Running Head: COMPARING P-MST, IAT, AND CRT

Measuring sexual interest using a pictorial modified Stroop task, a pictorial Implicit Association Test, and a Choice Reaction Time task

Caoilte Ó Ciardha[1]
[surname: Ó Ciardha]
and
Michael Gormley[1]
[1]School of Psychology, University of Dublin, Trinity College, Dublin, Ireland

Correspondence concerning this article should be addressed to Dr. Caoilte Ó Ciardha, School of Psychology, Keynes College, University of Kent, Canterbury, CT2 7NP.
E-Mail: C.C.OCiardha@kent.ac.uk

Abstract

Tasks that can successfully measure sexual interest have utility in forensic settings. Prior to use with problematic sexual interest however, work is needed in validating such tasks. This study focused on the measurement of non-deviant sexual interest. Eleven gay and fourteen straight participants each completed a pictorial Implicit Association Test (IAT), a pictorial modified Stroop task (P-MST) and a Choice Reaction Time (CRT) task. Each task was designed to tap into the sexual interest of participants. Stimuli were of males and females in bathing suits along with control images and sexual and non-sexual words. The IAT was most successful in differentiating between gay and straight participants. The P-MST also performed well, though the task's position in the battery of tasks seemed to affect the results. The CRT tasks did not successfully show group differences. Theoretical and methodological implications of the effectiveness of the three tasks in tapping into sexual interest are discussed.

Keywords

Indirect measures

Sexual interest

Implicit Association Test

Pictorial modified Stroop task

Choice Reaction Time task

Introduction

Recently there has been increased interest in the application of implicit or indirect tasks to measure different psychological phenomena. Methods such as the Implicit Association Test (Greenwald, McGhee, & Schwartz, 1998) and the modified Stroop task have been used to explore topics such as racial bias, phobias, addiction and many more. Such tasks may provide insight into the cognitive processes involved in these phenomena. Furthermore tasks of this nature may prove to be useful clinical tools for the assessment of individuals for whom phenomena such as phobias, racial bias etc. are problematic. Individuals researching sexual offending are increasingly taking using such indirect cognitive tasks (e.g. Banse, Schmidt, & Clarbour, 2010; Giotakos, 2005; Gray, Brown, MacCulloch, Smith, & Snowden, 2005; Gress, 2005; Mihailides, Devilly, & Ward, 2004; Mokros, et al., 2010; Nunes, Firestone, & Baldwin, 2007; Ó Ciardha & Gormley, in press; Price & Hanson, 2007; Smith & Waterman, 2004; Yoon & Knight, 2011).

This study takes three paradigms and compares how well they discriminate between males with gay and straight sexual interests. The study provides a starting point for further methodological refinement of such tasks for applied use and allows us to discuss some theoretical implications. The three tasks compared are: the Implicit Association Test (IAT), the pictorial modified Stroop task (P-MST) and the Choice Reaction Time (CRT) task. Versions of the IAT (Banse, et al., 2010; Brown, Gray, & Snowden, 2009; Gray, et al., 2005; Mihailides, et al., 2004; Nunes, et al., 2007; Ó Ciardha & Gormley, 2009), CRT (Giotakos, 2005; Mokros, Dombert, Osterheider, Zappalà, & Santtila, 2010; Santtila, et al., 2009; Wright & Adams, 1994, 1999) and P-MST (Ó Ciardha & Gormley, in press) tasks have been shown to tap into sexual interest or related cognitions for offenders and non-offenders. While these three tasks are similar, it is important to explore the differences between them in order to start to build up a fuller understanding of the cognitive processes engaged and, ultimately, the utility of the tasks for clinical use.

Several studies have found that an IAT can tap into associations related to offending (Banse, et al., 2010; Gray, et al., 2005; Mihailides, et al., 2004; Nunes, et al., 2007). The task compares reaction times to concept pairs that the participant may hold as congruous with reaction times to concept pairs that they may hold as incongruous. For example, a child molester may have faster reaction times when the concepts of *child* and *sex* are paired in comparison to when *adult* and *sex* are paired. This would suggest an association of children and sex in the offender's mind. However, from this task it would not be clear whether this

association was driven by a sexual interest in children or an association more related to cognitive distortions such as Ward and Keenan's (1999) *children as sexual beings* implicit theory. Another alternative explanation could involve the possible childhood sexual victimisation of the participant. There is therefore a clear need for a dialogue in the literature regarding what is being measured by the IAT and how methodological manipulations (in terms of stimuli, procedure etc) may impact on that.

