

Running Head: PICTORIAL MODIFIED STROOP TASK

Using a pictorial modified Stroop task to explore the sexual interests of sexual
offenders against children

Caoilte Ó Ciardha

and

Michael Gormley

School of Psychology, University of Dublin, Trinity College, Dublin, Ireland

Author Note:

Caoilte Ó Ciardha is now at School of Psychology, University of Kent, UK.

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

The results of two studies are reported examining the utility of a pictorial modified Stroop task in the assessment of sexual interest in a sample of non offending participants and of sexual offenders against children. A mixed factorial design was adopted for both. Nine gay and twelve straight participants took part in the first study which found that participants typically had attentional bias on the pictorial modified Stroop task that was in line with their stated sexual interests. Twenty four sexual offenders against children and twenty four control participants took part in the second study. Again results indicated that the task was tapping into the participants' stated sexual interests. Furthermore extrafamilial offenders and offenders with an admitted sexual interest in children demonstrated the greatest mean bias for child stimuli relative to adult stimuli. A cautious interpretation of the results was recommended, given the sample size in the study, the heterogeneity of the sample, differences in cognitive speed among offenders and controls and other methodological caveats.

Keywords

Pictorial Modified Stroop Task

Indirect Measures

Sexual Abuse

Sexual Interests

Applied Implicit Cognition

Introduction

To date the most widely used method of assessing the sexual interests of offenders has been penile plethysmography (PPG; Marshall & Fernandez, 2003). Partially as a response to criticisms of the PPG (see Kalmus & Beech, 2005) and also in an effort to get a broader understanding of the cognitive appraisal of sexual stimuli, researchers are increasingly looking at alternative methods of assessing sexual interest. These alternatives are mostly latency-based cognitive tasks (for a review of measures, see Thornton & Laws, 2009). The development of a sophisticated battery of standardized cognitive tasks could prove to be a useful adjunct or even an alternative to PPG measurement (Banse, Schmidt, & Clabour, 2010; Ó Ciardha & Gormley, 2009). Such a battery would not only offer a tool with practical application in assessment but also give researchers insight into the cognitive processes associated with sexual interest and sexual offending.

Currently there are several tasks that show promise in tapping into offence related cognition. The modified Stroop task (Ó Ciardha & Gormley, 2009; Price & Hanson, 2007; P. Smith & Waterman, 2004) involves presenting words or images to participants that may be sexually salient. Stimuli are presented in one of several colors—usually four— which participants must identify as quickly as possible. A consistent delay in responding to certain categories of stimuli is hypothesized to be a product of the salience of those stimuli to the individual. The pictorial modified Stroop task is examined in detail later. The choice reaction time task (CRT; Giotakos, 2005; Gress, 2008; Andreas Mokros, Dombert, Osterheider, Zappalà, & Santtila, 2010) is similar to the pictorial modified Stroop task but instead of identifying a color, the individual must identify the location of a dot superimposed on the image. Again, a

consistent delay in responding to a particular image type is taken to indicate a salience of those images.

The Implicit Association Test (IAT; Banse, et al., 2010; Brown, Gray, & Snowden, 2009; Gray, Brown, MacCulloch, Smith, & Snowden, 2005; Mihailides, Devilly, & Ward, 2004; Nunes, Firestone, & Baldwin, 2007; Ó Ciardha & Gormley, 2009; Steffens, Yundina, & Panning, 2008) involves categorizing stimuli using two buttons. Each button has two concepts allocated to it. The ease with which an individual categorizes items onto either button is seen as an indication of the relative strength with which the concepts allocated to that button are associated for the individual. An individual, for example, who categorizes words belonging to a sex/children button quicker than when sex and adult are paired may be hypothesized to have associations indicative of deviant interests or implicit theories involving sex and children.

Viewing time (Abel, et al., 2004; Glasgow, Osborne, & Croxen, 2003; Gress, 2005; Harris, Rice, Quinsey, & Chaplin, 1996) asks participants to rate an image on some dimension (typically how attractive they find the individual depicted) while recording the time it takes to respond. This task, therefore, includes an explicit self-report component (i.e. the image ratings) and a more implicit measurement of the time the individual spends viewing each image. Rapid serial visual presentation (RSVP; Beech, et al., 2008) examines the degree to which presentation of stimuli (e.g. an image of a child) results in an “attentional blink” whereby the participant fails to correctly respond to a task presented in rapid succession. Despite the promise shown by these tasks, several challenges face this field of research. There remains a lack of consensus on the best methodology to adopt for each paradigm. Additionally, there needs to be a greater focus on placing the results in a theoretical context (Imhoff, et

al., 2010). For example, it is not yet clear to what extent these tasks are measuring analogous processes or distinct cognitive phenomena relating to deviant sexual interest or sexual offending.

This article explores the utility of a pictorial modified Stroop task (P-MST) in the assessment of sexual interest; specifically whether images of adults and children can produce a systematic delay in responding that is related to sexual interest. Geer and Bellard (1996) refer to a delay in responding to sexual stimuli as Sexual Content Induced Delay (SCID). In their study Geer and Bellard used a lexical decision task where the sexual content of word stimuli was hypothesized to interfere with the task of identifying stimuli as words or non-words. Similarly delays in responding to stimuli in tasks such as the CRT, viewing time and pictorial and word versions of the modified Stroop task could be referred to as SCID. However this could imply that the phenomena measured by these tasks are analogous. These tasks may instead measure different cognitive processes associated with sexual interest.

