Legal and Criminological Psychology (2011), 16, 323–334

The British Psychological Society

www.wileyonlinelibrary.com

Judgment of age and attractiveness in a paired comparison task: Testing a picture set developed for diagnosing paedophilia

Andreas Mokros[1]*, Matthias Butz[1], Beate Dombert[1], Pekka Santtila[2], Karl-Heinz Bäuml[3] and Michael Osterheider[1]

[1]Department of Forensic Psychiatry and Psychotherapy, School of Medicine, University of Regensburg, Germany
[2]Department of Psychology, Åbo Akademi University, Turku, Finland
[3]Department of Psychology, University of Regensburg, Germany

Purpose. The Not-Real-People picture set serves as stimulus material for assessing paedophilic sexual interest. The pictures were generated according to five stages of sexual maturation from infancy to adulthood. While experts seem to be able to discern between the pictures based on maturity levels for a nude variant of the stimuli, it remained unclear whether lay persons would be able to reliably distinguish between picture of clothed individuals from various maturity levels.

Methods. Heterosexual university students (52 females, 50 males) participated in a paired comparison task, deciding which of the persons shown on two pictures was younger and which one was more attractive. Additionally, reaction times were recorded.

Results. Both male and female judges were able to differentiate between stimuli from the opposite sex with regard to age: the perception of the stages of pubertal development conformed to a measurement on a ratio scale. Similarly, the decisions on attractiveness of the categories were in accordance with a ratio scale. Male participants favoured adolescent and adult female stimuli. Contrary to expectation, female participants did not show a preference for any single age category.

Conclusions. The stimuli appear suitable for diagnostic purposes. The participants' decisions reflect the inherent maturity levels of the persons depicted. Reaction times in the paired comparison paradigm increase with task difficulty (i.e., similitude of pictures). The implications for indirect measures of sexual interest, based on reaction times, are discussed.

*Correspondence should be addressed to Dr Andreas Mokros, Department of Forensic Psychiatry and Psychotherapy, School of Medicine, University of Regensburg, District Hospital, Universitaetsstrasse 84, D-93053 Regensburg, Germany (e-mail: andreas.mokros@medbo.de).

DOI:10.1348/135532510X514104

All assessment methods for paedophilic sexual interest require suitable, child-related stimulus material. The use of picture stimuli for the assessment of paedophilic sexual interest is constrained by ethical and legal considerations (Abel, Huffman, Warberg, & Holland, 1998). The use of seized child pornographic pictures is clearly out of the question even though such material has been used in the past (Maletzky, 1995). Furthermore, there are various reasons why the stimulus material should be non-suggestive and non-pornographic. The main reason for this directive is to prevent any potentially detrimental effects on participants (Launay, 1999). Moreover, the legislation on child pornography has been tightened in most Western countries over the last couple of years, including the USA (Jones, 2006). As a consequence, researchers would commit a serious offence if they possessed such material and showed it to participants. Finally, the production of naturalistic photographs of children for the purpose of using them within diagnostic procedures concerning paedophilia is likely impossible to achieve.

As a consequence, Laws and Gress (2004) developed a picture set that was standardized for levels of sexual maturity, and contained only non-suggestive, modified pictures that did not show any natural persons. The resulting *Not-Real-People* (NRP; Pacific Psychological Assessment Corporation, 2004) picture set consists of a total of 160 images: 8 pictures for each of 5 categories of sexual maturity in either sex with a clothed and a nude variant. For example, eight pictures of nude infant girls, eight pictures of clothed male adults, and so forth. As a classification scheme for sexual maturity, Laws and Gress (2004) adopted the five stages of pubertal development as described by Tanner (1973). The changes in primary and secondary sexual characteristics reflect the level of sexual development. While these changes in bodily development are age-related they do not map onto an individual's age in a deterministic manner. Developmental and chronological age are not synonymous. Two teenagers of the same age, for example, can appear very different in terms of biological maturity (Greil & Kahl, 2005). The shift from Tanner stage 1 (infancy) to Tanner stage 2 marks the onset of puberty. Tanner stages 3 and 4 represent intermediate and late (adolescent) phases of pubertal development, while Tanner stage 5 refers to the sexually mature adult (Marshall, 1978).

