (2005). Viewing time measures and sexual interest: another piece of the puzzle. *Journal of Sexual Aggression, 11*(2), 117-125.

Gress, C. L. Z. (2008). *Delays in attentional processing when viewing sexual imagery : the development and comparison of two measures.* University of Victoria, Victoria.

Gress, C. L. Z., & Laws, D. R. (in press). Cognitive modeling of sexual arousal and interest: Choice reaction time measures. In D. Thornton & D. R. Laws (Eds.), *Cognitive Approaches to the Assessment of Sexual Interest in Sexual Offenders*.

Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*(2), 348-362.

Hanson, R. K., Gizzarelli, R., & Scott, H. (1994). The attitudes of incest offenders - sexual entitlement and acceptance of sex with children. *Criminal Justice and Behavior, 21*(2), 187-202.

Harris, G. T., Rice, M. E., Quinsey, V. L., & Chaplin, T. C. (1996). Viewing time as a measure of sexual interest among child molesters and normal heterosexual men. *Behaviour Research and Therapy, 34*(4), 389-394.

Harris, G. T., Rice, M. E., Quinsey, V. L., Chaplin, T. C., & Earls, C. (1992). Maximizing the discriminant validity of phallometric assessment data. *Psychological Assessment, 4*, 502-511.

Hecker, J. E., King, M. W., & Scoular, R. J. (in press). The startle-probe reflex: An alternative approach to the measurement of sexual interest. In D. Thornton & D. R. Laws (Eds.), *Cognitive Approaches to the Assessment of Sexual Interest in Sexual Offenders*.

Heim-Dreger, U., Kohlmann, C.-W., Eschenbeck, H., & Burkhardt, U. (2006). Attentional Biases for Threatening Faces in Children : Vigilant and Avoidant Processes. *Emotion, 6*(2), 320-325.

Hester, R., Dixon, V., & Garavan, H. (2006). A consistent attentional bias for drug-related material in active cocaine users across word and picture versions of the emotional Stroop task. *Drug and Alcohol Dependence, 81*(3), 251-257.

Holle, C., Neely, J. H., & Heimberg, R. G. (1997). The Effects of Blocked Versus Random Presentation and Semantic Relatedness of Stimulus Words on Response to a Modified Stroop Task Among Social Phobics. *Cognitive Therapy & Research, 21*(6), 681-697.

Janssen, E., Everaerd, W., Spiering, M., & Janssen, J. (2000). Automatic processes and the appraisal of sexual stimuli: Toward an information processing model of sexual arousal. *Journal of Sex Research, 37*(1), 8-23.

Janssen, E., Vorst, H., & Finn, P. (2002). The Sexual Inhibition (SIS) and Sexual Excitation (SES) Scales: II. Predicting Psychophysiological Response Patterns. *The Journal of Sex Research, 39*(2), 127-132.

Jessop, D. C., Rutter, D. R., Sharma, D., & Albery, I. P. (2004). Emotion and adherence to treatment in people with asthma: An application of the emotional Stroop paradigm. *British Journal of Psychology, 95*(2), 127-147.

Kalmus, E., & Beech, A. R. (2005). Forensic assessment of sexual interest: A review. *Aggression and Violent Behavior, 10*(2), 193-217.

Keenan, T., & Ward, T. (2003). Developmental antecedants of sexual offending. In T. Ward, D. R. Laws & S. M. Hudson (Eds.), *Sexual Deviance: Issues and Controversies* (pp. 119-134). London: Sage.

Kindt, M., & Brosschot, J. F. (1997). Phobia-Related Cognitive Bias for Pictorial and Linguistic Stimuli. *Journal of Abnormal Psychology, 106*(4), 644-648.

Kindt, M., & Brosschot, J. F. (1999). Cognitive bias in spider-phobic children: Comparison of a pictorial and a linguistic spider Stroop. *Journal of Psychopathology and Behavioral Assessment, 21*(3), 207-220.

Kolassa, I.-T., Musial, F., Mohr, A., Trippe, R. H., & Miltner, W. H. R. (2005). Electrophysiological correlates of threat processing in spider phobics. *Psychophysiology, 42*(5), 520-530.

Koukounas, E., & Over, R. (1999). Allocation of attentional resources during habituation and dishabituation of male sexual arousal. *Archives of Sexual Behavior, 28*(6), 539-552.

Koukounas, E., & Over, R. (2000). Changes in the magnitude of the eyeblink startle response during habituation of sexual arousal. *Behaviour Research and Therapy, 38*(6), 573-584.

Koukounas, E., & Over, R. (2001). Habituation of male sexual arousal: Effects of attentional focus. *Biological Psychology, 58*(1), 49-64.

Kramer, D. A., & Goldman, M. S. (2003). Using a Modified Stroop Task to Implicitly Discern the Cognitive Organization of Alcohol Expectancies. *Journal of Abnormal Psychology, 112*(1), 171-175.

Labarge, A. S., Cash, T. F., & Brown, T. A. (1998). Use of a Modified Stroop Task to Examine Appearance-Schematic Information Processing in College Women. *Cognitive Therapy & Research, 22*(2), 179-190.

