Indirect Measures in Forensic Contexts

Alexander F. Schmidt, Rainer Banse, & Roland Imhoff

University of Bonn

Author Note

Alexander F. Schmidt, Department of Social and Legal Psychology, University of Bonn, Germany; Rainer Banse, Department of Social and Legal Psychology, University of Bonn, Germany; Roland Imhoff, Sozialpsychologie: Social Cognition, University of Cologne, Germany.

Correspondence concerning this article should be adressed to Alexander F. Schmidt, Institute for Psychology, Department of Social and Legal Psychology, University of Bonn, Kaiser-Karl-Ring 9, 53111 Bonn, Germany. Email: afschmidt@uni-bonn.de.

Indirect Measures in Forensic Contexts

Whenever psychologists try to diagnose a condition or disposition, psychometrically precise assessment procedures are of paramount importance. This general need is amplified in applied forensic contexts where diagnostic decisions may collide with individual and societal rights and needs. Assessment outcomes have as far-reaching consequences for the respondent (e.g., restraint of individual freedom) as for members of society as a whole (e.g., risk of future victimization). These conflicting interests underscore the need for valid measures. However, classic self-report assessment procedures such as questionnaires and interviews are inherently transparent and can easily be faked by respondents who are aware of the personal consequences of the assessment outcome.

Another problem of (forensic) assessment based on self-reports is the high demand imposed on respondents' introspective abilities. Some forensically relevant constructs may lack introspective accessibility per se, such as situation-specific impulses and implicit offence-facilitating theories (cognitive distortions). Other constructs may in principle be open to introspection but the quality of self-report depends on certain cognitive skills that are not common among prototypical forensic populations, who usually have relatively low education levels and weak verbal skills. Due to these crucial validity threats, researchers and practicioners alike question the usefulness of self-report techniques (but see Grieger, Hosser, & Schmidt, 2012; Walters, 2006). Therefore, specifically in forensic contexts, there is a strong need for reliable and valid measurement paradigms. Ideally, these should not rely on explicit self-report and introspection and should be less transparent as well as less deliberately controllable.

What Are Direct/Indirect vs. Explicit/Implicit Measures?

The search for a solution to the drawbacks of self-report approaches has led to an increasing research interest in "implicit" and/or "indirect" measures. This measurement approach has shown remarkable predictive validity across different psychological subdisciplines (e.g., Friese, Hofmann, & Schmitt, 2008; Perugini, Richetin, & Zogmaister, 2010), including the forensic domain (Snowden &

Gray, 2010). The success of indirect measures has been attributed to the fact that these approaches benefit from being (a) inherently less transparent than self-report measures (due to the indirect character of the measurement procedure) and (b) able to tap into automatic attitudes and behavioral dispositions (because of the implicitness of the constructs to be measured). However, despite the immense popularity of these measures, the precise terminological differences between the theoretical attributes "indirect" and "implicit" do not always seem to be unequivocal as these terms are often used interchangeably.

For the remainder of the chapter we will rely on the terminological distinction proposed by De Houwer (2006). Accordingly, it has to be distinguished between two different uses of the term measure that either refers to the measurement outcome or the measurement process. The term *implicit* is reserved for various functional properties of the measurement outcome. These properties describe typical criteria of automaticity that specify the particular sense in which a measure is implicit (e.g., respondents' lack of awareness of the relationship between the assessed construct and the measurement outcome, lack of conscious access to the relevant construct, lack of voluntarily control over the assessment outcome). These properties do not necessarily co-occur and have to be demonstrated empirically rather than being mere presumptions. The term *indirect* refers to the procedural properties of the measurement that are always based on an explicit set of rules of how the measurement score is derived from the assessment (otherwise, it would not qualify as a measure in a scientific sense). Whereas *direct* measures rely on introspective self-descriptions or ratings of indicators (e.g., questionnaire items) of the relevant construct, *indirect* assessment procedures use the behavior exhibited in response to a stimulus (e.g., response latencies while categorizing images) to draw indirect inferences on the construct in question. Notably, the measurement outcome of an indirect measure is not necessarily implicit (De Houwer, 2006), as for example individuals might be fully aware of the purpose of the assessment (which is true for most indirect measures utilized in forensic psychology).

In line with De Houwer's (2006) abovementioned terminological distinction, in the remainder of this chapter we will focus on forensically relevant measures that draw their diagnostic inferences indirectly from task behavior (i.e., response latencies[1]). A wide range of forensically relevant, latency-based indirect measures have been introduced to tap into various domains of individual differences (Snowden & Gray, 2010). By far the most research utilizing indirect measures, however, has focused on deviant sexual interests (DSI) in children[2] (for recent overviews, see Snowden, Craig, & Gray, 2011; Thornton & Laws, 2009a). Indirect latency-based measures of DSI are often referred to as "attention-based measures" (Gress & Laws, 2009; Kalmus & Beech, 2005; Ò Ciardha, 2011), postulating that the underlying processes rely on the differential allocation of attentional resources. However, for most of these measures this remains hypothetical, as only a small body of empirical research into their procedural underpinnings exists. Also, empirical demonstrations of relevant implicitness criteria are missing. As a consequence, in focusing on the nature of the dependent variable we prefer to use *latency-based indirect measures* as the theoretically most parsimonious umbrella term.

Latency-Based Indirect Measures of Deviant Sexual Interests

An overview of latency-based indirect measures of DSI utilizing samples involving male sexual offender s against children is reported in Table 1. All of these measures capitalize on individual differences in (sexual) information processing and, as a result, also get framed as cognitive approaches to the assessment of DSI (Thornton & Laws, 2009a).