Mokros, Dombert, Osterheider, Zappalà, and Santtila (2009) used the complete NRP set of clothed and nude depictions in a choice reaction time experiment with paedophilic and non-paedophilic participants. Mokros *et al.* found a significant interaction effect for the stimulus category of sexual maturity with participants' group status: paedophilic subjects took more time for a dot location task if the dot was superimposed on pictures of infants than if the dot was on pictures of adults. Conversely, non-paedophilic subjects showed the opposite pattern (i.e., slower reaction times towards adult than towards child stimuli). Moreover, reaction times were longer if the stimuli depicted nude individuals than clothed ones, an example of sexual content induced delay (Geer & Bellard, 1996).

As the data reported by Mokros *et al.* (2009) reveal, the distinction between paedophilic and non-paedophilic subjects was most pronounced when comparing pictures of small children (Tanner stage 1) against pictures of adults (Tanner stage 5). This opens up the question that the intermediate stimulus age categories (i.e., Tanner stages 2–4) may not be as easy to discern as the extreme ones (Tanner stages 1 and 5). Consequently, assessing the reliability of the stimulus material is a necessary precursor to maximizing task-validity. If a standardized set of picture stimuli, such as the NRP, is to be used in identifying paedophilic sexual interest in a meaningful way, it is paramount to establish whether testees can implicitly identify the supposed allocation of individual pictures to the inherent Tanner stages of pubertal development.

While the results from Laws and Gress (2004) on *nude* stimuli seem to support the notion of sufficient inter-rater agreement at least for expert judges, the question remains whether judges also perceive the pictures of *clothed* individuals as distinct and in accordance with the presumable Tanner categories. As a consequence, the present study was carried out to test the following hypotheses:

> *(1)*: Participants are able to distinguish among NRP pictures of clothed individuals from the opposite sex in terms of age. The decision of choosing the *younger* person from two pictures in a repetitive paired comparison task maps onto the putative membership of the pictures to the five Tanner stages of sexual development. As a consequence, the relative frequency of choices between pictures from different categories are measurable on a ratio scale, ordered from Tanner category 1 (infants) as the youngest to Tanner category 5.

Hypothesis 1 was limited to pictures of the opposite sex with regard to the gender of the participants in order to incorporate the notion of sexual orientation. The analyses were limited to heterosexual participants. Heterosexual participants are easier to sample in sizable numbers from the student population that subjects were recruited from.

> *(2)*: The same holds for decisions on attractiveness: if heterosexual participants are requested to choose which of two persons from the opposite sex was more attractive, the supposed Tanner categories map onto a ratio scale of preference.

> *(2a)*: For male participants, the pictures of females from Tanner stages 4 and 5 (adolescents and adults) will be significantly more attractive than females from Tanner stages 1 to 3 (infants to pubescents).

> *(2b)*: For female participants, only pictures of *adult* males (Tanner stage 5) will be significantly more attractive than males from Tanner stages 1 to 4 (infants to adolescents).

Hypotheses 2a and 2b follow from evolutionary psychology (Buss, 2007): males favour younger female mating partners, whereas females prefer more mature male mating partners. Using an eye tracking procedure, Suschinsky, Elias, and Krupp (2007) found that young males fixated more often and longer on pictures of females that they rated most attractive. More specifically, Suschinsky *et al.* noted that body regions that are reproductively particularly relevant (i.e., head and breasts) received most visual attention. Furthermore, images with lower waist-to-hip ratios were looked at more intently – a finding that Suschinsky *et al.* interpreted in terms of the waist-to-hip ratio as an indication of reproductive fitness. Silverthorne and Quinsey (2000) assessed the age preferences of hetero- and homosexual males and females, using 30 pictures of human faces that ranged between about 18 and 60 years of age. Mean age of the participants was 33 years. Both hetero- and homosexual male participants showed stronger preferences for younger faces (from 18 to their mid-twenties). Female participants (hetero- and homosexual), in contrast, favoured older faces (i.e., in their mid-thirties or above). Moreover, stimulus age and participants' age were unrelated (except for a weak correlation of $r = .29$ within the subgroup of homosexual males).