In the P-MST, interference is possibly occurring at two points along Spiering, Everaerd and Laan's (2004) information processing model of sexual arousal. Spiering et al. (2004) hypothesize that sexual stimuli can be processed implicitly or unconsciously and that attentional mechanisms are triggered depending on contextual variables and physiological sensitivity. Once conscious appraisal of the stimuli begins, regulatory processes are engaged and the subjective experience of arousal can be experienced. The first point at which a pictorial modified Stroop interference could be occurring would be at a pre-attentive stimulus encoding stage and would be responsible for any so-called 'fast' component in the task. 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). A slow effect, on the other

hand operates between trials (McKenna & Sharma, 2004). It is likely that carry-over effects from the increased attention to one stimulus that continues onto a subsequent trial contribute to these slow-effects. It is also likely that slow effects are partly due to higher order rumination on the stimuli content and the activation of associated concepts related to that stimulus category. This rumination and activation 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, for example.

The results of two studies are presented here. In the first, the P-MST is used with a sample of non-offending participants to determine whether the task is tapping into non-deviant sexual interest. The second uses the task with a sample of offenders against children in order to assess the potential of the P-MST to measure pedophilic sexual interest.

Study 1

The goal of the first study was to examine the utility of the P-MST in tapping into the sexual interests of non-offending participants. The results of the task were hypothesized to relate to participants' sexual orientation. Specifically, 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. The design and results of study 1 were presented in detail in Ó Ciardha and Gormley (2009) and are summarized here to inform the reader of the methodological development that preceded study 2. Analyses which expand upon Ó Ciardha and Gormley (2009) are reproduced in full.

Before completing the P-MST, the participants carried out a traditional Stroop task. In addition to the standard congruency related hypotheses it was expected that

there would be no difference between gay and straight participants in their results on the traditional Stroop task and that age would significantly correlate with overall response times.

For the P-MST it was hypothesized 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 slower reaction times to male images relative to female images, while the opposite pattern would be true for straight participants. It was hypothesized that images of children would produce faster reaction times 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 control stimuli (images of large 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 participants' adult orientations.

Method

Design

A mixed factorial design was used with gay and straight participants completing the traditional Stroop task along with the P-MST. Two Implicit Association Tests (IATs) were also completed by participants after finishing the P-MST but are not included here.

Participants

Nine gay and twelve straight participants took part in the study (though one gay and one straight participant had non-exclusive sexual interests). All participants were college educated and had fluent English. 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).

Apparatus/Materials

Computerized tasks were presented on a Gateway Pentium III computer with a Gateway EV9108 19" cathode ray tube monitor. Participant responses were made via a Cedrus response pad (model RB-620) with four colored buttons (red, green blue and yellow). Tasks were run using the SuperLab4© stimulus presentation software.

Stimuli for the P-MST included images of male and female children and adults in bathing suits along with control images of large cats. All images of children and some of the adults were taken from the clothed version of the Not Real People (NRP) image set (Laws & Gress, 2004). The remaining images were created by the lead author using morphing software. The backgrounds of all images (NRP and novel morphs) were removed and replaced with a plain gray background (RGB code: 78,78,78) and the images were colored red (243, 57, 10), green (57, 243, 10), blue (10,57,243) and yellow (255, 255, 0) using the 'Colour Replacement Brush' in Adobe Photoshop Elements 3.1©. Images were 373 x 500 pixels in size and participants were seated approximately three feet away from the screen. Participants were asked to indicate on a questionnaire whether they had strong, some or no sexual interest in male and female adults.

Procedure

Responses were given via the response pad. First, participants completed a set of practice trials. Second, they completed a traditional Stroop task with congruent, control and incongruent font-color to word pairings. The three trial types were

presented randomly. Third, participants completed the pictorial modified 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). Block order was randomized across participants and trials were randomized within blocks. Each image was presented four times, once in each of the four colors. A fixation cross was presented prior to each stimulus for 500ms. A new fixation cross and trial was presented immediately following a response from the participant. No time limit was set for responses. Response time was recorded for every trial.

Analysis

Ó Ciardha and Gormley (2009) used several methods of data treatment in their analysis. However in a later reanalysis of the data (Ó Ciardha, 2010) it was discovered that using ipsative z -scores, instead of mean responses with outliers removed, seemed to negate some of the influence of age of participant on response times. Both viewing time and PPG measures of sexual interest sometimes use ipsative z -scores to standardize responses across individuals (Barbaree & Mewhort, 1994; Sachsenmaier & Gress, 2009), i.e. to account for individual differences in response times in viewing time tasks and individual differences in percentage erection in PPG. It was decided to adopt a similar approach to attempt to counteract the influence of individual differences in cognitive speed on results. Since there can be large individual variability in the impact of age on cognitive speed (G. A. Smith & Brewer, 1985), using an individual's own mean and standard deviation to attempt to control for variability in cognitive speed may prove more effective than including age as a covariate.

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 all images from the mean reaction time for each block and dividing by the overall standard deviation for reaction times to all 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 percentiles. Negative ipsative z -score values indicated trial type means that were quicker than the grand mean while positive scores were slower. After converting each of the experimental trial type means, age and values for trial types were no longer significantly correlated. This ipsative z -score approach, which differed from the results originally published by Ó Ciardha and Gormley (2009), was found to increase effect sizes for significant results (Ó Ciardha, 2010) while still demonstrating the same patterns of findings that had originally been found (Ó Ciardha & Gormley, 2009). This suggested that the ipsative z -score approach was removing some noise from the data.