The P-MST task and the CRT task are somewhat different to the IAT in that both interpret a slowing down of reaction times as potentially indicative of cognitive biases. This is in contrast to the speeding up associated with the IAT. The P-MST asks participants to identify, as quickly as possible, the colour in which a stimulus is presented. Any systematic slowing down of responses to a given stimulus category is interpreted as an attentional bias towards that category. The P-MST was shown to have potential in tapping into the sexual interest of non-offenders and in identifying group differences between offenders hypothesised to show high deviance and those expected to show low deviance (Ó Ciardha & Gormley, in press). While interpreted by Ó Ciardha and Gormley (in press) as an indirect measure of sexual interest or arousal it must be stressed that as with the IAT, alternative explanations are possible. Images may have been salient for other reasons. Interestingly word modified Stroop tasks (i.e. where stimuli are words, whose meaning is to be ignored, written in different coloured fonts) have found that both perpetrators (Price & Hanson, 2007; Smith & Waterman, 2004) and victims (Dubner & Motta, 1999; Foa, Feske, Murdock, Kozak, & McCarthy, 1991) of sexual violence may respond to offence related words in a similar manner.

With the CRT task 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). As with the P-MST, a systematic slowing down of reaction times to a given category of stimulus is interpreted as indicating an attentional bias towards this stimulus category. 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, 2011; Mokros, et al., 2010). Wright and Adams (1994, 1999) 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.

Gress (2011) did not find clear results indicating whether her CRT was tapping into sexual interest. Methodological issues and the heterogeneity of the offending sample may

have contributed to the inability to find a clear effect of sexual interest. Giotakos (2005) on the other hand found that extrafamilial sexual offenders against children 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. Mokros et al (2010) demonstrated an excellent ability in discriminating between those with paedophilic and non paedophilic interest using a similar methodology to the original Wright study. As with the P-MST, it is possible that the stimuli used in the CRT tasks were capable of capturing attention for reasons other than sexual interest. However, given the consistent findings across studies, the task seems to be relatively reliable and certainly promising in tapping into the cognitions related to sexual interest.

We acknowledge that exploring the use of indirect tasks using gay and straight sexual interest rather than deviant sexual interest limits somewhat the applicability of any findings to the assessment of offenders. However, in order to develop tasks with the potential to assess individual offenders with demonstrable validity sufficient to allow for their use in informing decisions about treatment, parole, risk etc it is important that those tasks be supported by a body of knowledge that speaks clearly to what constructs are being measured and which methodologies are optimum. As the number of offenders accessible for research participation is often limited, some research questions that are highly relevant to forensic researchers may first be investigated using normal sexual interest as an, admittedly imperfect, template for how tasks may later be used with deviant sexual interest.

Each of the previous CRT studies used a random presentation of images. Ó Ciardha and Gormley (2011) demonstrated that a blocked pictorial modified Stroop design yielded a larger effect size and better classified participants in terms of their sexual orientation relative to a randomised presentation of stimuli. This blocked approach, whereby all images of a certain category (e.g. adult females) are grouped together, was adopted with both the P-MST and the CRT used in this study. Clearly this is a departure from the randomised design of previous CRT studies but was included in order to keep the procedure for both tasks as similar as possible (and to test the utility of a blocked CRT design).

It was hypothesised that the P-MST and the CRT 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 results of the pictorial IAT would be significantly and negatively correlated with the P-MST and the CRT task. It was expected that the strength of the relationship between the three measures would inform discussion regarding the cognitive processes involved in each.

Method

Design

A mixed factorial design was used with gay and straight participants completing all experimental tasks; the P-MST, the pictorial IAT and the CRT task.

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 psychology students 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. All participants indicated on a questionnaire that they had an exclusive sexual interest in males or females. 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).

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 five buttons arranged in the shape of a quincunx (like the five spots used to represent a five on a die). Tasks were run using the SuperLab4© stimulus presentation software. A questionnaire was administered to determine age, education, color blindness and asked whether they had strong, some, or no sexual interest in male and female adults

Images for all tasks were of male and female adults in bathing suits. The P-MST and CRT tasks also included control images 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). Identical images were used in the CRT and P-MSTs with different images being used in the IAT. For the P-MST images, images were adapted using the *colour*

replacement brush using Adobe Photoshop Elements 3.1©, yielding red, green, blue and yellow images.