(Insert Table 1)

An important distinction is whether indirect measures assess deviant sexual *interest* (DSI) or deviant sexual *preferences* (DSP). *Interest* refers to the absolute level of sexual interest in a specific

---

[1] Due to space restrictions we will exclude indirect assessment paradigms that capitalize on physiological reactions such as polygraphy in the field of deception detection (for a critical overview see National Research Council, Committee to Review the Scientific Evidence on the Polygraph, 2003; but see Verschuere, Ben-Shakar, & Meijer, 2011) or penile plethysmography/phallometry (for an overview see Kalmus & Beech, 2005).
[2] Throughout this chapter the term *deviant* refers to sexual interest in prepubescent children (irrespective of other paraphilic or otherwise abnormal sexual interests) as indicated by corresponding fantasies or behavior.

target category (e.g., prepubescent children) irrespective of interest in other categories (e.g., postpubescent individuals), whereas *preference* denotes relative sexual preference for one target category over another target category and is usually based on a difference index of a target category minus a comparison category (e.g., prepubescent over postpubescent individuals). Notably, although representing two different conceptualizations, interest and preference are often used interchangeably in the literature. Several indirect measures – the most prominent example being the Implicit Association Test (IAT; Perugini, Costantini, Richetin, & Zogmaister, this volume) – are inherently conceptualized as DSP measures because they are calculated from latency differences based on sexually relevant versus sexually irrelevant trials. DSP measures do not convey diagnostic information about the absolute level of DSI, because the baseline level of DSI gets eliminated in the computation. Thus, a person with strong interest in both children and adults will have DSP scores comparable to a person with weak interest in both target categories.

In addition to the DSI/DSP distinction, latency-based indirect measures of DSI can be grouped into two distinct measurement approaches: task-relevant and task-irrelevant paradigms. *Task-relevant indirect DSI/DSP measures* involve the explicit categorization of sexual target categories – either as sexually relevant themselves or in combination with classification trials of sexual attributes. Due to the explicit task requirement to process sexual relevance, it is fairly transparent to the respondent that DSI is the diagnostic construct of interest. However, as respondents are not usually informed that response latencies are the central dependent variable and the underlying rationale of the diagnostic inference is unknown to them, these measures qualify as indirect. *Task-irrelevant measures* are based on the detection of sexually irrelevant characteristics (e.g., location, colour, semantic meaning) of target stimuli that are presented together with distracting sexually relevant vs. irrelevant background stimuli (e.g., adults, children). The underlying rationale is that sexually relevant background stimuli interfere with the primary detection task due to attentional capture resulting in increased response latencies.

Task-relevant indirect measures of deviant sexual interests

5

Viewing time tasks. Viewing times (VT) were the first ever latency-based indirect measure of sexual interest (Rosenzweig, 1942). In the standard VT procedure, participants are asked to evaluate pictures of target individuals on a graded scale of sexual attractiveness/arousal. The response latency of this judgment is unobtrusively measured. It is a robust finding that this response latency is longer for sexually attractive as compared to sexually unattractive targets and, in turn, VT measures can be used to discriminate between participants with respect to sexual preference (e.g., Imhoff, Schmidt, Nordsiek, Luzar, Young, & Banse, 2010).

The underlying processes driving the robust effects of longer response latencies for sexually attractive targets are not entirely clear. Three mechanisms have been frequently proposed: (a) deliberate delay due to the hedonic quality of sexually preferred targets, (b) automatic attentional adhesion, (c) slowing down of decision making processes after the presentation of explicit erotic stimuli (sexual content induced delay; Geer & Bellard, 1996). We conducted a series of experiments providing the first causal tests of VT processes (Imhoff et al., 2010). Deliberate delay to keep pleasant stimuli in sight longer was ruled out because prolonged response latencies also emerged for relevant sexual attractiveness ratings in the absence of target pictures that had been presented beforehand for a fixed amount of time. Additionally, VT effects emerged when restricted response windows of 1000 *ms* were utilized. Furthermore, attentional adhesion to sexually attractive stimuli also could not fully explain VT effects: If sexually attractive stimuli lead to longer response latencies because the stimuli automatically capture participants' attention and distract them from the rating task, VT effects should vanish when the sexually attractive stimuli are taken away from the participants prior to rating sexual attractiveness. However, as described above, this was not the case. Finally, VT effects still emerged for stimuli depicting target stimuli's heads without any further indications of erotic content, thereby ruling out sexual content induced delay as a causal expalanation. These results raise the question of whether the term "viewing time" is a misnomer, as participants saw all stimuli for the same amount of time under restricted viewing conditions but still differed in their response latencies.

Accordingly, standard VT effects should be described as prolonged decision latencies for sexually attractive targets (Imhoff et al., 2010).