> *(3)*: The reaction times for the age comparison (Who is younger?) are linearly related to the difficulty of the task: the more disparate any two pictures are in terms

of their corresponding levels on the ratio scale for age judgments, the less time a participant will need to reach a decision.

Hypothesis 3 follows from the general finding in research on decision making that more difficult decisions require more time to be completed (e.g., Wright & Ayton, 1988). Palmer, Tepe Nasman, and Wilson (1994), for instance, found significantly slower reaction times for a match/mismatch procedure using same or different letters if the task difficulty was higher (i.e., categorical rather than physical similarity). If the same held for paired comparisons of pictures with regard to judgments of age, task difficulty could be a relevant confound for applications such as the dot probe paradigm (MacLeod, Mathews, & Tata, 1986). The dot probe task includes the simultaneous display of two pictures and has been used to assess sexual orientation (Prause, Janssen, & Hetrick, 2008). The dot probe task could possibly be adapted to assess paedophilic sexual interest. Then, the task difficulty of comparing pictures from various age ranges may become a relevant issue.

Method

Participants

Participants were 102 heterosexual university students (52 females, 50 males). The students took part in the study voluntarily for course credit or for two cinema vouchers, depending on personal choice. Mean age of the subjects was 24.5 years ($SD = 4.7$ years, range: 19–49). Male and female participants did not differ significantly with regard to age: $t_{(100)} = 1.76$, $p = .08$ (ns), $d = .35$. The female students ($M = 23.7$, $SD = 5.3$ years) were slightly younger than the male students ($M = 25.3$, $SD = 3.9$ years). Most likely, the small-to-moderate difference (effect size: $d = .35$) is not relevant given the observation of Silverthorne and Quinsey (2000) that attractiveness ratings of stimulus age categories were not correlated with participants' age.

Sexual orientation of the participants was assessed using a self-report questionnaire (Sell, 1996). Participants were included only if they rated themselves as 'exclusively heterosexual' on a seven-point Likert-type scale. Another 10 volunteers (5 females, 5 males) took part in the experiment but their data were excluded from subsequent analysis because the participants indicated a homosexual orientation in the questionnaire (Sell, 1996).

Instruments and procedure

The experiment was conducted using a desktop computer with a 24-in. (61 cm) thin-film-transistor display. The participants were placed in front of the display at a distance of 90 cm. Testing was done individually. On the screen, the pictures were about 20 cm high and 15.5 cm wide (750 × 575 pixels), with the two pictures of a given pair presented at a distance of 0.8 cm from each other. Thus, the visual angle was about 12.7° for the height of a picture and some 20° for the width of a pair of pictures presented simultaneously. The sequencing of the experimental trials was controlled through the software Presentation®, version 12.1 (Neurobehavioral Systems, Inc., 2008). The instruction made clear that the participants were about to see a sequence of pairs of pictures showing different individuals and that they would have to decide whom they found younger (phase 1) and whom they found more attractive (phase 2). The participants used a computer mouse to record their decisions through pressing the corresponding mouse button.

We used the 80 pictures from subsets A2 and B2 of the NRP picture set (Pacific Psychological Assessment Corporation, 2004; cf. Laws & Gress, 2004). That is, only the

clothed stimulus variant was used: in subsets A2 and B2, the persons are depicted in swimwear. That is, males are depicted in trunks, females are shown in one-piece swimsuits or wearing bikinis. Each stimulus trial consisted of the simultaneous presentation of two pictures from *different* Tanner categories (either female/female, male/male, or female/male). First, a fixation cross appeared in the centre of the screen for 0.5 sec, then the pair of picture stimuli was shown (until the participant responded through pressing the mouse button). The inter-stimulus interval was 3 sec.

In the first phase of the experiment, the participant was instructed to decide who was younger. After a short break, participants were asked to decide whom they found more attractive during the second phase of the experiment. Both experimental phases consisted of 40 trials (10 female/female, 10 male/male, and 20 female/male comparisons) and took about 20 min to complete. The sequence of trials was balanced and randomized within each phase with regard to pertinent combinations of stimulus Tanner category, sex, and position on the screen (left/right). Each of the 80 available pictures was shown only once within each experimental phase. In this way, effects of memory or habituation could be kept at a minimum. For each trial, the specifications were recorded (e.g., phase 1 [age]: male/Tanner 5 vs. female/Tanner 3). In addition, the participant's decision was registered as well as the reaction time (i.e., the interval between stimulus onset and pressing the mouse button).