IAT word stimuli were generated by compiling a list of previously used words (e.g. Banse, Schmidt, & Clarbour, 2010; Gray, et al., 2005; Mihailides, et al., 2004; Nunes, et al., 2007; Smith & Waterman, 2004) and expanding on those lists using the authors' own judgement and collaboration with other researchers carrying out similar research. Words were then excluded from that list on the grounds of word length, frequency and also on *a priori* grounds. While many words could conceivably be considered non sexual we selected words that formed a coherent group of words that could be used to describe an unattractive or nonsexual person. 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. There were six female and six male images used in the task.

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 (Wright & 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.

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 P-MST, a pictorial IAT and a CRT task. The IAT was always completed second with the order of the P-MST 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 P-MST. 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.

Previous P-MSTs exploring sexual interest (Ó Ciardha & Gormley, in press) included a traditional Stroop task at the start of the procedure. This was intended to train the participants how to carry out the task and also as a measure whereby performance on this could be compared across groups. However, in this study, in order to be as methodologically similar to the CRT task, the traditional Stroop task was removed from the beginning of the task. 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 P-MST. 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 P-MST and included the same feedback for incorrect responses for practice trials only

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 paired). If participants answered incorrectly on any trial 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.

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.

Treatment of data

In analysing the results of the CRT task and of the P-MST a method of data treatment based on ipsative z-scores was adopted to minimise any potential influence of differences in cognitive speed, as a result of the significant difference in age or other factors. Ó Ciardha (2010) found that this method is likely to moderate the effects of cognitive speed on the results of a P-MST. P-MST and CRT task data were treated identically. For each participant, an ipsative *z*-score was calculated for each of the experimental conditions (adult female, adult male and cat) by subtracting the mean overall reaction time to all images from the mean reaction time for each block and dividing by the overall standard deviation for reaction times to all images. Each individual's 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 percentiles. Negative values indicated trial type means that were quicker than the grand mean while positive scores were slower. Analyses were also carried out using mean reaction times and the results of these analyses are reported where findings differ from the z-score results.

There were many different ways in which the IAT data could be analysed. The developers of the IAT suggest removing extreme values based on certain thresholds, adding error latencies for incorrect responses (or using time to a correct response in designs where the participant must correct their mistakes before continuing) and calculating a difference score between experimental blocks, taking into account the standard deviation of those blocks (Greenwald, Nosek, & Banaji, 2003). However it was decided to use an ipsative *z*-score approach to analysing the data in order to maximise the ability of the study to compare results across the IAT, P-MST, and the CRT task. As with the other tasks in the study, different methods of analysis were used, including different iterations of the Greenwald et al. method, but only reported if they differed from the ipsative *z*-score findings. 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). Outliers were removed using the same criteria as for the P-MST and the CRT task.

Results of the P-MST

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 other measures of central tendency, i.e. means with outliers removed, found a statistically significant interaction, so it seems safe to assume that this lack of significance using the ipsative 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 presentation is taken into account, by carrying out two separate two-way ANOVAs, the interaction reported above is only found when the P-MST was presented last. There was no significant interaction when the P-MST was presented before the IAT and CRT. Order of block presentation within the P-MST does not have significant impact on response times, though response to the last presented block are typically faster than the other two. An ROC analysis revealed that a difference score calculated from the response times to male and female images from the P-MST 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 .917 found by Ó Ciardha and Gormley (in press) using a similar design to compare the reaction times of gay and straight male participants. However, when order of task is taken into account, the difference score when the P-MST is completed first yields an ROC with poor discriminant ability, AUC = .571, p=.685, SE=.179, while the difference score when P-MST is completed last has an excellent (Tape, 1999) discriminant ability, AUC = .905, p=.015, SE=.086.

Results of the Choice Reaction Time task

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[1]) = 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 and this is clearly depicted in Figure 1.

INSERT FIGURE 1 APPROXIMATELY HERE

Looking separately at CRT results for each order of presentation (i.e. whether the CRT task was presented as the first or third experimental task) did not yield any interaction between orientation and gender of trial. However, there was evidence for an influence of the order in which stimulus blocks were presented within the task. A one-way repeated measures ANOVA found a significant effect of block order, F (2, 48) = 9.463, $p < .001$, partial η^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 η^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

[1] 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

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.