Lang, R. A., Black, E. L., Frenzel, R. R., & Checkley, K. L. (1988). Aggression and erotic attraction toward children in incestuous and pedophilic men. *Annals of Sex Research, 1*(3), 417-441.

Langevin, R., & Curnoe, S. (2007). The Therapeutic Challenge of the Learning Impaired Sex Offender [Electronic Version]. *Sexual Offender Treatment*, 2. Retrieved December 17, 2007 from http://www.sexual-offender-treatment.org/56.0.html.

Langevin, R., & Watson, R. (1991). A comparison of incestuous biological and stepfathers. *Annals of Sex Research, 4*(2), 141-150.

Langton, C. M. (2007). Rape-related cognition: current research. In T. A. Gannon, T. Ward, A. Beech & D. Fisher (Eds.), *Aggressive Offenders' Cognition: Theory, Research and Practice*. Sussex, UK: Wiley.

Lavy, E., & Van den Hout, M. (1993). Selective attention evidenced by pictorial and linguistic Stroop tasks. *Behavior Therapy, 24*(4), 645-657.

Laws, D. R., & Gress, C. L. Z. (2004). Seeing things differently: The viewing time alternative to penile plethysmography. *Legal and Criminological Psychology, 9*, 183-196.

Laws, D. R., & Marshall, W. L. (1990). A conditioning theory of the etiology and maintenance of deviant sexual preference and behavior. In H. E. Barbaree (Ed.), *Handbook of sexual assault: Issues, theories, and treatment of the offender.* (pp. 209-229). New York, NY, US: Plenum Press.

Lecci, L., & Cohen, D. J. (2002). Perceptual consequences of an illness-concern induction and its relation to hypochondriacal tendencies. *Health Psychology, 21*(2), 147-156.

Letourneau, E. J. (2002). A Comparison of Objective Measures of Sexual Arousal and Interest: Visual Reaction Time and Penile Plethysmography. *Sexual Abuse: Journal of Research and Treatment, 14*(3), 207-223.

Lundh, L.-G., & Öst, L.-G. (2001). Attentional bias, self-consciousness and perfectionism in social phobia before and after cognitive-behaviour therapy. *Scandinavian Journal of Behaviour Therapy, 30*(1), 4-16.

MacLeod, C. M. (1991). Half a Century of Research on the Stroop Effect : An Integrative Review. *Psychological Bulletin, 109*(2), 163-203.

Macmillan, N. A., & Creelman, C. D. (1990). *Detection theory : a user's guide*. Cambridge: Cambridge University Press.

Mahamedi, F., & Heatherton, T. F. (1993). Effects of high calorie preloads on selective processing of food and body shape stimuli among dieters and nondieters. *International Journal of Eating Disorders, 13*(3), 305-314.

Maletzky, B. M. (2003). Letter to the Editor. *Sexual Abuse: Journal of Research and Treatment, 15*(4), 393-393.

Mann, R., & Beech, A. (2003). Cognitive Distortions, Schemas, and Implicit Theories. In T. Ward, D. R. Laws & S. M. Hudson (Eds.), *Sexual Deviance: Issues and Controversies* (pp. 135-153). London: Sage.

Mann, R., Webster, S., Wakeling, H., & Marshall, W. (2007). The measurement and influence of child sexual abuse supportive beliefs. *Psychology, Crime and Law, 13*, 443-458.

Marshall, W. L., Barbaree, H. E., & Butt, J. (1988). Sexual offenders against male children: Sexual preferences. *Behaviour Research and Therapy, 26*(5), 383-391.

Marshall, W. L., Barbaree, H. E., & Christophe, D. (1986). Sexual offenders against female children: Sexual preferences for age of victims and type of behaviour. *Canadian Journal of Behavioural Science/Revue canadienne des Sciences du comportement, 18*(4), 424-439.

Marshall, W. L., & Fernandez, Y. M. (2000). Phallometric testing with sexual offenders: Limits to its value. *Clinical Psychology Review, 20*(7), 807-822.

Marshall, W. L., & Fernandez, Y. M. (2003). Sexual preferences: Are they useful in the assessment and treatment of sexual offenders? *Aggression and Violent Behavior, 8*(2), 131-143.

Marshall, W. L., Marshall, L. E., Serran, G. A., & Fernandez, Y. M. (2006). *Treating Sexual Offenders*. London: Routledge.

Marshall, W. L., Payne, K., Barbaree, H. E., & Eccles, A. (1991). Exhibitionists: Sexual preferences for exposing. *Behaviour Research and Therapy, 29*(1), 37-40.

Maruna, S., & Mann, R. E. (2006). A fundamental attribution error? Rethinking cognitive distortions. *Legal and Criminological Psychology, 11*(2), 155-177.

Marziano, V., Ward, T., Beech, A. R., & Pattison, P. (2006). Identification of five fundamental implicit theories underlying cognitive distortions in child abusers: A preliminary study. *Psychology, Crime & Law, 12*(1), 97-105.

Mattia, J. I., Heimberg, R. G., & Hope, D. A. (1993). The revised Stroop color-naming task in social phobics. *Behaviour Research and Therapy, 31*(3), 305-313.