The statistical analyses were limited to the decisions for concordant opposite sex comparisons of heterosexual participants: data from male participants looking at two female pictures from different Tanner categories and data from female participants looking at two male pictures from different Tanner categories. This yielded 10 comparisons per participant and task (age or attractiveness), and 500 comparisons in total per sex of participant and task.

The choice data were collapsed into asymmetric, complementary preference tables, separately for task (age/attractiveness), and participants' sex (males judging female stimuli/females judging male stimuli). The Bradley–Terry–Luce (BTL) model was employed to analyse these tables (Bradley & Terry, 1952; Luce, 1959; Zermelo, 1929). Preference parameters v_i were estimated through Ford's (1957) algorithm set to 100 iterations. Comparative scaling techniques (such as the paired comparison method) originated in psychophysics and allow direct ordering of objects with regard to a particular property (i.e., without using some external frame of reference such as a Likert scale). If the data meet particular model assumptions, the objects can be compared on an interval (Thurstone Case V) or even ratio scale (BTL model; Suppes & Zinnes, 1963).

If the BTL model holds, the relative frequencies of choosing any stimulus i over any stimulus j can be measured on a ratio scale, with parameter values v_i and v_j. The preference probabilities $P_{i,j}$ of choosing stimulus i instead of stimulus j can then be calculated for each pair of stimuli (i, j) by $P_{i,j} = v_i/(v_i + v_j)$ (cf. Bäuml, 1994). An example from our data: male participants found pictures from the category *female infants* (Tanner 1) younger-looking than images from the category *female adolescents* (Tanner 4) in nearly four-fifths of the pairings (78%). Comparison of the preference parameters (see Table 1, second row of data) yielded an estimated frequency of 78% that exactly matched the observed data: $P_{i,j} = v_i/(v_i + v_j) = 7.69/(7.69/2.17) = 7.69/9.86 \approx 0.78$.

Goodness of fit was assessed globally (i.e., accordance of model predictions with observed data) and specifically (i.e., equality of preference parameters) through likelihood-ratio tests (G^2) that are asymptotically distributed as chi-square (Bradley, 1984). The

Table 1. Parameter estimates from the BTL model for judgments of stimulus age (Who is younger?) and stimulus attractiveness (Who is more attractive?), with 95% confidence intervals (± CI)

			Tanner stage				
Criterion	Participants sex	Stimulus sex	1	2	3	4	5
Age	Female	Male	4.10 (± 0.84)	3.75 (± 0.79)	2.56 (± 0.59)	1.81 (± 0.45)	1.00 (± 0.29)
	Male	Female	7.69 (± 1.45)	4.13 (± 1.01)	2.64 (± 0.72)	2.17 (± 0.61)	1.00 (± 0.34)
Attractiveness	Female	Male	1.20 (± 0.26)	1.00 (± 0.23)	1.26 (± 0.27)	1.41 (± 0.30)	1.19 (± 0.26)
	Male	Female	1.05 (± 0.25)	1.12 (± 0.26)	1.00 (± 0.24)	1.62 (± 0.35)	1.79 (± 0.38)

reaction time data were analysed with mixed regression models (random intercepts, random slopes, and the combination of random intercepts and slopes).

Results

Table 1 summarizes the parameter estimates from the BTL model for the five stimulus categories (Tanner stages 1–5) for male and female judges with regard to age and attractiveness. Parameter values were rescaled through proportional transformations in each case so that the lowest parameter value was set to 1.

When judging the stimuli of the opposite sex in terms of *age*, the responses from both male and female participants were commensurate with the BTL model. The data did not differ from the predictions of the model; hence, the likelihood-ratio goodness-of-fit tests failed to reach statistical significance: $G^2_{(6)} = 10.33$ ($p = .11$) for female participants and $G^2_{(6)} = 8.08$ ($p = .23$) for male participants.