Results

Results of the traditional Stroop task

Consistent with typical traditional Stroop findings (MacLeod, 1991) a one-way repeated measures ANOVA found significant differences among the reaction times to control, congruous and incongruous words; $F(1.45, 29.002) = 27.915, p < .001$, partial $\eta^2 = .583$. Degrees of freedom were adjusted as sphericity could not be assumed. Using the Bonferroni method, post-hoc test found the incongruent word reaction times (Mean = 845.83ms, $SD = 198.36$ ms) to be significantly slower than both the control words (Mean = 747.1ms, $SD = 164.62$ ms) and the congruent words (Mean = 720.1ms, $SD = 166.35$ ms). Congruent and control words did not differ

significantly. The inclusion of sexual orientation as a between-subjects factor indicated no interaction between sexual orientation and the interference effect of the traditional Stroop task. Age correlated positively and significantly with the average reaction times for each of the three Stroop 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. Because of this influence of age, ipsative z -scores were used in analyzing the P-MST results.

Results of the pictorial modified Stroop task

A 2x5 mixed factorial ANOVA was carried out on the results of the P-MST with sexual orientation as the between-subjects factor and trial type as the within-groups factor (i.e. adult female, adult male, child female, child male and cat). Figure 1 presents the results of this analysis and Table 1 shows the means and standard deviations of the reaction times and ipsative z -scores across each of the stimulus categories for both sexual orientations. There were significant main effects of sexual orientation and trial type. However, since there was a significant interaction, all findings were interpreted in light of this; $F(4, 76) = 6.145$, $p < .001$, partial $\eta^2 = .244$. In order to identify the source of the interaction, further analyses were carried out. First, separate 2x2 mixed factorial ANOVAs were conducted, looking at child and adult stimuli separately. In both of these ANOVAs, sexual orientation was the between groups variable while gender of stimulus was the within groups variable.

Approximate location of Figure 1

Approximate location of Table 1

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 than images of female children; $F(1,19) = 6.85, p = .032$, partial $\eta^2 = .265$. Looking at adult images, there was a significant disordinal interaction between gender of stimuli and sexual orientation indicating that participants took longer to respond to orientation-consistent stimuli $F(1,19) = 16.318, p = .002$, partial $\eta^2 = .432$. Both significant ANOVA results above included a Bonferroni correction for familywise error.

The difference between responses to adult female and adult male images among gay men was significant; $t(8) = -2.724, p = .026$, two-tailed, $d = 1.928$. The same comparison among straight men also yielded a significant t value; $t(11) = 2.947, p = .013$, two-tailed, $d = 1.776$. Given the inevitable loss of power produced by effectively splitting the sample in two, it was considered inappropriate to further limit the power by applying post tests which are designed to protect against type I error; e.g. t -tests with Bonferroni conversions. Statistically, this is not an ideal solution, but it is an attempt to balance out the likelihood of committing a type I or type II error, and any interpretation of the results should be cognizant of this.

The difference between orientation-consistent adult images and orientation-consistent child images were compared across all participants and the results found that responses to adult images were significantly longer; $t(20) = 4.857, p < .001$, two-tailed, $d = 2.174$. Reaction times to control big cat stimuli were consistent across gay and straight participants; $t(19) = .638, p = .531$, two-tailed, $d = .293$.

A Receiver Operating Characteristic (ROC) curve was plotted to assess how well a score based on the difference between ipsative z -scores of adult female stimuli versus adult male stimuli response times was capable of predicting self reported sexual orientation. This analysis shows an area under the ROC curve (AUC) of .917

which represents a predictive ability that differs significantly ($p = .001$; $SE = .068$) from .5.

The influence of order on the observed data

The P-MST used here incorporated a blocked design to maximize group differences when reacting to potentially sexually salient stimuli. However, the disadvantage of this technique is that it may introduce order effects. 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 $\eta^2 = .203$. Response times quickened over successive blocks. When pairwise comparisons were carried out, with a Bonferroni correction, the reaction times to the final block (Mean = 659.41ms, $SD = 94.08$ ms) were found to be significantly faster than first in the block (Mean = 777.2ms, $SD = 189.66$ ms); $t(19) = 3.785$, $p = .001$, two-tailed.

Discussion

Results of the traditional Stroop task indicated the presence of a traditional Stroop effect. Age and response time across all trial types were positively and significantly correlated. The treatment of the P-MST results used analyses based on ipsative z -scores to counter the effect of individual differences in cognitive speed (suggested by the correlation between age and response times).

A mixed factorial ANOVA indicated an interaction of orientation of participant and stimulus type. A much larger sample would be needed to properly tease out the source of this interaction but results did indicate that adult stimuli produced significantly longer reaction times than child stimuli. In addition, further testing indicated that participants had significantly slower reaction times to sexual orientation-consistent adult images than orientation-inconsistent adult images and also

orientation-consistent child images. These patterns were in line with the hypothesis, though the risk of familywise error must be taken into account when interpreting results. Following support for the main research question (i.e. whether the P-MST would tap into sexual interest) the data were explored to determine how well they discriminated gay participants from straight ones. Using ROC analysis the P-MST demonstrated an excellent (Tape, 1999) ability to discriminate between gay and straight participants.

The P-MST produced an unexpected main effect of gender for the child stimuli when an ANOVA was carried out looking at reaction times 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. It is also possible that for these participants, images of male children were salient to them for reasons unrelated to sexual interest.