As a consequence, two further mechanisms remained as plausible explanations for VT effects even in absence of a stimulus while responding: (a) automatic time-consuming schematic processes triggered by sexually attractive stimuli (stimulus-specific effects) and (b) cognitive demands associated with the task of rating sexual attractiveness (task-specific effects). Imhoff, Schmidt, Weiß, Young, and Banse (2012) disentangled stimulus- and task-specific effects by manipulating the sexual orientation perspective under which male participants responded to standard VT tasks. It was shown that VT effects were predominantly a function of the assigned perspective (task-specific account) and not dependent on participants' sexual orientation (stimulus-specific account). In other words, sexual attractiveness ratings from a vicarious (e.g., heterosexual) perspective took longest when the targets were adult females, regardless of participants' actual hetero- or homosexual orientation. This is at odds with the notion that VT measures primarily tap into hot automatic processes elicited by sexually attractive stimuli rather than task-dependent response strategies (e.g., scrutinizing whether the stimulus exhibits the right age, sex, and/or attractiveness for being a sexually exciting stimulus). The latter process is based on the assumption that the more criteria for endorsing sexual attraction have to be affirmed, the longer the scrutinizing takes (i.e., VTs increase), whereas rejection of any criterion allows the process to stop immediately (i.e., VTs decrease). In line with this account of task- vs. stimuli-specific effects, Glasgow (2009) reported that neither perceived sexual competition of female mate rivals nor filial affection to children in heterosexual women – both stimulus characteristics hypothesized to increase VTs for sexually irrelevant categories – conflated standard VT effects while rating sexual attractiveness.

Since the first demonstrations of VT effects for child sexual offenders (Harris et al., 1996), there have been numerous independent forensic replications (for an overview of VT studies on DSI see Table 1). Viewing time tasks have been shown to distinguish between child sexual offenders and non-offending controls (Glasgow, 2009; Harris et al., 1996; ds = 1.61 and 1.00), mixed community and

7

offender controls (Banse et al., 2010; Fromberger et al., in press; *d*s from 0.76 to 0.82), as well as

varying offender control groups such as non-sexual offenders (Babchishin, Nunes, & Kessous, 2012;

Banse et al., 2010; *d*s from 0.86 to 1.84) and adult sexual offenders (Abel et al., 2004; Gress, 2005;

Worling, 2006; *d*s from 0.51 to 1.08). They also differentiated between different subtypes of child

sexual offenders (e.g., child sexual offenders who victimized boys or boys and girls vs. only girls;

Gress, 2005; Schmidt, Gykiere, Vanhoeck, Mann, & Banse, in press; *d*s from 0.84 to 1.65).

Reports of internal consistency on raw latencies (Cronbach's α) for forensic VT measures

generally ranged between .72 and .93 (with only two notable exceptions of .60 [male child stimuli;

Letourneau, 2002] and .62 [male adolescent stimuli; Worling, 2006]; see Table 1). Sets of African

American stimuli were tested in two studies (Abel et al., 1998; Letourneau, 2002), but only the latter

author reported origin-specific analyses (internal consistency of the African American categories was

comparable to the Caucasian stimulus set with αs ranging from .72 to .87). These generally satisfying

to good coefficients might overestimate the reliability of VT measures as the calculations might be

confounded with general executive functioning (i.e., reaction speed). However, general classification

speed as assessed by a different task was found to be unrelated to VT DSI/DSP measures (Schmidt et

al., in press). At present, no data on retest reliability are available for VT DSI measures.

Viewing time DSI measures oftentimes converge with corresponding self-report measures

(e.g., Abel et al., 1998; Babchishin, et al., 2012; Banse et al., 2010; Glasgow, 2009; Harris et al., 1996;

Worling, 2006) as well as DSP Implicit Association Tests (Babchishin, Nunes, & Hermann, in press;

Banse et al., 2010). However, particularly in light of problems with self-reports, comparing VT with

other non-self-reported measures seems advisable. A prime candidate for this is penile

plethysmography (PPG), often regarded as the most valid measure of DSI (Seto, 2008; regarding

methodological shortcomings see, e.g., Kalmus & Beech, 2005). Two studies that concurrently

utilized VT and PPG measures of DSI/DSP confirmed their convergent validity (Letourneau, 2002;

Stinson & Becker, 2008; *r*s between .28 and .61). However, another study reported a negative

association between VT and PPG indexes (*r* = -.47; Babchishin et al., 2012), rendering the findings on

convergence of VT with PPG assessments inconclusive. The Screening Scale for Pedophilic Interests (Seto & Lalumière, 2001) is another indicator of pedophilic interest based on an index of offending behavior. It is phallometrically validated and also associated with recidivism risk (Seto, Harris, Rice, & Barbaree, 2004). VT measures were reported to converge with the Screening Scale for Pedophilic Interests (Banse et al., 2010; Schmidt et al., in press). Additionally, Schmidt et al. (in press) reported preliminary support for positive correlations of VT DSI/DSP measures with recidivism risk as assessed by standard actuarial risk indicators (Static-99R; Helmus, Thornton, Hanson, & Babchishin, 2012).

At present, no data have been published on the fakeability of VT paradigms when used in forensic contexts (as this would risk informing at least a number of offenders of the underlying scoring procedures). Obviously, VT tasks are easy to fake once the measurement rationale is known. In line with this, naïvely dissimulating pedophiles exhibited were significantly less accurately classified than non-dissimulating pedophilic child sexual offenders (Gray & Plaud, 2005; d = -2.13). This finding can be criticized in terms of its post-hoc classification algorithms for the dissimulators and its strong sample selection effects (Sachsenmaier & Gress, 2009). Opposing evidence stems from another study: VT tasks did not show differences between non-informed deniers vs. admitters of child sexual offending such that both deniers and admitters could be discriminated from non-sexual offender controls (Babchishin et al., 2012; ds = 1.22 and 1.32, respectively). The results from Babchishin et al. (2012) thus provide the first evidence of VT tasks' robustness against non-informed dissimulation although this finding awaits replication.