INSERT FIGURE 2 APPROXIMATELY HERE

Results of the Implicit Association Test

Using ipsative *z*-scores for the IAT results, 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 this analysis the first trial type consisted of trials that were consistent with a heterosexual sexual orientation, i.e. trials where a word or image was presented that the participant had to classify using either the *female/sexual* button or the *male/nonsexual* button. The second trial type consisted of trials consistent with a homosexual sexual orientation, where the participant classified words or images using the *male/sexual* or the *female/nonsexual* buttons. The interaction between these trial types is plotted in Figure 2. After calculating a difference score using the ipsative *z*-scores, an ROC analysis revealed that the IAT had an excellent and statistically significant ability to discriminate between gay and straight participants, AUC = .974, $p <.001$, $SE = .026$. Greenwald et al.'s (2003) algorithms and other methods yielded similarly significant AUC values.

Relationship between the three tasks

There was no correlation between difference scores on the P-MST and the CRT task. Furthermore 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. P-MST response times overall (Mean = 727.63, SD = 192.46) were slower than overall response times 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 ipsative *z*-scores for the P-MST and IAT, $r = .448$, $p = .025$, two-tailed.

Two forced entry binary logistic regressions were carried out to assess whether a combination of both the IAT and P-MST scores yielded a better categorisation of participants as either gay or straight than either did individually. For the first regression, the IAT difference score served as the predictor variable in the first block and the P-MST 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 P-MST score in terms of categorisation of individuals based on their sexual orientation (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 P-MST (R^2 = .36). The combination of both predictor variables yielded a 17% increase in total variance explained which accounted for the total variance in the orientation variable (R^2 = 1). The IAT score accounted for 64% of unique variance while the P-MST accounted for 17%. Nineteen percent of the variance in categorisation was attributable to overlapping variance among the IAT and P-MST. The addition of the P-MST 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.

Discussion

Results indicated that orientation did not seem to have much of an impact on response times in the CRT task. There was no interaction between trial type and orientation and a 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. Results suggested that sexual orientation may have moderated the impact of order, though this conclusion is made cautiously. Sexual orientation/interest had a clearer effect on the P-MST. There was an interaction between trial type and orientation and a difference score based on the P-MST results for male and female images demonstrated a fair (Tape, 1999) ability to discriminate between gay and straight participants. However, this discriminant ability was not as accurate as had been found by Ó Ciardha and Gormley (in press). Order of task presentation seemed to affect the strength of the modified Stroop effect, in that when the task was presented last it produced the most robust effects. The IAT used in this study

demonstrated a clear interaction between trial type and sexual orientation. In addition the IAT difference score had an excellent ability to discriminate straight and gay participants. Response times to the IAT and P-MST both correlated significantly and when both were combined in a logistic regression model, the model correctly classified all participants in terms of their sexual orientation. Neither the P-MST 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 P-MST 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 P-MST was the last task presented is interesting. Any interpretation of this finding must be made cautiously given the reduction in power produced by splitting the sample into those who had been administered the P-MST first or last. In accepting that caveat, we surmise that 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 in other studies 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 P-MST, that the participant presented with the P-MST last were quicker to master this and therefore had less noise in their data. However, there was no significant difference in overall response time between participants where the P-MST 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 by Ó Ciardha and Gormley (in press) when the P-MST 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 pictorial modified Stroop effect when the participant is 'depleted' rather than fresh. Further investigation using a larger sample will allow the exploration of this finding in more detail. It should be noted that order of task presentation did not improve performance on the CRT task when presented last.

As mentioned, there was no clear evidence that the CRT task was able to tap into sexual interest though 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 (Mokros, et al., 2010; 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 P-MST. The P-MST produced significantly longer reaction times across trials than did the CRT task. This indicates that the P-MST 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 P-MST 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 P-MST 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 P-MST 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 P-MST 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 by Ó Ciardha and Gormley (2011) for a randomised P-MST. In addition, in the current study, the P-MST was able to identify group differences based on sexual orientation but the CRT wasn't. These results taken together suggest that the P-MST and the CRT are not measuring the same processes. McKenna and Sharma (2004) make the distinction between fast and slow components in modified Stroop designs where fast effects refer to the instantaneous grabbing of attention that might occur on the presentation of a single stimulus. Slow effects on the other hand refer to a slow down that occurs between and across trials possibly related to the participants' rumination of the stimulus categories. While slow effects appear to drive the results of the P-MST (Ó Ciardha & Gormley, 2011), they appear not to be present (or are not as pronounced) in the CRT task. While a randomised CRT task and a blocked P-MST 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 fast effects related to the process of sexual arousal and that the P-MST is better suited to capturing slow effects.