Some caveats are necessary regarding the ipsative z -score method used here. When the method is used with individuals with very large or very small standard deviations, the size of an effect may be distorted. Caution must therefore be exercised in the 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 I error and influence power when individuals with high or low variability of arousal are in a sample. Sachsenmaier and Gress (2009) 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 minimize or exaggerate an effect. In PPG small standard deviations are due to a uniformity of

arousal across tasks. A uniformity of reaction time across the amount of stimulus presentations in the P-MST would be much less likely. With a reaction time task there are several ways to check for indicators that the data might be distorted by a z -score transformation. The first is to look at the mean reaction times of an individual and their standard deviations. Very fast mean responses with a small standard deviation would suggest the participant was responding as rapidly as possible with little regard for accuracy. A check of the error rate would easily establish whether this was the case. Slow mean responses with a large standard deviation could indicate that the participant was not fully attending or was prioritizing accuracy over speed to too great an extent. Again, checking for errors and also looking at the rate of outliers could indicate whether the data should be considered problematic. Viewing time tasks do not have these additional checks available since participants can typically take as much time as they want to respond and there is no “correct” answer if the individual is rating how attractive they find an image.

The appropriateness of an ipsative z -score method in analyzing the results of a P-MST is an issue that needs further attention but is of primary concern if the measure is to be used as a clinical tool. The ipsative z -score method is used here to attempt to reduce extraneous noise in order to examine the construct validity of the task. Reducing this noise is especially important given the relatively small sample sizes of the two studies reported here. Future studies looking at further validating the measure by testing fakeability etc. should explore the influence of unusual response patterns on the utility of the z -score method.

Given the presence of an order effect in the data it seems likely that the removal of that effect would have produced data with an even better ability to discriminate the sexual interests of participants. Two methods may reduce the impact

of order. The first, quite obviously, is to do away with the blocked design, and completely randomize all stimuli. This would ensure 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. 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. A pseudorandom presentation could 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. Study 2 adopted such a design.

Study 2

Having demonstrated in study 1 that the P-MST was able to tap into the sexual interests of non-offending participants, the objective of study 2 was to explore whether the measure could demonstrate differences between men who had committed sexual offences against children and men who hadn't. The P-MST used in the current study was identical to that used in study 1 except that images were presented in clusters of the same trial type instead of larger blocks. As discussed, this was done in order to reduce potential order effects of the larger block design.

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 a number of possible ways in which researchers could attempt to categorise whether offenders 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, Haynes, Stalgaitis, & Flanagan, 1986; Seto, Lalumiere, & Kuban, 1999). In addition, admitted sexual interest in children was also taken into account.

The study set out first of all to explore whether there would be group differences in response times on the P-MST between offenders and non-offenders. It was then planned to look at the response times for those offenders deemed more likely to demonstrate a sexual interest in children relative to adults. It was hypothesized 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 to non-offending controls and most incest offenders (i.e. those without an admitted sexual interest in children).

The current study also set out to replicate the findings of study 1 by investigating whether the pictorial Stroop task was able to tap into the adult sexual preferences of participants and whether it was reliably able to discriminate between participants based on sexual orientation. It was hypothesized that participants would have longer reaction times to orientation-consistent images in the task. It was also hypothesized that the task would discriminate between the orientation of participants with an accuracy significantly greater than chance. Receiver Operating Characteristic (ROC) analysis were to be used to explore the sensitivity and specificity of the task in discriminating between offenders and controls, groups of offenders and also participants of different sexual orientations. While it was hypothesized that ROCs

would significantly discriminate between gay and straight participants, given the results of study 1, all other comparisons were exploratory.

As with study 1 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, Sherman, & Spreen, 2006). While the Stroop task 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.

Method

Design

A mixed factorial design was used with offending and non-offending control participants completing all experimental tasks. As with study 1, two Implicit Association Tests (IATs) were also completed by participants but are not included in this article. In all cases the Stroop task was completed before the IATs.

Participants

Twenty four men who had committed sexual offences involving children and 24 non-offending control participants took part in the study. All offenders had offences involving at least one child under the age of 16. Ten offenders were recruited from a community treatment setting, while 14 were incarcerated offenders. 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. Offenders from the community treatment setting 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.

Incarcerated participants were all convicted of sexual assault, rape, indecent assault, gross indecency or a combination of these offences. One inmate was serving a life sentence; the remaining 13 had a mean sentence of 6.6 years ($SD = 2.78$ years). Offending 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 second level education (equivalent to high school education). Fourteen offenders had completed their second level education, while ten had not.

On completion of the Stroop tasks, one offending participant indicated that he had adopted a strategy to minimize his response times on the task (most likely squinting or not looking directly at the screen). His Stroop task results were therefore excluded.

Apparatus/Materials

Computer apparatus and stimuli were the same as for study 1 except that some participants were tested using a Gateway Solo 9300 Laptop. A questionnaire was also administered to each participant. There were three versions of the questionnaire. All versions contained questions asking the participants about age, color 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 received 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 but not for incarcerated participants.

Procedure

Participants first read the information letter and signed the consent form. They then carried out the computerized tasks. Procedure for these tasks were identical to study 1 except that stimuli for the P-MST were grouped in smaller clusters (instead of blocks) according to stimulus types, with 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 colors, yielding a total of 16 images per cluster. Novel images were used in each cluster and the clusters of child stimuli contained one image from each of four Tanner stages. As with the traditional Stroop task, participants had to identify, using a button box, the color with which each image had been tinted. Cluster order was randomized across participants, with the condition that one cluster of each trial type had to be presented before a second cluster of any type could be presented. Trials were randomized within cluster. Response time was recorded for every trial along with whether that response was correct or not.