In addition, the order of the Tanner categories was reflected in the participants' decisions: stimuli from Tanner category 1 were regarded as youngest, while stimuli from Tanner category 5 were assumed to be oldest. Consequently, the sequence of parameter values in Table 1 follows the predicted pattern, with Tanner 1 stimuli obtaining the highest score in terms of youngness, whereas the lowest are afforded to Tanner 5 stimuli. More specifically, the parameter values for the stimulus categories (Tanner stages 1–5) were different from each other at $p < .001$: $G^2_{(4)} = 74.23$ ($p < .001$) for female participants and $G^2_{(4)} = 112.67$ ($p < .001$) for male participants.

Thus, the Tanner categories of the stimuli were regarded as distinct. This effect was more pronounced for male than for female participants, however. The geometric mean for the differences between stimulus Tanner categories was 1.67 for male participants and 1.42 for female participants, respectively. For the male participants judging female stimuli, the discreteness of categories is due to Tanner categories 1 and 5: their 95% confidence intervals do not overlap with any of the remaining categories, respectively. Concerning female participants' judgments of male stimuli, only Tanner category 5 is perceived as significantly distinct while the 95% confidence intervals of the remaining categories do overlap.

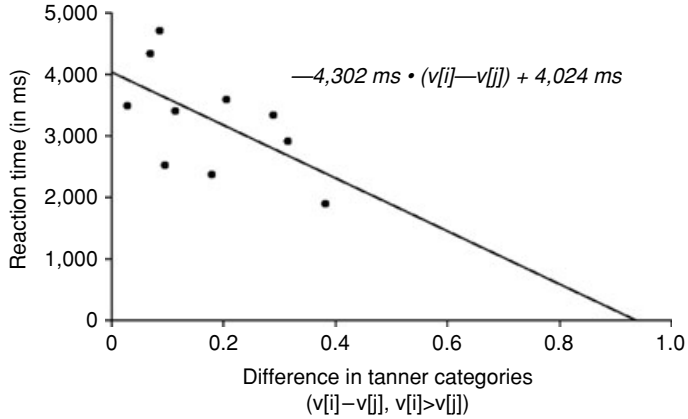

Figure 1. Mean reaction times of male participants for different increments of female stimulus categories on the v-scale with fixed regression slope and intercept.

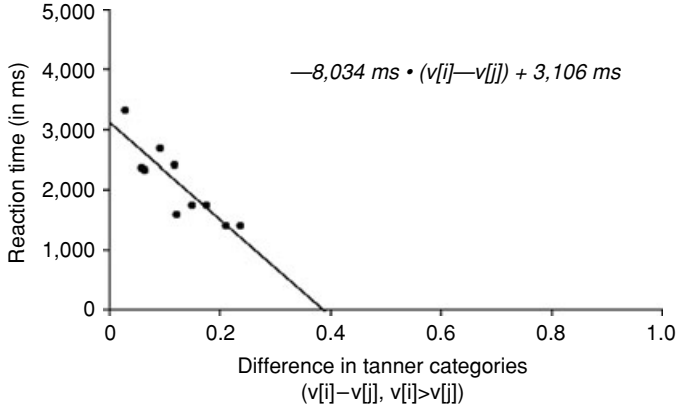

Figure 2. Mean reaction times of female participants for different increments of male stimulus categories on the v-scale with fixed regression slope and intercept.

The appraisal of the stimuli with regard to *attractiveness* was also in accordance with the specifications of the BTL model: $G^2_{(6)} = 7.45$ ($p = .28$) for female participants and $G^2_{(6)} = 2.70$ ($p = .85$) for male participants. As predicted, male participants tended to judge both adolescent and adult female participants (Tanner categories 4 and 5) as significantly more attractive than female infants or pubescents from Tanner categories 1 to 3: $G^2_{(1)} = 16.48$ ($p < .001$) for male participants. Contrary to expectation, female participants did not display a preference for any stimulus category of male pictures: $G^2_{(4)} = 3.29$ ($p = .42$).