McCusker, C. G., & Gettings, B. (1997). Automaticity of cognitive biases in addictive behaviours: Further evidence with gamblers. *British Journal of Clinical Psychology, 36*(4), 543-554.

McKenna, F. P., & Sharma, D. (1995). Intrusive cognitions: An investigation of the emotional Stroop task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(6), 1595-1607.

McKenna, F. P., & Sharma, D. (2004). Reversing the Emotional Stroop Effect Reveals That It Is Not What It Seems: The Role of Fast and Slow Components. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 382-392.

McNally, R. J., Riemann, B. C., Louro, C. E., & Lukach, B. M. (1992). Cognitive processing of emotional information in panic disorder. *Behaviour Research and Therapy, 30*(2), 143-149.

Mihailides, S., Devilly, G. J., & Ward, T. (2004). Implicit cognitive distortions and sexual offending. *Sexual Abuse: Journal of Research and Treatment, 16*(4), 333-350.

Mogg, K., & Bradley, B. P. (2002). Selective processing of smoking-related cues in smokers: Manipulation of deprivation level and comparison of three measures of processing bias. *Journal of Psychopharmacology, 16*(4), 385-392.

Mogg, K., Bradley, B. P., Dixon, C., Fisher, S., Twelftree, H., & McWilliams, A. (2000). Trait anxiety, defensiveness and selective processing of threat: An investigation using two measures of attentional bias. *Personality and Individual Differences, 28*(6), 1063-1077.

Munafò, M., Mogg, K., Roberts, S., Bradley, B. P., & Murphy, M. (2003). Selective processing of smoking-related cues in current smokers, ex-smokers and never-smokers on the modified Stroop task. *Journal of Psychopharmacology, 17*(3), 310-316.

Murphy, W. D., & Barbaree, H. E. (1994). *Assessments of Sex Offenders by Measures of Erectile Response: Psychometric Properties and Decision Making*. Brandon, VT: Safer Society Press.

Murphy, W. D., Haynes, M. R., Stalgaitis, S. J., & Flanagan, B. (1986). Differential sexual responding among four groups of sexual offenders against children. *Journal of Psychopathology and Behavioral Assessment, 8*(4), 339-353.

Neidigh, L., & Krop, H. (1992). Cognitive distortions among child sexual offenders. *Journal of Sex Education and Therapy, 18*, 208–215.

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-go Association Task. *Social Cognition, 19*(6), 625-666.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and Using the Implicit Association Test: II. Method Variables and Construct Validity. *Personality and Social Psychology Bulletin, 31*(2), 166-180.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at Age 7: A Methodological and Conceptual Review. In J. A. Bargh (Ed.), *Automatic processes in social thinking and behavior* (pp. 265-292). New York, NY, US: Psychology Press.

Nunes, K. L., Firestone, P., & Baldwin, M. W. (2007). Indirect assessment of cognitions of child sexual abusers with the Implicit Association Test. *Criminal Justice and Behavior, 34*(4), 454-475.

Pennington, B. F. (2002). *The development of psychopathology : nature and nurture*. New York ; London: The Guilford Press.

Perpiña, C., Leonard, T., Treasure, J., Bond, A., & Baños, R. (1998). Selective processing of food- and body-related information and autonomic arousal in patients with eating disorders. *The Spanish Journal of Psychology, 1*(1), 3-10.

Phaf, R. H., & Kan, K.-J. (2007). The automaticity of emotional Stroop: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 38*(2), 184-199.

Polaschek, D. L. L., & Gannon, T. A. (2004). The implicit theories of rapists: What convicted offenders tell us. *Sexual Abuse: Journal of Research and Treatment, 16*(4), 299-314.

Polaschek, D. L. L., & Ward, T. (2002). The implicit theories of potential rapists: What our questionnaires tell us. *Aggression and Violent Behavior, 7*(4), 385-406.

Powell, J., Tait, S., & Lessiter, J. (2002). Cigarette smoking and attention to signals of reward and threat in the Stroop paradigm. *Addiction, 97*(9), 1163.

Price, S. (2005). *Measuring the Deviant Schema of Sexual Offenders: A Stroop Replication Study.* Unpublished Masters, Carleton University.

Quinsey, V. L., Chaplin, T. C., & Carrigan, W. F. (1979). Sexual preferences among incestuous and nonincestuous child molesters. *Behavior Therapy, 10*(4), 562-565.

Rabinowitz Greenberg, S. R., Firestone, P., Bradford, J. M., & Greenberg, D. M. (2002). Prediction of Recidivism in Exhibitionists: Psychological, Phallometric, and Offense Factors. *Sexual Abuse: Journal of Research and Treatment, 14*(4), 329-347.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*(3), 510-532.

Rice, M., & Harris, G. (2005). Comparing Effect Sizes in Follow-Up Studies: ROC Area, Cohen's d, and r. *Law and Human Behavior, 29*(5), 615-620.

Rice, M. E., & Harris, G. T. (2002). Men who molest their sexually immature daughters: Is a special explanation required? *Journal of Abnormal Psychology, 111*(2), 329-339.

Richards, A., French, C. C., Johnson, W., & Naparstek, J. (1992). Effects of mood manipulation and anxiety on performance of an emotional Stroop task. *British Journal of Psychology, 83*(4), 479-491.