Results

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 and non-offending participants to congruous, control and incongruous trials on the traditional Stroop task. These means and their standard deviations are presented in Table 2. As with study 1 a main effect of trial type indicated 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. However, there was a significant between groups effect indicating that the offending group performed slower overall, $F(1, 44) = 10.725, p = .002$, 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 analyzing the results of the P-MST, in order to minimize the impact of individual differences in cognitive speed.

Approximate location of Table 2

The pictorial modified Stroop task

Ipsative z -score were calculated using an identical method to study 1. The method yielded values that indicated how many standard deviations an individual's trial type mean was away from their overall mean. As with study 1, negative values indicated trial type means that were quicker than the overall mean while positive scores were slower.

A 2x2x5 mixed factorial ANOVA was carried out on the results of the P-MST, 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 in adults 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. Unfortunately this form of 'double check' was not available for the incarcerated

participants. 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. The ANOVA demonstrated a significant interaction of trial type and sexual orientation, $F(4, 168) = 3.163, p = .015$, partial $\eta^2 = .07$ as depicted in Figure 2. There were no interactions involving offender status indicating that offending and non-offending participants yielded similar patterns of responding. However, it was hypothesized that all offenders would not demonstrate a sexual interest in children.

Approximate location of Figure 2

To test whether subgroups of offenders differed in their patterns of responding, participants were categorized as *high deviance* if they had admitted a sexual interest in children or had extrafamilial victims (this included exhibitionism and child pornography offenders along with incest offenders with extrafamilial victims or and admitted sexual interest in children under 16 years of age). The remaining participants, categorized as *low deviance* comprised of offenders with intrafamilial only victims and no self reported deviant sexual interest. Given the size of the sample and to simplify the analysis, the response times of participants were re-organized to remove the need to add sexual orientation and gender of victim as variables in the analysis. PPG research has found that stated preferred gender of adult typically corresponds to arousal across different stimulus age categories (e.g. Blanchard, Klassen, Dickey, Kuban, & Blak, 2001). 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 3 graphically shows the pattern of responses across the four trial types for the three groups of participants: control, low deviance and high deviance.

Approximate location of Figure 3

A 3x4 mixed factorial ANOVA was carried out comparing the response times of the three groups (control, high deviance and low deviance) across the four new trial types. There was a main effect of trial type; $F(3, 129) = 7.668, p < .001, \text{partial } \eta^2 = .151$. There was no interaction of group and trial type; $F(6, 129) = .605, p = .726, \text{partial } \eta^2 = .151$. Given that this was an exploratory study with limited power, specific analyses of theoretical interest were conducted as opposed to a full complement of *a posteriori* pair-wise comparisons. First, paired samples *t*-tests were carried out to compare response times to adult and child orientation/offence-consistent stimuli. Control participants had significantly slower reaction times to adult stimuli (mean ipsative *z*-score = .17, $SD = .272$) compared with child stimuli (mean ipsative *z*-score = -.002, $SD = .242$), $t(23) = 2.076, p = .049, d = .873$, two-tailed. The high deviance group did not differ significantly in their response times to age-appropriate (mean ipsative *z*-score = .04, $SD = .299$) and age-inappropriate stimuli (mean ipsative *z*-score = .048, $SD = .253$), $t(13) = -.066, p = .948, d = .04$, two-tailed. Offenders in the low deviance group also had a result that indicated no significant difference between reaction times to adult (mean ipsative *z*-score = .163, $SD = .285$) and child stimuli (mean ipsative *z*-score = -.041, $SD = .208$), $t(7) = 1.268, p = .245, d = .952$, 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 3 that the response pattern of the low

deviance group and the control participants mirror each other quite closely. A small sample size in the low deviance 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 orientation-consistent versus inconsistent adult stimuli and whether this pattern was the same for child stimuli. These results were then compared with those of low deviance and high deviance individuals since it was hypothesized that the high deviance group would differ from controls while the low deviance group would not. For control participants, response times to orientation-consistent images of adults (mean ipsative *z*-score = .17, *SD* = .272) were significantly longer than those for orientation-inconsistent adult stimuli (mean ipsative *z*-score = -.14, *SD* = .204); $t(23) = 4.099$, $p < .001$, $d = 1.666$, two-tailed, while the difference between orientation-consistent (mean ipsative *z*-score = -.002, *SD* = .242) and inconsistent (mean ipsative *z*-score = -.101, *SD* = .209) child images was not significantly different; $t(23) = 1.505$, $p = .146$, $d = .629$, two-tailed. The low deviance group showed 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 (mean ipsative *z*-score = .163, *SD* = .285) and inconsistent (mean ipsative *z*-score = -.088, *SD* = .238) adult images was not significant; $t(7) = 1.595$, $p = .155$, $d = 1.186$, two-tailed, though this was likely due to size of the sample of low deviance individuals. The difference between offence consistent (mean ipsative *z*-score = -.041, *SD* = .208) and inconsistent child images (mean ipsative *z*-score = -.072, *SD* = .21) was also non-significant but, as mentioned, had a small effect size; $t(7) = .229$, $p = .826$, $d = .2$, two-tailed.