As separate repeated measures regression analyses revealed, the reaction times for decisions on stimulus age (Who is younger?) receded in a linear fashion with task simplicity (see Figures 1 and 2): the further apart any two stimuli were on the ratio scale of their respective Tanner categories, the quicker both male and female participants would reach a decision. Following from the coefficient estimate from the regression analysis, the average reduction in reaction times for male participants from the most difficult (Tanner

Table 2. Repeated measures regression (random intercept model) of differences between age estimates on reaction times (in ms)

	Males ($N = 50$)			Females ($N = 52$)		
Effect	est.[a]	*SE*	*p*	*est.*	*SE*	*p*
Fixed	Coefficient			Coefficient		
Intercept	4,024	323	$<.001$	3,106	181	$<.001$
ΔScale[b]	−4,302	1,038	$<.001$	−8,034	897	$<.001$
Random	Variance			Variance		
Intercept	6.83×10^6	4.56×10^5	$<.001$	1.72×10^6	1.13×10^5	$<.001$
Subject	2.89×10^6	7.24×10^5	$<.001$	9.09×10^5	2.14×10^5	$<.001$

[a] est., regression estimate (coefficient/variance).

[b] ΔScale, difference between BTL-parameter estimates v_i and v_j for the age judgments on Tanner categories i and j ($i < j$, $i = 1 \ldots 4$, $j = 2 \ldots 5$), re-scaled to $\sum_{i=1}^{5} v_i = 1$.

category 3 vs. Tanner category 4) to the easiest comparison (Tanner category 1 vs. Tanner category 5) was 1,518 ms. For female participants, the corresponding value was estimated at 1,427 ms (see Table 2). Compared with other random coefficient models, the random intercept model showed best model fit for both male and female participants (Akaike's information criterion [AIC] = 9,343.62 for male participants and AIC = 8,994.42 for female participants, respectively).

Discussion

The decisions of both groups of participants (males and females) can be described by the BTL model for preference data (Bradley & Terry, 1952; Luce, 1959) with regard to judgments on both age and attractiveness. First, the participants ordered the stimulus categories on a single continuum of age appearance. The locations of the Tanner categories on this continuum were distinct. Furthermore, male judges perceived stronger differences between female stimulus categories than female judges between male picture categories. Most decisively, however, the order of the stimulus categories on the continuum matched the order of the Tanner categories. That is, persons on pictures from Tanner category 1 were perceived as the youngest, whereas individuals on images from Tanner category 5 were perceived as the oldest. Non-expert participants are able to distinguish between the inherent stages of sexual maturity, at least for the extreme categories from both ends of the spectrum (Tanner 1, prepubertal, vs. Tanner 5, adult). Pictures purportedly showing individuals from Tanner categories 2 to 4 (early puberty to adolescence) were not discernible from each other to a significant degree. Hence, the usefulness of these intermediate stimulus categories (Tanner 2–4) for validation research towards assessing paedophilic sexual interest remains doubtful. One might consider collapsing these images into one category in further research. This would simplify the range of levels of sexual maturity to three stages: infancy, puberty, and adulthood.

Participants' decisions concerning *attractiveness* were also in accordance with the predictions from the BTL model. There was no distinctive pattern of preference, however, among the female participants for any given age category. Male participants, in contrast, showed the expected preference for adolescent and adult women. The latter outcome of ascribing attractiveness not only to adults but also to adolescents is in line with differential parental investment as a cause for sex-specific mating preferences (Trivers, 1972). According to this theory, men are theorized as favouring younger female mating partners in order to maximize reproductive success – a prediction that has repeatedly been demonstrated (e.g., Bailey, Gaulin, Agyei, & Gladue, 1994; Buss, 1989; Kenrick & Keefe, 1992; Silverthorne & Quinsey, 2000). The observation that female participants did not distinguish between male stimuli of varying ages with respect to attractiveness may be explicable with research that indicated a rather non-specific sexual arousal response in women (Chivers & Bailey, 2005; Lykins, Meana, & Strauss, 2008) and with less consensus among women's attractiveness ratings of male photographs (Wood & Brumbaugh, 2009). More cogently, evolutionary psychology research indicates that male attractiveness is less relevant as a cue in mate selection for females (e.g., Buss & Barnes, 1986), a finding that can also be explained through the parental investment theory of sexual selection (Trivers, 1972): from the perspective of the female, the willingness and ability of the male partner to provide for the offspring is more relevant to ensure survival (cf. Jankowiak, Hill, & Donovan, 1992).