In summary, VT tasks can be regarded as among the most frequently researched latency-based measures of DSI. There have been numerous reports from different labs demonstrating that VT measures are satisfactorily reliable and valid indicators of DSI in forensic contexts. The VT effect has been regarded as so robust that there are commercially distributed VT paradigms (e.g., Abel, 1995). However, from a scientific perspective data based on Abel's VT routines have to be treated with some caution as crucial methodological details have not been published (for a critical overview see

Sachsenmaier & Gress, 2009). Furthermore, VT tasks have been shown to be robust against deniers (Babchishin et al., 2012) uninformed about successful faking strategies.

On the other hand, the predominantly task-driven nature of VT effects is a potential threat to the diagnostic validity of VT paradigms. The task-dependency also cautions against the interpretation of VT effects as caused by automatic processes elicited by sexually attractive stimuli and outside of conscious control (i.e., attentional adhesion; Imhoff et al., 2010). Thus, participants' compliance in completing the secondary rating task from their own self-relevant perspective is of crucial importance. VT measures result in good differentiations of sexually deviant from non-deviant samples only as long as participants comply with the instructions to rate how subjectively sexually attractive targets are. However, when participants (naïvely or knowingly) complete the task from a perspective other than their own (Imhoff et al., 2012) or with a completely different task (e.g., rating age of the target; Petruschke, Imhoff, Banse, & Weber, in preparation), latency patterns in standard VT paradigms will most likely be nondiagnostic.

Implicit Association Tests. The Implicit Association Test (IAT) introduced by Greenwald et al. (1998) is another prominently researched latency-based indirect measure. In forensic contexts, the prototypical Children/Adults DSP IAT is based on two double discrimination tasks – the so-called critical blocks – assessing associative strengths between target categories (e.g., *Children* vs. *Adults*) and attribute categories (e.g., *Sexually exciting* vs. *Sexually unexciting*), both arranged on bipolar dimensions (for a detailed description of the assessment procedure and underlying processes see Perugini et al., this volume). Classical IATs are usually scored by calculating the difference between the mean response latencies of compatible and incompatible critical blocks, divided by the pooled standard deviation of the response latencies (Greenwald, Nosek, & Banaji, 2003). Given that this calculation depends on a standardized difference index, DSP IATs are inherently effect size measures (analogous to Cohen's *d*) of relative DSP (as opposed to absolute DSI measures such as raw VTs).

There have been several independent reports of DSP IAT effects in forensic populations (for an overview see Table 1). DSP IATs differentiated between child sexual offenders and non-offending controls (Mihailides et al., 2004; Nunes et al., 2007; ds 0.71 and 0.92), mixed community and offender controls (Banse et al., 2010; ds from 0.43 to 0.82), as well as varying offender control groups such as non-sexual offenders (Brown et al., 2009; Mihailides et al., 2004; ds 0.92 and 0.95, respectively) or adult sexual offenders (Gray et al., 2005; d = 0.84). Furthermore, DSP IATs distinguished between child sexual offenders who victimized either only boys or boys and girls vs. only girls (Schmidt et al., in press; ds from 0.64 to 0.72) and child sexual offenders whose victims were under twelve years of age vs. twelve years and older (Brown et al., 2009; d = 0.77). Van Leeuwen et al. (2012) reported strong DSP IAT differences between self-identified community pedophiles and non-offending controls (d = 1.74). DSP IATs were not confounded by general classification speed abilities (Schmidt et al., in press). In a meta-analysis, Babchishin et al. (in press) reported a mean DSP IAT effect of d = 0.63 between child molesters and non-molesters. As expected, group differences were largest for comparisons of child sexual offenders to non-offenders (d = 0.96), followed by comparisons to non-sexual offenders (d = 0.58) and to rapists (d = 0.48). Notably, treatment participation was a significant moderator of IAT effects: DSP IATs showed larger effects for child sexual offenders who had not undergone treatment than for treated child sexual abusers when compared with control groups. These findings corroborate either treatment effects on indirectly assessed DSP or confounds associated with child sexual offender treatment (e.g., group selection effects on behalf of suspected lower DSI levels).

Retest-reliability was tested only once for DSP IATs (r_{tt} = 0.63; Brown et al., 2009). Reports of internal consistency (Cronbach's α) for DSP IATs comparing sexual interest in children vs. adults ranged between .72 and .83 (Table 1), thereby corroborating the satisfactory reliability of these measures. However, the only two studies using sex-specific DSP IATs (Banse et al., 2010; Schmidt et al., in press) reported lower alphas for *Boys/Men* (.61 and .65, respectively) in comparison to *Girls/Women* IATs (.79 and .82). This difference might be attributed to variance restrictions in typical

forensic samples: Homosexual orientation is usually overrepresented in randomly selected child sexual offender samples. This results in less clearly differentiated DSP patterns for *Boys/Men* IATs between child sexual offender and control groups: Homosexually oriented child sexual abusers show less DSP as they are likely to be interested in both boys and men, whereas heterosexual controls are interested in neither boys nor men and thus show less DSP as well (Banse et al., 2010; Schmidt et al., in press). Whether the underlying rationale of sex-specific DSP IATs to disentangle sexual orientation from sexual maturity preferences is a viable option to increase criterion validity is an open empirical question as sex-differentiated IATs produced smaller DSP effects than generic *Children/Adults* IATs (Babchishin et al., in press).