In contrast to the results of the low deviance group and the control participants, high deviance participants demonstrated larger differences between response times to offence consistent (mean ipsative z -score = .048, SD = .253) and offence inconsistent (mean ipsative z -score = -.188, SD = .161) child images than they did to orientation-consistent (mean ipsative z -score = .04, SD = .299) and inconsistent (mean ipsative z -score = -.13, SD = .223) adult images. The difference between child images were significant; $t(13) = 2.427, p = .03, d = 1.3.$, two-tailed, while they were not for adult images; $t(13) = 1.447, p = .172, d = .79$, two-tailed. As with study 1, multiple analyses were carried out which increased the risk of familywise error. Again, given the sample sizes involved in the subgroups measured, conversions were not applied in an attempt to balance out the likelihood of committing a type I or type II error. Results were therefore interpreted cautiously.

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, $AUC = .557, p = .509, SE = .087$. The measure fared only slightly better when discriminating between the high deviance group and non-offenders; $AUC = .592, p = .358, SE = .102$.

Study 1 explored the utility of an ROC curve in establishing whether the results of the P-MST could differentiate between participants based on sexual orientation. To demonstrate whether this was the case with the current data, an ROC was plotted to see how well the 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.

Discussion

Study 2 found that offenders overall did not differ from non-offenders in their patterns of responding on the P-MST. 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 only offenders or non-offenders. ROC analyses showed some utility in differentiating between gay and straight participants based on their P-MST results but not in differentiating between offending and non-offending participants.

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 P-MST. 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 P-MST 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. As would be expected, the reaction times of control participants 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 offenders deemed likely to show low deviance, though the sample lacked sufficient size to achieve significance in comparing the difference between the age-appropriate and age-inappropriate orientation/offence consistent

stimuli. Individuals deemed likely to demonstrate high deviance 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 many reasons. The first is that the task itself does not tap into sexual interests. This explanation seems unlikely, given the patterns found in this study and in study 1, albeit using a different version of the P-MST. 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. It could be that the presence of older children in the stimulus sets had a disproportionate influence on the increase in reaction times. Additionally, if a rumination effect is implicated in the P-MST for sexually salient stimuli, as is hypothesized by Ó Ciardha and Gormley (2010) 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 P-MST as a clinical tool and therefore highlights the need for further research and validation. Study 1 found an unexpected attentional bias towards male children over female children across all participants, regardless of orientation. This finding was not replicated by study 2. This discrepancy highlights the need of further testing and validation of the P-MST for sexually salient stimuli.

In line with PPG research there was a difference between incest offenders and non-incest offenders in their responses to the P-MST. 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 suggest that this 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 etc. These comparisons would have been possible with a larger sample. However the goal of the current study was to determine if the P-MST had initial promise in exploring deviant sexual interest in children. The finding that offenders with extrafamilial victims 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.

Though the results of the P-MST indicated that the task is measuring attentional processes related to sexual interest, the results were not as clear-cut as had

been expected. A clustered approach was taken instead of the blocked one used in study 1 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. Ó Ciardha and Gormley (2010) have found that, using a between subjects experimental design, blocked presentation of the P-MST for sexual interest significantly outperforms both clustered and random presentation with non-offending participants. Despite this, the concern regarding the influence of order still remains. As a minimum, larger clusters should be used to allow maximal rumination or the activation of concepts associated with sexual interest. Caveats regarding the use of ipsative z -scores should be taken also into account in any future studies. In addition future studies should consider dividing the child image blocks or clusters into younger and older children. This would allow an exploration of both pedophilic and hebophilic sexual interest. Furthermore, it may be difficult to distinguish clearly between the ages of children in the NRP set so using only the extreme age categories may produce clearer differentiation (Mokros, et al., in press). While this study's findings suggest that the P-MST is not as effective as other tasks such as the IAT (e.g. Brown, et al., 2009) or CRT (e.g. Mokros, et al., 2010) in discriminating between offenders and non-offenders, further refinement of the task and administration with larger more defined samples may improve the measure's utility and ability to predict group membership.

References

- 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.
- 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
- 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.
- 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.
- Blanchard, R., Klassen, P., Dickey, R., Kuban, M. E., & Blak, T. (2001). Sensitivity and specificity of the phallometric test for pedophilia in nonadmitting sex offenders. *Psychological Assessment*, 13(1), 118-126.
- 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.
- 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

- 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.
- 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.
- 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.
- 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.
- 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* (Doctoral thesis, University of Victoria, British Columbia, Canada).
- 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.

- Imhoff, R., Schmidt, A. F., Nordsiek, U., Luzar, C., Young, A. W., & Banse, R. (2010). Viewing time effects revisited: Prolonged response latencies for sexually attractive targets under restricted task conditions. *Archives of Sexual Behavior, 39*(6), 1275-1288. doi:10.1007/s10508-009-9595-2
- Kalmus, E., & Beech, A. R. (2005). Forensic assessment of sexual interest: A review. *Aggression and Violent Behavior, 10*(2), 193-217.
- 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.
- MacLeod, C. M. (1991). Half a Century of Research on the Stroop Effect : An Integrative Review. *Psychological Bulletin, 109*(2), 163-203.
- 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.
- 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: Journal of Research and Treatment, 16*(4), 333-350.
- Mokros, A., Butz, M., Dombert, B., Santtila, P., Bäuml, K.-H., & Osterheider, M. (in press). Judgement of age and attractiveness in a paired comparison task: Testing a picture set developed for diagnosing paedophilia. *Legal and Criminological Psychology*. doi:10.1348/135532510X514104

- 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
- 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.
- 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.
- 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. (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*. (pp. 177-201): Wiley-Blackwell.
- Ó Ciardha, C., & Gormley, M. (2010). *A comparison of blocked, clustered and random stimuli presentation in the pictorial modified Stroop task for sexually salient stimuli*. Manuscript submitted for publication.