In addition, the results corroborate the notion that task difficulty is linearly related to reaction times. In this way, the present study complements the extensive literature on decision making and task difficulty (e.g., Wright & Ayton, 1988) with findings from a paired comparison paradigm. But the results also indicate that task difficulty should be closely scrutinized in all latency-based measures of deviant sexual interest. Otherwise, artifacts might occur insofar as incremental changes in reaction time may not only be due to disorders of sexual preference, but also due to implicit differences in task difficulty.

Finally, while the clothed pictures from the NRP set appears to be a viable stimulus material for research on and possibly for diagnostics of paedophilia, future developments should explore the potential of completely computer-generated, virtual images (Renaud, Rouleau, Granger, Barsetti, & Bouchard, 2002) to avoid the ethical dilemma of presenting naturalistic pictures in the assessment of sexual preference disorders. Further research on stimulus sets for assessing paedophilic sexual interest (such as the NRP) should strive for calibrating the stimulus set by asking paedophilic individuals to participate in the study. It is conceivable that paedophilic subjects are more sensitive to subtle peculiarities of the child images than experts (Laws & Gress, 2004) or the student laypeople in the present study. Furthermore, it would be sensible to include both hetero- and homosexual participants into future studies on the reliability of picture stimuli.

Acknowledgements

This research was supported by an Academy of Finland Grant (121232) to the first and fourth author and a Center of Excellence Grant from the Stiftelsen för Åbo Akademi Foundation to the fourth author. Parts of this research were presented at the 51st Conference of Experimental Psychologists (TeaP), Jena, Germany, March 31–April 1, 2009. The authors want to thank Teemu Laine from the University of Turku, Finland, for his help with programming the stimulus presentation script. None of the authors have any financial interest in or financial conflict with the subject matter or the materials discussed in this manuscript.

References

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

Bailey, J. M., Gaulin, S., Agyei, Y., & Gladue, B. A. (1994). Effects of gender and sexual orientation on evolutionarily relevant aspects of human mating psychology. *Journal of Personality and Social Psychology*, *66*, 1081–1093. doi:10.1037/0022-3514.66.6.1081

Bäuml, K.-H. (1994). Upright versus upside-down faces: How interface attractiveness varies with orientation. *Perception and Psychophysics*, *56*, 163–172.

Bradley, R. A. (1984). Paired comparisons: Some basic procedures and examples. In P. R. Krishnaiah & P. K. Sen (Eds.), *Handbook of statistics: Nonparametric methods*, (Vol. 4, pp. 299–326). Amsterdam: Elsevier.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, *39*, 324–345.

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences*, *12*, 1–49. doi:10.1017/S0140525X00023992

Buss, D. M. (2007). *Evolutionary psychology: The new science of the mind* (3rd ed.). Boston, MA: Allyn & Bacon.

Buss, D. M., & Barnes, M. F. (1986). Preferences in human mate selection. *Journal of Personality and Social Psychology*, *50*, 559–570. doi:10.1037/0022-3514.50.3.559

Chivers, M. L., & Bailey, J. M. (2005). A sex difference in features that elicit a genital response. *Biological Psychology*, *70*, 115–120. doi:10.1016/j.biopsycho.2004.12.002

Ford, L. R. Jr. (1957). Solution of a ranking problem from binary comparisons. *American Mathematical Monthly*, *64*, 28–33. doi:10.2307/2308513

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

Greil, H., & Kahl, H. (2005). Assessment of developmental age: Cross-sectional analysis of secondary sexual characteristics. *Anthropologischer Anzeiger*, *63*, 63–75.

Jankowiak, W. R., Hill, E. M., & Donovan, J. M. (1992). The effect of sex and sexual orientation on attractiveness judgments: An evolutionary interpretation. *Ethology and Sociobiology*, *13*, 73–85. doi:10.1016/0162-3095(92)90019-Z

Jones, C. (2006, April 10). *Porn law goes too far, 11th circuit rules*. Retrieved from http://www.law.com/jsp/article.jsp?id=1144414533557

Kenrick, D. T., & Keefe, R. C. (1992). Age preferences in mates reflect sex differences in human reproductive strategies. *Behavioral and Brain Sciences*, *15*, 75–133.