Convergent validity with other DSP measures was shown meta-analytically (Babchishin et al., in press): DSP IATs converged with moderate effect sizes with self-report, VT, and offence-behavioral measures of DSI (Screening Scale for Pedophilic Interests) as well as with actuarial estimates of recidivism risk ($r = .27$). No convincing DSP IAT associations with corresponding PPG indexes have been reported so far.

Although the IAT has been considered resistant to deliberate faking attempts, it has repeatedly been shown to be fakeable when respondents are informed about successful faking strategies (e.g., slowing of latencies in consistent blocks), are experienced with the paradigm, and/or are strongly motivated to fake results (Teige-Mocigemba et al., 2010). Nevertheless, there is preliminary evidence (Brown et al., 2009) that DSP IATs can distinguish between denying child sexual offenders and non-sexual offending control groups ($d = 1.01$) but not between denying and admitting child sexual offenders ($d = .27$; non-significant). On the other hand Babchishin et al. (2012) failed to show any differentiation between either deniers or admitters vs. non-sexual offender controls, respectively.

In summary, in addition to VT tasks, IATs have emerged as a second robust paradigm to indirectly assess DSI/DSP. Multiple studies from independent labs as well as a first meta-analysis (Babchishin et al., in press) demonstrated that IATs are reliable and valid indicators of DSP for

children over adults in forensic contexts. Still, some issues need further research. First, given that the interference effects that drive DSP IATs are based on the associative strength between concepts such as *Children* and *Sexual excitement,* where these associations originate from is an interesting question (Snowden et al., 2011). Is an association of children and sex a valid indicator of genuine DSP or an indicator of childhood experiences of sexual abuse – a condition quite prevalent among child sexual abusers (Seto, 2008)? Second, IATs have repeatedly been proven fakeable. Attempts to develop methodological strategies to detect and even correct deliberate manipulations of IAT results in forensic populations (Cvencek, Greenwald, Brown, Gray, & Snowden, 2010) need to be viewed with some caution. Statistics on the discriminatory power of detection algorithms are based on group means, which may have limited validity in single case diagnostics (i.e., relative group differences used to classify dissimulation are not available in single case assessments and are based on sample characteristics that might not be relevant to the actual case in question). Also, as these statistics are based on comparisons with response-latencies in consistent blocks of an uncritical baseline IAT (e.g., Gender/Self IAT; Cvencek et al., 2010), respondents who know this could easily start to fake the baseline IAT as well (by slowing latencies in the consistent block of the baseline IAT).It seems somewhat of a paradox to derive faking-resistant countermeasures from a measure that is fakeable in itself.

Implicit Relational Assessment Procedure. A fairly recent, task-relevant indirect measure is the Implicit Relational Assessment Procedure (IRAP) introduced by Dawson et al. (2009). In the IRAP, target categories (C*hildren* vs. *Adults*) and target words representing *sexual* vs. *non-sexual* attributes are presented in either consistent (in accordance with the respondent's individual associations) or inconsistent (at odds with the respondent's associations) pairings. During the task, participants are forced to categorize the presented pairings as either *true* or *false* according to predetermined contingencies: During one type of blocks (consistent for non-deviant individuals), participants are required to categorize adults as sexual (e.g., *Adult – Sexual – True*; *Adult – Nonsexual – False*) and children as nonsexual (e.g., *Child – Nonsexual – True*; *Child – Sexual – False*) as opposed to

(inconsistent) blocks during which the feedback contingencies were reversed, and participants were required to categorize adults as nonsexual (e.g., *Adult – Sexual – False*; *Adult – Nonsexual – True*) and children as sexual (e.g., *Child – Sexual – True*; *Child – Nonsexual – False*). The rationale of the IRAP is that it takes less time to respond positively to pairings that are consistent with beliefs than to pairings inconsistent with beliefs, because during consistent trials answer keys and initial responses are matched, whereas in inconsistent trials the initial response has to be inhibited and overcome with an alternative response that is incompatible with automatic individual associations. Similar to the IAT, the IRAP DSP index is calculated as a *d*-measure from the difference of response latencies in consistent vs. inconsistent blocks (Dawson et al., 2009). Dawson et al. (2009) were able to differentiate between child sexual abusers and non-offender controls (d = .91) and the IRAP DSP index was unrelated to years of education in their sample. No further psychometric properties were reported. Hence, the IRAP has to be regarded as among the least researched indirect measures of DSI/DSP with only preliminary findings concerning its validity.

Eye Movement Tracking Task. Fromberger et al. (in press) recently demonstrated the potential of assessing eye movements as another indirect measure of DSP. In a paired comparison task, pictures of girls vs. women and boys vs. men had to be classified according to which of the stimuli was more sexually attractive. Initial fixation latencies as well as relative fixation times aggregated into DSP indexes differentiated between pedophilic child sexual abusers and non-pedophilic (healthy and forensic non-child sexual offending) controls (ds = 1.84 and 1.34, respectively). Initial fixation latencies showed good diagnostic accuracy in terms of sensitivity and specificity. This preliminary finding holds promise for forensic purposes because initial fixation latencies are deemed quite robust against faking attempts as they are regarded as an indicator of automatic bottom-up information processing. However, average mean initial fixation latencies in the Fromberger et al. (in press) study were roughly one second, which cannot be regarded as indicating initial automatic information processing. Clearly, more research is needed on the reliability of these

eye movement measures as well as their potential to distinguish deniers and non-deniers from relevant control groups.