Ryan, F. (2002). Attentional bias and alcohol dependence: A controlled study using the modified Stroop paradigm. *Addictive Behaviors, 27*(4), 471-482.

Sachsenmaier, S. J., & Gress, C. L. Z. (in press). The Abel Assessment for Sexual Interests-2™ (AASI-2): A Critical Review. In D. Thornton & D. R. Laws (Eds.), *Cognitive Approaches to the Assessment of Sexual Interest in Sexual Offenders*.

Santtila, P., Mokros, A., Viljanen, K., Koivisto, M., Sandnabba, N. K., Zappalà, A., et al. (2009). Assessment of sexual interest using a choice reaction time task and priming: A feasibility study. *Legal and Criminological Psychology, 14*, 65-82.

Segal, Z. V., & Stermac, L. E. (1990). The role of cognition in sexual assault. In W. L. Marshall, D. R. Laws & H. E. Barbaree (Eds.), *Handbook of sexual assault: Issues, theories, and treatment of the offender.* (pp. 161-174). New York, NY, US: Plenum Press.

Sekuler, R., & Blake, R. (1994). *Perception* (3rd ed.). New York ; London: McGraw-Hill.

Seto, M. C. (2008). *Pedophilia and sexual offending against children: Theory, assessment, and intervention*. Washington, DC, US: American Psychological Association.

Seto, M. C., Cantor, J. M., & Blanchard, R. (2006). Child Pornography Offenses Are a Valid Diagnostic Indicator of Pedophilia. *Journal of Abnormal Psychology, 115*(3), 610-615.

Seto, M. C., Harris, G. T., Rice, M. E., & Barbaree, H. E. (2004). The Screening Scale for Pedophilic Interests Predicts Recidivism Among Adult Sex Offenders With Child Victims. *Archives of Sexual Behavior, 33*(5), 455-466.

Seto, M. C., & Lalumière, M. L. (2001). A brief screening scale to identify pedophilic interests among child molesters. *Sexual Abuse: Journal of Research and Treatment, 13*(1), 15-25.

Seto, M. C., Lalumiere, M. L., & Kuban, M. (1999). The sexual preferences of incest offenders. *Journal of Abnormal Psychology, 108*(2), 267-272.

Sharma, D., Albery, I. P., & Cook, C. (2001). Selective attentional bias to alcohol related stimuli in problem drinkers and non-problem drinkers. *Addiction, 96*(2), 285-295.

Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences, 44*(3), 535-551.

Singer, B. (1984). Conceptualizing sexual arousal and attraction. *Journal of Sex Research, 20*(3), 230-240.

Smith, P. (2006). *Cognitive bias in sexual offenders*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.

Smith, P. (in press). Assessing Sexual Interest with the Emotional Stroop Test. In D. Thornton & D. R. Laws (Eds.), *Cognitive Approaches to the Assessment of Sexual Interest in Sexual Offenders*.

Smith, P., & Waterman, M. (2003). Processing bias for aggression words in forensic and nonforensic samples. *Cognition & Emotion, 17*(5), 681.

Smith, P., & Waterman, M. (2004). Processing Bias for Sexual Material: The Emotional Stroop and Sexual Offenders. *Sexual Abuse: Journal of Research and Treatment, 16*(2), 163-171.

Smith, P., & Waterman, M. (2005). Sex Differences in Processing Aggression Words Using the Emotional Stroop Task. *Aggressive Behavior, 31*(3), 271-282.

Snowden, R., Wichter, J., & Gray, N. (2008). 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*, 558-565.

Spiering, M., & Everaerd, W. (2007). The sexual unconscious. In E. Janssen (Ed.), *The psychophysiology of sex.* (pp. 166-184). Bloomington, IN US: Indiana University Press.

Spiering, M., Everaerd, W., & Laan, E. (2004). Conscious Processing of Sexual Information: Mechanisms of Appraisal. *Archives of Sexual Behavior, 33*(4), 369-380.

Stormark, K. M., & Torkildsen, O. (2004). Selective processing of linguistic and pictorial food stimuli in females with anorexia and bulimia nervosa. *Eating Behaviors, 5*(1), 27-33.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. New York: Oxford University Press.

Tape, T. G. (1999). Interpreting diagnostic tests.   Retrieved 7th December, 2007, from http://gim.unmc.edu/dxtests/roc3.htm

Thornhill, R., & Palmer, C. T. (2000). *A natural history of rape: Biological bases of sexual coercion*. Cambridge, MA US: The MIT Press.

Trivedi, N., & Sabini, J. (1998). Volunteer Bias, Sexuality, and Personality. *Archives of Sexual Behavior, 27*(2), 181-195.