- 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.
- Sachsenmaier, S. J., & Gress, C. L. Z. (2009). 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*.
- Seto, M. C. (2008). *Pedophilia and sexual offending against children: Theory, assessment, and intervention*. Washington, DC, US: American Psychological Association.
- Seto, M. C., Lalumiere, M. L., & Kuban, M. (1999). The sexual preferences of incest offenders. *Journal of Abnormal Psychology, 108*(2), 267-272.
- Smith, G. A., & Brewer, N. (1985). Age and individual differences in correct and error reaction times. *British Journal of Psychology, 76*(2), 199-203.
- 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.
- Spiering, M., Everaerd, W., & Laan, E. (2004). Conscious Processing of Sexual Information: Mechanisms of Appraisal. *Archives of Sexual Behavior, 33*(4), 369-380.
- Steffens, M. C., Yundina, E., & Panning, M. (2008). Automatic associations with "erotic" in child sexual offenders: Identifying those in danger of reoffence. *Sexual Offender Treatment, 3*, 1-9.
- 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>

Thornton, D., & Laws, D. R. (Eds.). (2009). *Cognitive Approaches to the Assessment of Sexual Interest in Sexual Offenders*. Chichester, UK: Wiley.

Table 1: Means and standard deviations for response times and ipsative z-scores across categories of pictorial stimuli for gay and straight participants.

Sexual Orientation	Stimulus Type	Mean Reaction Time ^a	SD of Mean Reaction Time	Mean Ipsative z-score ^b	SD of Mean Ipsative z-score
Gay	Female Child	672.953	81.464	-.232	.268
	Male Child	729.143	153.13	-.041	.292
	Female Adult	692.299	105.035	-.157	.241
	Male Adult	913.563	404.8	.496	.583
	Cat	697.592	124.699	-.15	.308
Straight	Female Child	700.147	92.858	-.188	.207
	Male Child	758.49	107.05	.106	.412
	Female Adult	825.443	164.971	.457	.402
	Male Adult	734.98	141.023	-.058	.293
	Cat	689.126	116.927	-.238	.315

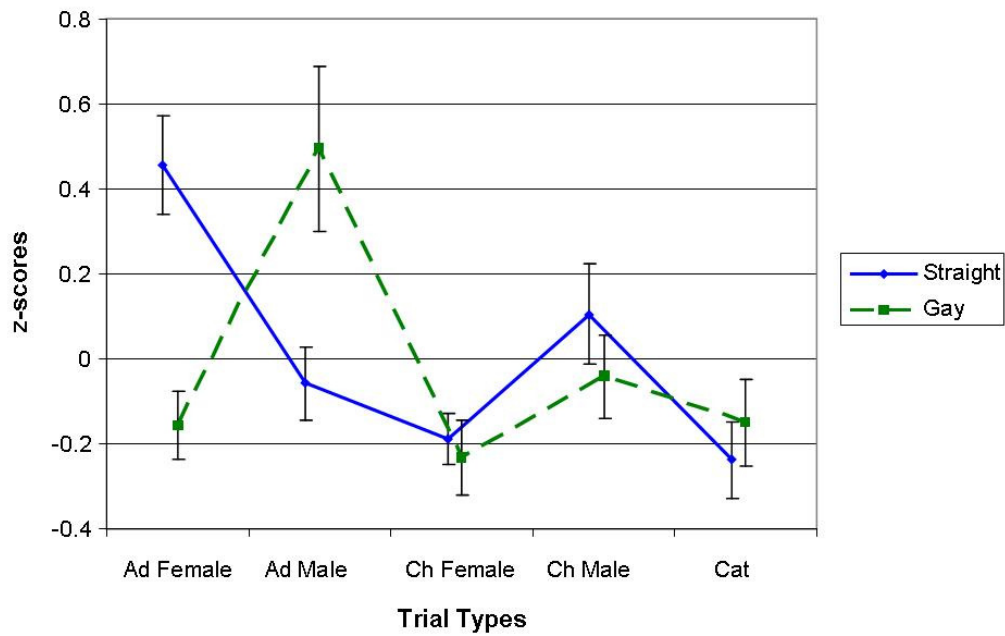
- a. Mean reaction times with outliers (in individual data) over 3 times the interquartile range beyond the 25th and 75th percentiles removed
- b. Minus values indicate that individuals' response times to the category are typically slower than their mean response time to all pictorial stimuli.

Table 2: Means and standard deviations for response times across categories of traditional Stroop task stimuli for offending and non-offending participants.