Task-irrelevant indirect measures of deviant sexual interests

Emotional Stroop Tasks. The classical Stroop interference paradigm (Stroop, 1935) has been among the first to be adapted to an indirect measure of DSP in forensic populations (for an overview see Price, Beech, Mitchell, & Humphreys, in press). Initially, Emotional Stroop variants in which participants had to classify the print colour of *sexual* vs. *neutral* word stimuli were utilized. Sexual words are hypothesized to produce longer response latencies compared to neutral words due to increased emotional salience that interferes with the colour classification task. In an initial study, Smith and Waterman (2004) found no such effect between child sexual offenders and rapists for sexual words representing child molesting and rape themes (only five offenders in each group). On the other hand, Price and Hanson (2007) were able to discriminate child sexual offenders from non-offending controls (d = 0.82) utilizing the same stimulus words as Smith and Waterman (2004) but failed to show any effects with an alternative, more offence-specific stimulus set. Another Emotional Stroop variant utilizing differently coloured pictorial stimuli of children vs. adults did not distinguish between child sexual offenders and non-offending control groups (Ó Ciardha & Gormley, 2012). Van Leeuwen et al. (2012) introduced a Picture-Word Stroop variant during which words superimposed on pictures of children vs. adults had to be classified as either sexual or neutral. Notably, contrary to the classic Emotional Stroop variants, sexually relevant pictures in this Picture-Word Stroop were presumed to facilitate classifications of sexual words. Self-identified pedophiles' response latencies were shown to differ from non-offending controls' on this DSP index in the expected directions (d = 1.41). As van Leeuwen et al. (2012) provide evidence for a facilitatory (as opposed to the traditional inhibitory) effect of sexually relevant images, it remains unclear whether the heterogeneity of these effects (Price et al., in press) are due to methodological factors (e.g., differing stimulus sets and Stroop variants) or to sample characteristics (e.g., intra- vs. extrafamiliar child sexual abusers). In

summary, there is inconclusive evidence for DSP Stroop paradigms as valid indicators of individual differences in forensic contexts.

Attentional Blink. Based on the attentional blink phenomenon (Raymond, Shapiro, & Arnell, 1992), Beech et al. (2008) introduced the Rapid Serial Visual Presentation Task (RSVP). This paradigm capitalizes on the fact that if the first target presented is sexually relevant, it interferes with the perception of the second target when the targets are presented in rapid succession. Beech and his colleagues showed that intra- and extrafamilial child molesters in contrast to non-sexual offender controls made more errors in reporting a second target when they were presented target pictures of children vs. pictures of animals (*d*s = 1.00 and 1.28, respectively) in an RSVP task. However, Crooks et al. (2009) did not replicate these findings in a sample of adolescent child molesters, leaving open the question of whether the findings might be explained by sample differences (i.e., adolescent child sexual offenders are deemed to exhibit lower DSI levels than adult child molesters) or by a lack of task validity.

Choice Reaction Time Task. In the prototypical Choice Reaction Time Task (Wright & Adams, 1994), individuals have to detect target probes (e.g., dots) that are superimposed on either sexually relevant or irrelevant pictorial stimuli (e.g., pictures of adults vs. children). It has been shown twice that a DSP index of mean response latencies for infants vs. adults discriminated between child sexual offenders and non-sex offending controls (Mokros et al., 2010; Pöppl et al., 2011; *d*s = 1.41 and 0.99, respectively) without further psychometric properties being reported.

In summary, attention-based/task-irrelevant paradigms have been quite successfully used in clinical populations where avoidance of threat or negative valence is claimed to be the source of the attentional bias (Cisler, Bacon, & Williams, 2009). Yet, corresponding DSI/DSP tasks lack a consistent pattern of effects congruent with the supposed rationale of selective attention. This might result from the fact that in the case of DSI/DSP positive valence and approach behavior associated with sexual interest might facilitate rather than divert attention allocation. Additionally, most attention-

based paradigms cannot disentangle initial attentional capture and subsequent difficulties in disengagement from the relevant stimuli (Fox, Derakshan, & Standage, 2011). This leads to theoretical problems in predetermining the directedness of the hypothesized attention biases in research with sexually relevant stimuli (Prause, Janssen, & Hetrick, 2008), leaving it unclear whether potential group differences represent sexual interest or other sources of attention biases (e.g., phobic avoidance). More elaborate theoretical frameworks that clarify the relationship between attention biases and DSI are needed. Importantly, data on reliability – a common problem with attention-based measures of individual differences (Cisler et al., 2009) – are generally missing for task-irrelevant measures. Hence, it is fair to conclude that task-irrelevant approaches are currently the least developed indirect DSI measures.

What Goes Up Must Come Down – Implicit Assumptions about Implicit Measures

New developments often foster excessive and to some degree illusory expectations. This is certainly true for indirect/implicit measures (Perugini & Banse, 2007). It is thus necessary to thoroughly examine the empirical foundations of the implicit assumptions about implicit measures (for an overview see Gawronski, 2009). Probably the most common beliefs about implicit measures are that they assess subconscious associations not accessible through introspection, and, relatedly, that they are therefore not fakeable. Likewise, it is often claimed that implicit measures circumvent problems of social desirability because respondent s are supposedly not able to adjust their responses on indirect measures. But lack of introspective access does not necessarily imply that the associations are subconscious (De Houwer, 2006). In fact, empirical results point quite to the contrary (e.g., Gawronski, Hofmann, & Wilbur, 2006). Furthermore, although indirect measures are obviously not as easy to fake as self-report measures, both VTs and IATs are fakeable under specific boundary conditions as laid out in the sections above (see http://www.innocentdads.org/abel.htm for detailed instruction on how to fake VT DSI measures; Cvencek et al., 2010). Ultimately, it is likely that no scientific, psychological measure ever will be completely immune to faking attempts, although resistance across measures will vary along a continuum and indirect measures are obviously

situated on the more resistant end. Paradigms that tap into early levels of bottom-up/stimulus-driven processes such as startle probe reflex (Hecker, King, & Scoular, 2009) or anti-saccade tasks (Fox et al., 2011) seem to be promising future options as they might be even more resilient against faking attempts and deliberative top-down regulation.