United States v. Powers, 59 1460 (4th Cir. 1995).

van Honk, J., Tuiten, A., de Haan, E., van den Hout, M., & Stam, H. (2001). Attentional biases for angry faces: Relationships to trait anger and anxiety. *Cognition & Emotion, 15*(3), 279-297.

van Honk, J., Tuiten, A., van den Hout, M., Koppeschaar, H., Thijssen, J., de Haan, E., et al. (1998). Baseline salivary cortisol levels and preconscious selective attention for threat: A pilot study. *Psychoneuroendocrinology, 23*(7), 741-747.

van Honk, J., Tuiten, A., van den Hout, M., Koppeschaar, H., Thijssen, J., de Haan, E., et al. (2000). Conscious and preconscious selective attention to social threat: Different neuroendocrine response patterns. *Psychoneuroendocrinology, 25*(6), 577-591.

van Honk, J., Tuiten, A., Verbaten, R., van den Hout, M., Koppeschaar, H., Thijssen, J., et al. (1999). Correlations among salivary testosterone, mood, and selective attention to threat in humans. *Hormones and Behavior, 36*(1), 17-24.

Walker, M. K., Ben-Tovim, D. I., Paddick, S., & McNamara, J. (1995). Pictorial adaptation of Stroop measures of body-related concerns in eating disorders. *International Journal of Eating Disorders, 17*(3), 309-311.

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

Ward, T., & Beech, A. (2006). An integrated theory of sexual offending. *Aggression and Violent Behavior, 11*(1), 44-63.

Ward, T., & Beech, A. (2008). An integrated theory of sexual offending. In D. R. Laws & W. T. O'Donohue (Eds.), *Sexual deviance: Theory, assessment, and treatment (2nd ed.).* (pp. 21-36). New York, NY US: Guilford Press.

Ward, T., & Brown, M. (2004). The Good Lives model and conceptual issues in offender rehabilitation. *Psychology, Crime & Law, 10*(3), 243-257.

Ward, T., Fon, C., Hudson, S. M., & McCormack, J. (1998). A descriptive model of dysfunctional cognitions in child molesters. *Journal of Interpersonal Violence, 13*, 129-155.

Ward, T., Gannon, T. A., & Keown, K. (2006). Beliefs, values, and action: The judgment model of cognitive distortions in sexual offenders. *Aggression and Violent Behavior, 11*(4), 323-340.

Ward, T., Hudson, S. M., Johnston, L., & Marshall, W. L. (1997). Cognitive distortions in sex offenders: An integrative review. *Clinical Psychology Review, 17*(5), 479-507.

Ward, T., & Keenan, T. (1999). Child molesters' implicit theories. *Journal of Interpersonal Violence, 14*(8), 821-838.

Ward, T., Keown, K., & Gannon, T. A. (2007). Cognitive distortions as belief, value and action judgements. In T. A. Gannon, T. Ward, A. Beech & D. Fisher (Eds.), *Aggressive Offenders' Cognition: Theory, Research and Practice*. Sussex, UK: Wiley.

Ward, T., Polaschek, D. L. L., & Beech, A. (2006). *Theories of Sexual Offending*. Sussex, UK: Wiley.

Waters, A. J., & Feyerabend, C. (2000). Determinants and Effects of Attentional Bias in Smokers. *Psychology of Addictive Behaviors, 14*(2), 111-120.

Waters, A. J., Sayette, M. A., & Wertz, J. M. (2003). Carry-over effects can modulate emotional Stroop effects. *Cognition & Emotion, 17*(3), 501-509.

Waters, A. J., Shiffman, S., Sayette, M. A., Paty, J. A., Gwaltney, C. J., & Balabanis, M. H. (2003). Attentional bias predicts outcome in smoking cessation. *Health Psychology, 22*(4), 378-387.

Wertz, J. M., & Sayette, M. A. (2001). Effects of Smoking Opportunity on Attentional Bias in Smokers. *Psychology of Addictive Behaviors, 15*(3), 268-271.

Whelan, R. (2008). Effective analysis or reaction time data. *Psychological Record, 58*(3), 475-482.

Wiederman, M. W. (1999). Volunteer Bias in Sexuality Research Using College Student Participants. *Journal of Sex Research, 36*(1), 59-66.

Wikström, J., Lundh, L.-G., & Westerlund, J. (2003). Stroop effects for masked threat words: Pre-attentive bias or selective awareness? *Cognition & Emotion, 17*(6), 827-842.

Williams, K. M. (2003). Two techniques for assessment of sexual interest: A discussion of the clinical utility of penile plethysmography and visual reaction time. *Forensic Examiner, 12*(1), 35-38.

Wright, L. W., Jr., & Adams, H. E. (1994). Assessment of sexual preference using a choice reaction time task. *Journal of Psychopathology and Behavioral Assessment, 16*(3), 221-231.

Wright, L. W., Jr., & Adams, H. E. (1999). The Effects of Stimuli That Vary in Erotic Content on Cognitive Processes. *Journal of Sex Research, 36*(2), 145-151.

Zeitlin, S. B., Bradburn, D. O., & Lawson-Kerr, K. (1995). Selective processing of epilepsy-related cues in patients with high fear of seizures. *Journal of Epilepsy, 8*(1), 16-22.

Appendix 1: Publication, presentations etc. arising from the present work

Book Chapter

Ó Ciardha, C., & Gormley, M. (2009). Comparing two implicit cognitive measures of sexual interest: a pictorial modified Stroop task and the Implicit Association Test. In D. Thornton & D. R. Laws (Eds.), *Cognitive Approaches to the Assessment of Sexual Interest in Sexual Offenders*. Chichester, UK: Wiley.