Launay, G. (1999). The phallometric assessment of sex offenders: An update. *Criminal Behaviour and Mental Health*, *9*, 254–274. doi:10.1002/cbm.317

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. doi:10.1348/1355325041719338

Luce, R. D. (1959). *Individual choice behaviour: A theoretical analysis*. New York: Wiley.

Lykins, A. D., Meana, M., & Strauss, G. P. (2008). Sex differences in visual attention to erotic and non-erotic stimuli. *Archives of Sexual Behavior*, *37*, 219–228. doi:10.1007/s10508-007-9208-x

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology*, *95*, 15–20. doi:10.1037/0021-843X.95.1.15

Maletzky, B. M. (1995). Editorial: Stimulus materials and the protection of victims. *Sexual Abuse: A Journal of Research and Treatment*, *7*, 109–111.

Marshall, W. A. (1978). Puberty. In F. Falkner & J. M. Tanner (Eds.), *Human growth: Postnatal growth*, (Vol. 2, pp. 152–155). London: Balliere Tindall.

Mokros, A., Dombert, B., Osterheider, M., Zappalà, A., & Santtila, P. (2009). Assessment of pedophilic sexual interest with an attentional choice reaction time task. *Archives of Sexual Behavior*. Advance online publication. doi:10.1007/s10508-009-9530-6

Neurobehavioral Systems, Inc. (2008). Presentation (Version 12.1) [Computer software]. Retrieved from http://www.neurobs.com/

Pacific Psychological Assessment Corporation (2004). *The NRP (Not Real People) stimulus set for assessment of sexual interest*. Victoria: Author.

Palmer, B., Tepe Nasman, V., & Wilson, G. F. (1994). Task decision difficulty: Effects on ERPs in a same–different letter classification task. *Biological Psychology*, *38*, 199–214. doi:10.1016/0301-0511(94)90039-6

Prause, N., Janssen, E., & Hetrick, W. P. (2008). Attention and emotional responses to sexual stimuli and their relationship to sexual desire. *Archives of Sexual Behavior*, *37*, 934–949. doi:10.1007/s10508-007-9236-6

Renaud, P., Rouleau, J. L., Granger, L., Barsetti, I., & Bouchard, S. (2002). Measuring sexual preferences in virtual reality: A pilot study. *Cyberpsychology and Behavior*, *5*, 1–9. doi:10.1089/109493102753685836

Sell, R. L. (1996). The Sell assessment of sexual orientation: Background and scoring. *Journal of Lesbian, Gay and Bisexual Identity.*, *1*, 295–310.

Silverthorne, Z. A., & Quinsey, V. L. (2000). Sexual partner age preferences of homosexual and heterosexual men and women. *Archives of Sexual Behavior*, *29*, 67–76. doi:10.1023/A:1001886521449

Suppes, P., & Zinnes, J. L. (1963). Basic measurement theory. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology*, (Vol. 1, pp. 1–76). New York: Wiley.

Suschinsky, K. D., Elias, L. J., & Krupp, D. B. (2007). Looking for Ms. Right: Allocating attention to facilitate mate choice decisions. *Evolutionary Psychology*, *5*, 428–441.

Tanner, J. M. (1973). Growing up. *Scientific American*, *229*(3), 34–43. doi:10.1038/scientificamerican0973-34

Trivers, R. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of man* (pp. 136–179). Chicago, IL: Aldine-Atherton.

Wood, D., & Brumbaugh, C. C. (2009). Using revealed mate preferences to evaluate market force and differential preference explanations for mate selection. *Journal of Personality and Social Psychology*, *96*, 1226–1244. doi:10.1037/a0015300

Wright, G., & Ayton, P. (1988). Decision time, subjective probability and task difficulty. *Memory and Cognition*, *16*, 176–185.

Zermelo, E. (1929). Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung [The calculation of the results from a tournament as a maximum problem within the calculus of probability]. *Mathematische Zeitschrift*, *29*, 436–460. doi:10.1007/BF01180541

Received 22 December 2009; revised version received 7 May 2010