Another interesting conundrum concerns the idea of the "true value" or the "true self". All the issues concerning social desirability and fakeability imply that indirect/implicit measures are able to tap into individuals' genuine attitudes, opinions, or sexual interests that the respondent in forensic contexts is motivated to conceal from self-reports. However, this is dependent on what one would psychologically regard as the "true self". On the one hand, it might be assumed that the "true self" is revealed under circumstances of failing deliberate control (e.g., disinhibition from alcohol consumption giving rise to the true self). On the other hand, it might be hypothesized that the "true self" can be inferred from what a person deliberately does and explicitly chooses in a controlled mode (Gawronski, 2009). Theoretically, from a dual-systems perspective it can be claimed that indirect measures should predict spontaneous behaviors whereas explicit measures should be related to deliberate behaviors. However, there is a whole set of situational and personal moderators of these relationships (Friese et al., 2008; Perugini et al., 2010) underscoring that especially in applied forensic diagnostics one has to be cautious not to draw diagnostic inferences from single direct or indirect measures. All these sources of measurement error conflate criterion/predictive validity and thus it is questionable to interpret any single measure as an absolute index of a specific psychological attribute. Diagnostic conclusions are much safer if they are derived from multiple valid, convergent, and conceptually different measures that tap into unique parts of criterion variance. Corroborating this, combining direct and indirect measures into test batteries has been proven as incrementally valid over and above single (direct and indirect) DSI indicators in child sexual abuser populations (Babchishin et al., 2012; Banse et al., 2010).Therefore, in forensic contexts the pressing question remains when and under what boundary conditions implicit/indirect measures are incrementally valid predictors of specific behaviors above and beyond explicit measures.

What Should We Aim for? A Research Outlook

Important steps on the way to developing DSI measurement tasks have been achieved (Thornton & Laws, 2009b). The current state of research suggests that VTs and IATs are the most promising and best validated indirect measures of DSI/DSP. Several replications have been successfully conducted with independent samples and by different research labs. Preliminary work on the effects of various moderators on task performance (e.g., general reaction speed; Schmidt et al., in press; denial; Babchishin et al., 2012; Brown et al., 2009) has been undertaken.

Methodological aims. In terms of methodology, future research should aim to increase the reliability of indirect DSI/DSP measures and routinely report relevant coefficients that are based on corresponding test subsets (e.g., split-half coefficients, Cronbach's α). An effective strategy to optimize reliability is to use a sufficient number of trials. Additionally, in order to maximize differences between individuals (as opposed to experimental conditions), a fixed random stimulus order identical for all participants should be used because a fully randomized stimuli order adds unnecessary portions of random error to the measure.

More research focusing on convergent and discriminant validity with other established measures of DSI is needed. Each single validation criterion is plagued with its own set of problems. Self-reported DSI is regarded as amenable to various impression management influences. Sexual delinquent behavior such as child molesting represents a criminological/judicial category rather than a specific indication of a psychological construct such as DSI/DSP (empirically only up to 50% of child sexual offenders exhibit pedophilic DSP; Seto, 2008). Clinical pedophilia diagnoses suffer from low reliability and/or validity as these are usually based on inferences from offence behavior rendering them tautological. Thus, future research should preferably focus on behavioral measures based on PPG or sexual behavior on the internet. Assessing these additional behaviors might be the only way to solve the paradox of not having a directly accessible validation criterion for DSI.

Standardization is another important aim. Phallometric assessments have been extensively criticized for a lack of standardization of stimuli, assessment procedural, and scoring (e.g., Kalmus & Beech, 2005). This critique has been routinely named as one of the main reasons for the development of indirect DSI measures. However, at present indirect measures of DSI are far from being standardized, too – either in terms of stimuli or scoring algorithms. Such idiosyncrasies constitute barriers against comparing results and undermine knowledge accumulation. An easily accomplished way of standardizing latency-based measures might be to use an analog to the *d*-measure of the IAT (Greenwald et al., 2003) based on the standardized mean differences of the relevant sexual interest categories as most measures rely on a DSP difference index comparing responses to child vs. adult stimuli.

Theoretical aims. Future research needs to address questions pertaining to why and how indirect DSI/DSP measures differentiate between offender subgroups, as well as the boundary conditions affecting their validity. Therefore, it is highly advisable to control for factors such as intra- vs. extrafamilial child sexual offending, antisociality/psychopathy, pre- vs. postpubescent victims, victim sex, sexual orientation, and/or risk levels to disentangle the influence of sample characteristics from methodological issues. Apart from these issues, there is need for theoretical clarification of exactly what indirect measures assess and how implicit/indirect sexual interest indicators relate to sexual behavior when behavior contradicts explicit sexual interests. The exact relationship between latency-based DSI measures and actual (for example, physiologically assessed) sexual arousal is as of yet unclear (Ó Ciardha, 2011). As latency-based sexual interest indications cannot be regarded as the same as physiological sexual arousal, the relation between these two observational levels should be clarified for each latency-based measure and indirect measures as a whole.