Participant	Stimulus Type	Mean Reaction	SD of Mean
-------------	---------------	---------------	------------

Group		Time ^a	Reaction Time
Control	Congruous	738.77	173.334
	Neutral	748.596	155.73
	Incongruous	892.552	239.821
Offender	Congruous	954.759	280.514
	Neutral	997.622	305.341
	Incongruous	1169.211	424.485

a. Mean reaction times with outliers (in individual data) over 3 times the interquartile range beyond the 25th and 75th percentiles removed



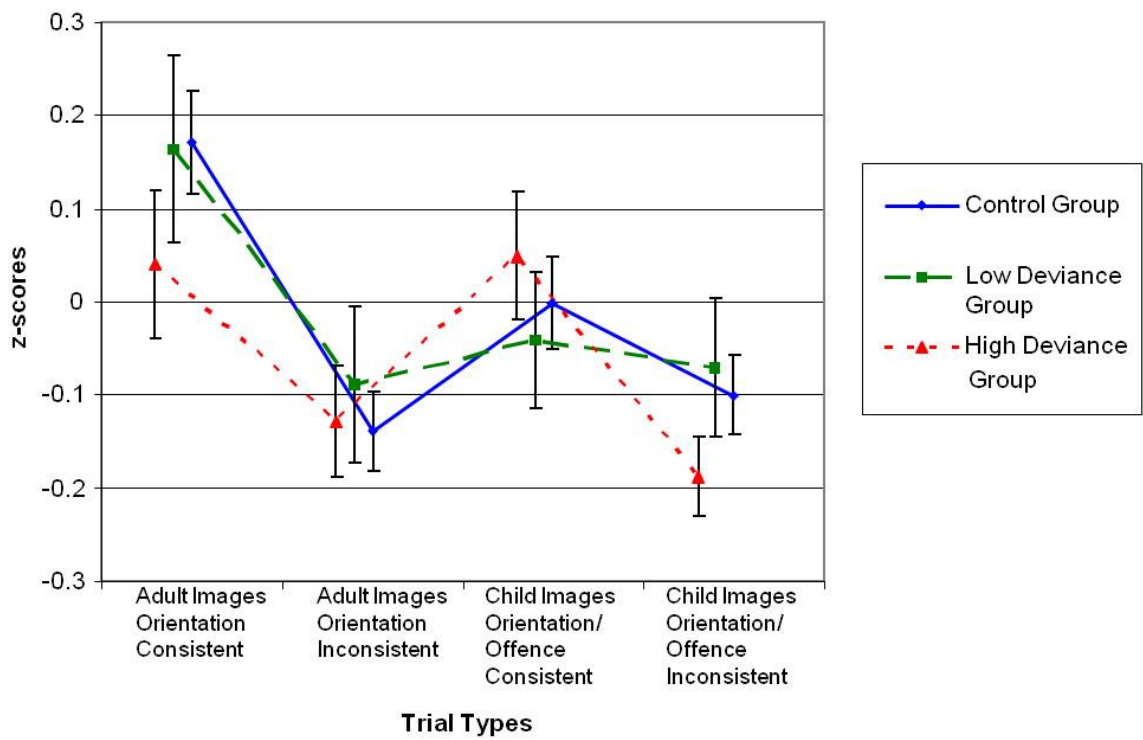
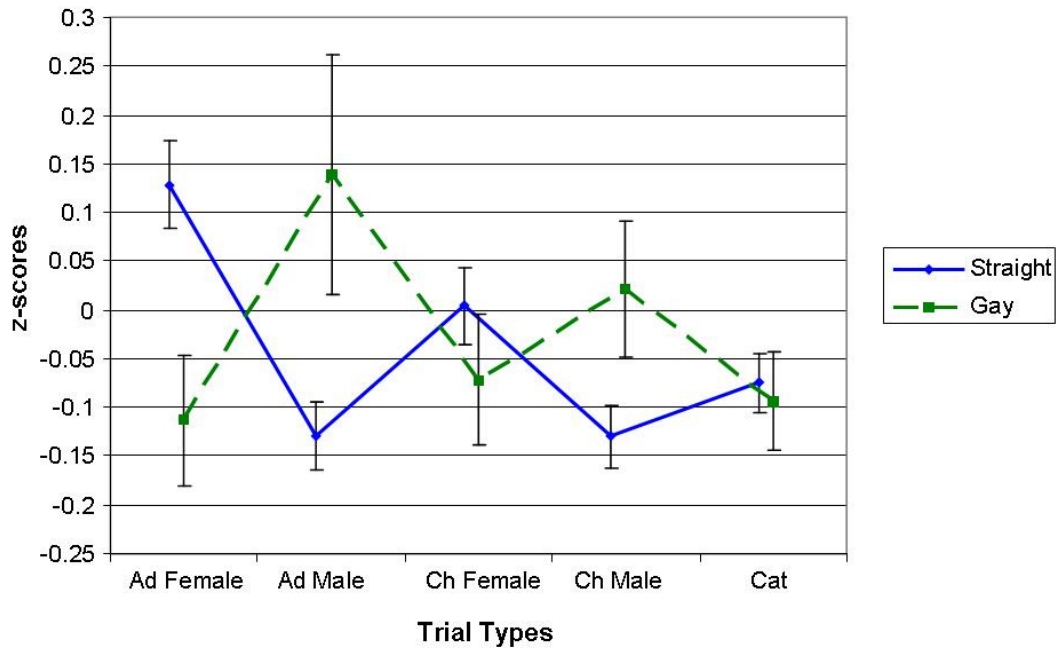


Figure 1: Reaction times (as ipsative z-scores) across trial types for gay and straight participants in Study 1 (error bars indicate one standard error).

Figure 2: Reaction times (as ipsative z-scores) across trial types for gay and straight participants in Study 2 (error bars indicate one standard error).

Figure 3: Response times to stimulus types for control participants, offenders deemed likely to show low sexual deviance and offenders deemed likely to show high sexual deviance (error bars indicate one standard error).