Award

Graduate Research Award for paper titled: *Using a pictorial modified Stroop task and two Implicit Association Tests in the exploration of sexual interests and associated cognitions of sexual offenders against children.* Awarded at the 28th annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Dallas, TX.

Conference Presentations

Ó Ciardha, C. & Gormley, M. (2009, October). Developing clinically useful implicit tasks of sexual interest: What are they measuring? In C. Ó Ciardha (Chair), *Indirect measures of Assessment: clinical applications and theoretical implications*. Symposium conducted at the 28th annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Dallas, TX.

Thornton, D., McKee, R & Ó Ciardha, C. (2009, October). Convergent and divergent validity of implicit measures of sexual interest among SVP. In C. Ó Ciardha (Chair), *Indirect measures of Assessment: clinical applications and theoretical implications*. Symposium conducted at the 28th annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Dallas, TX.

Ó Ciardha, C. & Gormley, M. (2008, October). The use of a pictorial modified Stroop task and two Implicit Association Tests in the assessment of sexual interest among sexual offenders against children. In D.R. Laws (Chair), *Attention-based assessment: an international perspective*. Symposium conducted at the 27th

annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Atlanta, GA.

Poster Presentations

Ó Ciardha, C., & Gormley, M. (2008). *Assessment of sexual interest: comparisons between a pictorial modified Stroop task and the Implicit Association Test*. Poster presented at the Annual Conference of the British Psychological Society, Dublin, Ireland.

Ó Ciardha, C., & Gormley, M. (2007). *Utility of Two Implicit Measures of Sexual Interest and Sexual Associations*. Poster presented at the 26th annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Diego, CA.

Appendix 2: Letters granting ethical approval from the School of Psychology, St John of God Hospitaller Service and the Irish Prison Service

School of Psychology
University of Dublin, Trinity College
Dublin 2, Ireland

Tel: +353 1 608 1886
Fax: +353 1 671 2006

F.A.O: Caoilte ÓCiardha

School of Psychology
Ethics Committee

15 June, 2006

Dear Caoilte,

The School of Psychology Ethics Committee met recently to consider your application entitled 'Evaluating the utility of implicit cognitive measures in the assessment of sexual offenders against children'.

I am pleased to inform you that the Committee has approved your application, subject to you making the following changes.

- Typos in information sheet.
- Providing debriefing.
- Subject to approval of Granada and a letter of agreement form them.

Your's sincerely,

Kevin Tierney
Chairperson,
School of Psychology Ethics Committee

http://www.tcd.ie/Psychology

214

St. John of God Hospitaller Services

Provincial Administration
Programme Development Department
Stillorgan Co. Dublin Ireland
Tel: (01) 2771500 Fax: (01) 2783024
e-mail: programmes@sjog.ie   web: www.sjog.ie

20 October 2006

Mr Caoilte O'Ciardha
Trinity College Dublin
School of Psychology
Aras an Phiarsaigh
Trinity College
Dublin 2

Re: Proposal - *"Evaluating the utility of implicit cognitive measures in the assessment of sexual offenders against children"*

Dear Caoilte,

Thank you for attending the meeting of the Provincial Ethics Committee on Tuesday, 17th October and presenting the above proposal.
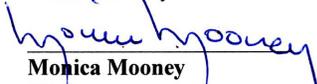
The Provincial Ethics Committee is happy to approve your proposal pending receipt of the following amendments:

- Letters of Information and Consent Forms for the Granada Institute clients and the control group printed on relevant letterhead (TCD and Granada Institute) including statements advising both groups of the participation of the other (Granada Institute participation to remain anonymous).
- Access to Files: Confirmation that the process of gathering of data will ensure anonymity of clients.

Please forward your amendments for the attention of Jackie Hall at the address above. Please note there is no need to return the entire application again. Only the amendments, clearly highlighted are required. Following receipt of these amendments the Provincial Ethics Committee will be in a position to formally approve this study. If you have any further queries please do not hesitate to contact Jackie Hall on 01 277 1500 Ext 1694.

Kind regards

Yours sincerely

Monica Mooney
Chair
Provincial Ethics Committee

The Order, founded in 1539, is a Catholic organisation providing health care, special education, rehabilitation and welfare services in 50 countries. The Province operates services in Ireland, New Jersey USA, and Malawi, Central Africa.

IRISH PRISON SERVICE

Seirbhís Phríosúin na hÉireann
Ceanncheathrú
Lárionad Gnó an IDA
Bóthar Bhéal Átha na Lao
An Longfort
Co. an Longfoirt

Website/Láithreán gréasáin: www.irishprisons.ie
Email/Ríomhphoist: info@irishprisons.ie

Telephone/Teileafón: (043) 35100
Fax/Facs: (043) 35371

Caoilte Ó Ciardha
School of Psychology
Áras an Phiarsaigh
Trinity College
Dublin 2

08 October 2008

Re.   Research Proposal 'Evaluating the utility of implicit cognitive measures in the assessment of sexual offenders against children'.

Dear Caoilte,

Your application to carry out the above research project was considered by the Prisoner Based Research Ethics Committee a number of times and by the Director General of the Irish Prison Service.

I am pleased to advise you that the research application is approved, and I attach some comments from four of the Committee members that you might find useful, X's and Y's comments in particular.