While performing at a level that would be considered excellent (Tape, 1999), the IAT did not correctly classify participants as gay or straight in every case. Similarly the P-MST 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 P-MST results 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, 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 P-MST 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. Further validation and manipulation of the tasks is therefore required to understand the relationship of the tasks to underlying cognitive processes and to establish their utility as clinical tools.

The ultimate goals of research using indirect measures to assess sexual interest is to develop tools with diagnostic utility for the assessment of individuals who's sexual behaviour violates society's norms and in addition to develop an understanding of the cognitive processes involved. The current study identifies key methodological considerations and explores some theoretical issues surrounding sexual interest-related cognition. However, there are clear limitations to using normal variability in sexual interest in adults as a cognate for deviant sexual interest. There have been several recent studies which have been demonstrated very promising results in using indirect tasks in the assessment of deviant sexual interest (especially Mokros, et al., 2010). There have however also been null findings adopting similar methodologies (e.g. Gress, 2011) which indicates the need to comprehensively explore the methodological and theoretical underpinnings of tasks

measuring cognitive responses to sexually salient stimuli. This study offers part of the picture on those underpinnings. These conclusions may inform further similar research that focuses on deviant sexual interest in offending populations. However given difficulties in accessing offending populations in sufficient numbers to carry out fine-grained adjustments of these indirect measures to explore theoretical and methodological questions, it is likely that the field will still rely on using non-deviant sexual interest to address certain questions. Fortunately it is possible to additionally use methods of offending proclivity (e.g. Gannon & O'Connor, 2011) to begin to build a fuller picture of sexual interest in offending along with providing an evidence-base for the validity of indirect measures of assessment. Thus, future studies should continue to explore the processing of both deviant and non-deviant sexual stimuli in non-offending populations while seeking out opportunities to test the most robustly validated paradigms with offending populations.

References

Banse, R., Schmidt, A. F., & Clarbour, J. (2010). Indirect Measures of Sexual Interest in Child Sex Offenders: A Multimethod Approach. *Criminal Justice and Behavior, 37*(3), 319-335. doi:10.1177/0093854809357598

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.

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. doi:10.1177/1079063209332234

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.

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.

Gannon, T. A., & O'Connor, A. (2011). The Development of the Interest in Child Molestation Scale. *Sexual Abuse: A Journal of Research and Treatment*. doi:10.1177/1079063211412390

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.

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. doi:10.1037/0021-843k114.2.304

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., Anderson, J. O., & Laws, D. R. (2011). Delays in attentional processing when viewing sexual imagery: The development and comparison of two measures. *Legal and Criminological Psychology*. doi:10.1111/j.2044-8333.2011.02032.x

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.

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.

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

Mokros, A., Dombert, B., Osterheider, M., Zappalà, A., & Santtila, P. (2010). Assessment of pedophilic sexual interest with an attentional choice reaction time task. *Archives of Sexual Behavior, 39*(5), 1081-1090. doi:10.1007/s10508-009-9530-6

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. doi:10.1177/0093854806291703

Ó Ciardha, C., & Gormley, M. (2011). *A comparison of blocked, clustered and randomly presented sexually salient stimuli in a pictorial modified Stroop task*. Manuscript in preparation.

Ó Ciardha, C. (2010). *The use of implicit cognitive measures in the assessment of sexual offenders (Doctoral thesis).* University of Dublin, Trinity College, Dublin Ireland.

Ó Ciardha, C., & Gormley, M. (in press). Using a pictorial modified Stroop task to explore the sexual interests of sexual offenders against children. *Sexual Abuse: A Journal of Research and Treatment*.

Price, S. A., & Hanson, R. K. (2007). A modified Stroop task with sexual offenders: Replication of a study. *Journal of Sexual Aggression, 13*(3), 203-216. doi:10.1080/13552600701785505

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.

Smith, P., & Waterman, M. (2004). Processing Bias for Sexual Material: The Emotional Stroop and Sexual Offenders. *Sexual Abuse: A Journal of Research and Treatment, 16*(2), 163-171. doi:10.1023/B:SEBU.0000023064.54397.d6

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

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

Wright, L. W., & 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., & Adams, H. E. (1999). The effects of stimuli that vary in erotic content on cognitive processes. *Journal of Sex Research, 36*(2), 145-151.

Yoon, J., & Knight, R. A. (2011). Sexual Material Perception in Sexually Coercive Men: Disattending Deficit and Its Covariates. *Sexual Abuse: A Journal of Research and Treatment, 23*(2), 275-291. doi:10.1177/1079063210391104