Please provide a copy of this letter to the Governor of the institution/s that you wish to visit as entry to the prison/institution for the purpose of the study is contingent on the agreement of individual governors and appropriate security clearance.

* Please note that the Governor must be contacted in advance of your proposed attendance at the prison/institution.

The IPS would be interested in seeing the outcome of this research.

Yours sincerely,

Séamus Beirne
Secretary
Prisoner Based Research Ethics Committee

Appendix 3: Questionnaire given to all participants in study 2 (chapter 3)

Note: the answers you give below are confidential and will be matched up with your results on the computerised test only through the use of a number. The experimenter will not know whose answers these are.

1. What age are you? _____

2. Which of the following most accurately describes your sexual preference (tick one box only)

    a) Exclusive preference for males   ☐

    b) Prefer men but have some sexual attraction towards females   ☐

    c) Equally attracted to both sexes   ☐

    d) Prefer females but have some sexual attraction towards males   ☐

    e) Exclusive preference for females   ☐

    f) Other preference. Please specify _____   ☐

Appendix 4: Questionnaire given to Granada Institute participants in study 3 (chapter 4)

General Information Questionnaire

---

1) What age are you? _____

2) What is the highest level of education you have achieved?

Did not finish Primary School	□
Finished Primary School	□
Finished Junior Certificate or equivalent (Inter Cert, GCSE's etc.)	□
Finished Leaving Certificate or equivalent (A-levels etc)	□
Completed higher education diploma/certificate	□
Completed bachelors/undergraduate degree	□
Completed postgraduate degree/diploma	□

Other (please specify) _____
_____
_____

---

3) To your knowledge, are you colour blind?   Yes   No
                                                  □   □

---

4) How would you describe your sexual interest towards adults? (TICK MORE THAN ONE IF APPLICABLE)

No sexual interest in adults	□
Some sexual interest in female adults	□
Strong sexual interest in female adults	□
Some sexual interest in male adults	□
Strong sexual interest in male adults	□

b) Roughly how many, if any, different adult sexual partners have you had?

Males   _____
Females   _____

c) How long was the longest sexual relationship that you had with an adult (leave blank if no relationships)? _____

5) How would you describe your sexual interest towards children? (TICK MORE THAN ONE IF APPLICABLE)

No sexual interest in children	☐
Some sexual interest in female children	☐
Strong sexual interest in female children	☐
Some sexual interest in male children	☐
Strong sexual interest in male children	☐

b) If you've described any sexual interest in children, what age of child are you typically interested in? _____

6) How long have you been attending group therapy in Granada? _____

7) Before joining the group did you...

attend Granada for individual therapy	☐
attend Granada for assessment	☐
attend Granada for assessment and therapy	☐
attend somewhere else for assessment or therapy	☐
none of the above	☐

b) If you've received any individual therapy, how long did this last?

_____

8a) What type of abuse were you involved in carrying out? (TICK MORE THAN ONE IF APPLICABLE AND LEAVE BLANK ANY QUESTIONS THAT DON'T APPLY)

Abuse where physical contact occurred	☐
Abuse where no physical contact occurred	
Child pornography	☐
Exhibitionism/exposing	☐

b) In the case of any times where physical contact occurred:

What was the gender of the child/children?
Male	☐
Female	☐
Both	☐
How old was/were the child/children?	_____
Were you related to the child/children	
Yes	☐
No	☐

219

c) In the case of <u>viewing images or other material containing child abuse</u> (child pornography)

What gender were the children in the material?
- Male   ☐
- Female   ☐
- Both   ☐

What ages (or age ranges) were the children?   _____

d) In the case of <u>exhibitionism/exposing</u>

What gender was/were the child/children?
- Male   ☐
- Female   ☐
- Both   ☐

What age(s) was/were the child/children?   _____

Were you related to the child/children
- Yes   ☐
- No   ☐

Appendix 5: Questionnaire given to Arbour Hill prison participants in study 3 (chapter 4)

---

1) What age are you? _____

---

2) What is the <u>highest level</u> of education you have achieved?

Did not finish Primary School	☐
Finished Primary School	☐
Finished Junior Certificate or equivalent (Inter Cert, GCSE's etc.)	☐
Finished Leaving Certificate or equivalent (A-levels etc)	☐
Completed higher education diploma/certificate	☐
Completed bachelors/undergraduate degree	☐
Completed postgraduate degree/diploma	☐

Other (please specify) _____
_____

---

3) To your knowledge, are you colour blind?　　　Yes　　No
　　　　　　　　　　　　　　　　　　　　　　　　　　☐　　　☐

---

4) How would you describe your sexual interest towards adults? (TICK MORE THAN ONE IF APPLICABLE)

No sexual interest in adults	☐
Some sexual interest in female adults	☐
Strong sexual interest in female adults	☐
Some sexual interest in male adults	☐
Strong sexual interest in male adults	☐

5) Roughly how many, if any, different adult sexual partners have you had?

　　　Males　　_____
　　　Females　_____

6) How long was the longest sexual relationship that you had with an adult (leave blank if no relationships)?　　_____

---

7) How would you describe your sexual interest towards children? (TICK MORE THAN ONE IF APPLICABLE)

No sexual interest in children	☐
Some sexual interest in female children	☐
Strong sexual interest in female children	☐
Some sexual interest in male children	☐
Strong sexual interest in male children	☐

8) If you've described any sexual interest in children, what age of child are you typically interested in?
_____

The next questions are regarding the <u>sentence you are currently serving</u>.

TICK MORE THAN ONE IF APPLICABLE AND LEAVE BLANK ANY QUESTIONS THAT DON'T APPLY

9) What type of offence involving a child/children were you convicted of carrying out?

Abuse where physical contact occurred	□
Child pornography offences	
Viewing/downloading pornography	□
Producing pornography	□

10) What sentence were you given? _____

---

11) In the case of <u>any times where physical</u> contact occurred:

What was the gender of the child/children?

Male	□
Female	□
Both	□

How old was/were the child/children? _____

Were you related to the child/children?

Yes (father/step-father/guardian)	□
Yes (other relation)	□
No	□
Both related and unrelated victims	□

Approximately how old were you at the time of the offence(s) _____

Did the offence you were convicted of involve:

	Yes	No
Sexual activity with the same child on more than one occasion?	□	□
Sexual activity with different children on separate occasions?	□	□
Filming or otherwise recording the contact?	□	□

---

13) In the case of <u>viewing/downloading images or other material containing child abuse</u> (child pornography)

What gender were the children in the material?

Male	□
Female	□
Both	□

What ages (or age ranges) were the children? _____

---

14) Have you ever received treatment or therapy as a result of offending against children

	Yes	No
	□	□

If yes, how long did you receive therapy for? _____

---

15) Does your current conviction involve of any sexual offences against adults?

	Yes	No
	□	□

If yes, were the victims

Male	□
Female	□
Both	□

Appendix 6: Questionnaire given to control participants in study 3 (chapter 4)

General Information Questionnaire

1) What age are you? _____

2) What is the highest level of education you have achieved?

Did not finish Primary School	☐
Finished Primary School	☐
Finished Junior Certificate or equivalent (Inter Cert, GCSE's etc.)	☐
Finished Leaving Certificate or equivalent (A-levels etc)	☐
Completed higher education diploma/certificate	☐
Completed bachelors/undergraduate degree	☐
Completed postgraduate degree/diploma	☐

Other (please specify) _____
_____
_____

3) To your knowledge, are you colour blind?   Yes   No
                                                  ☐   ☐

4) How would you describe your sexual interest towards adults? (TICK MORE THAN ONE IF APPLICABLE)

No sexual interest in adults	☐
Some sexual interest in female adults	☐
Strong sexual interest in female adults	☐
Some sexual interest in male adults	☐
Strong sexual interest in male adults	☐

b) Roughly how many, if any, different adult sexual partners have you had?

Males   _____
Females   _____

c) How long was the longest sexual relationship that you had with an adult (leave blank if no relationships)?   _____

4) Have you ever been convicted of a sexual offence involving a child?   Yes   No
                                                                            ☐   ☐

Appendix 7: Questionnaire given to participants in study 5 (chapter 5)

General Information Questionnaire

1) What age are you? _____

2) What is the highest level of education you have achieved?

Did not finish Primary School   □
Finished Primary School   □
Finished Junior Certificate or equivalent (Inter Cert, GCSE's etc.)   □
Finished Leaving Certificate or equivalent (A-levels etc)   □
Completed higher education diploma/certificate   □
Completed bachelors/undergraduate degree   □
Completed postgraduate degree/diploma   □
Other (please specify) _____
_____
_____

3) To your knowledge, are you colour blind?   Yes   No
                                                 □   □

4) How would you describe your sexual interest towards adults? (TICK MORE THAN ONE IF APPLICABLE)

No sexual interest in adults   □
Some sexual interest in female adults   □
Strong sexual interest in female adults   □
Some sexual interest in male adults   □
Strong sexual interest in male adults   □

Appendix 8: Questionnaire given to participants in study 6 (chapter 6)

General Information Questionnaire

1) What age are you? _____

2) What is the <u>highest level</u> of education you have achieved?

Did not finish Primary School	☐
Finished Primary School	☐
Finished Junior Certificate or equivalent (Inter Cert, GCSE's etc.)	☐
Finished Leaving Certificate or equivalent (A-levels etc)	☐
Completed higher education diploma/certificate	☐
Completed bachelors/undergraduate degree	☐
Completed postgraduate degree/diploma	☐

Other (please specify) _____
_____
_____

3) To your knowledge, are you colour blind?   Yes   No
                                                 ☐   ☐

4) How would you describe your sexual interest towards adults? (TICK MORE THAN ONE IF APPLICABLE)

No sexual interest in adults	☐
Some sexual interest in female adults	☐
Strong sexual interest in female adults	☐
Some sexual interest in male adults	☐
Strong sexual interest in male adults	☐

b) Roughly how many, if any, different adult sexual partners have you had?

Males   _____
Females   _____

c) How long was the longest sexual relationship that you had with an adult (leave blank if no relationships)?   _____

Appendix 9: Sample pictorial stimuli

Adult Male

Adult Female

Large Cat

Child Male (from the NRP set)