Clinical aims. Ultimately, a desirable goal would be not only to methodologically improve and theoretically better understand measures of DSI/DSP, but to make them more accessible to clinical usage. As one example of an approach towards applied usability, Banse et al. (2010) have created the *Explicit and Implicit Sexual Interest Profile (EISIP)* – a user-friendly test battery that produces profile

outputs that can be interpreted by clinicians outside of research laboratory settings immediately

after the assessment. Additionally, future work might focus on the development of norms for

relevant offender and non-offender populations. Finally, data on predictive validity, the most

relevant piece of the puzzle for applied purposes, are still missing. Nevertheless, as phallometrically

assessed DSI has been proven as among the best predictors of sexual reoffending (Mann et al., 2010),

high hopes might be put into VTs and IATs – the most valid DSI/DSP measures – as less costly and

laborious adjuncts to PPG assessments. Preliminary cross-sectional reports of correlations with

actuarial risk assessment instruments (Schmidt et al., in press) and convergence of VT with PPG

measures (Letourneau, 2002; Stinson & Becker, 2008) corroborate that this long-term research effort

is worthwhile to pursue.

References

Abel, G. G. (1995). *The Abel Assessment for Sexual Interest–2 (AASI–2)*. Atlanta, GA: Abel Screening Inc.

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 of Research and Treatment, 10*, 81–95. doi:10.1177/107906329801000202

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: A Journal of Research and Treatment, 16*, 255–265. doi:10.1177/107906320401600306

Babchishin, K. M., Nunes, K. L., & Hermann, C. (in press). The validity of Implicit Association Test (IAT) measures of sexual attraction to Children: A meta-analysis*. Archives of Sexual Behavior.* doi:10.1007/s10508-012-0022-8

Babchishin, K. M., Nunes, K. L., & Kessous, N. (2012). *A multimodal examination of sexual interest in children.* Manuscript submitted for publication.

Banse, R., Schmidt, A. F., & Clarbour, J. (2010). Indirect measures of sexual interest in child sex offenders: A multi-method approach. *Criminal Justice and Behavior, 37*, 319–335. doi:10.1177/0093854809357598

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*, 397-402. doi:10.1037/a0013587

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*, 166-180. doi:10.1177/1079063209332234

Cisler, J. M., Bacon, A. K., Williams, N. L. (2009). Phenomenological characteristics of attentional biases towards threat: A critical review. *Cognitive Therapy and Research, 33*, 221-234. doi:10.1007/s10608-007-9161-y

Crooks, V. L., Rostill-Brookes, H., Beech, A. R., & Bickley, J. A. (2009). Applying Rapid Serial Visual

Presentation to adolescent sexual offenders: Attentional bias as a measure of deviant sexual

interest? *Sexual Abuse: A Journal of Research and Treatment, 21*, 135–148.

doi:10.1177/1079063208328677

Cvencek, D., Greenwald, A. G., Brown, A. S., Gray, N. S., & Snowden, R. J. (2010). Faking of the Implicit

Association Test is statistically detectable and partly correctable. *Basic and Applied Social

Psychology, 32*, 302–314. doi:10.1080/01973533.2010.519236

Dawson, D. L., Barnes-Holmes, D., Gresswell, D. M., Hart, A. J. P., & Gore, N. J. (2009). Assessing the

implicit beliefs of sexual offenders using the Implicit Relational Assessment Procedure: A first

study. *Sexual Abuse: A Journal of Research and Treatment, 21*, 57–75.

doi:10.1177/1079063208326928

De Houwer, J. (2006). What are implicit measures and why are we using them. In R. W. Wiers & A. W.

Stacy (Eds.), *The handbook of implicit cognition and addiction* (pp. 11-28). Thousand Oaks, CA:

Sage Publishers.

Friese, M., Hofmann, W., & Schmitt, M. (2008). When and why do implicit measures predict

behaviour?: Empirical evidence for the moderating role of opportunity, motivation, and

process reliance. *European Review of Social Psychology, 19*, 285-338.

doi:10.1080/10463280802556958

Fox, E., Derakshan, N., & Standage, H. (2011). The assessment of human attention. In K. C. Klauer, A.

Voss, & C. Stahl (Eds.), *Cognitive methods in social psychology* (pp. 15-47). New York, NY:

Wiley.

Fromberger, P., Jordan, K., Steinkrauss, H., von Herder, J., Witzel, J., Stolpmann, G., Kröner-Herwig,

B., & Müller, J. L. (in press). Diagnostic accuracy of eye movements in assessing pedophilia.

Journal of Sexual Medicine. doi: 10.1111/j.1743-6109.2012.02754.x

Gawronski, B. (2009). Ten frequently asked questions about implicit measures and their frequently

supposed, but not entirely correct answers. *Canadian Psychology, 50*, 141-150.

doi:10.1037/a0013848

Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are "implicit" attitudes unconscious?

Consciousness and Cognition, 15, 485–499. doi:10.1016/j.concog.2005.11.007

Geer, J. H., & Bellard, H. S. (1996). Sexual content induced delays in unprimed lexical decisions:

Gender and context effects. *Archives of Sexual Behavior, 25, 91–107.* doi:10.1007/BF02437581

Glasgow, D. V. (2009). Affinity: The development of a self-report assessement of paedophile sexual

interest incorporating a viewing time validity measure. In D. Thornton & D. R. Laws (Eds.),

Cognitive approaches to the assessment of sexual interest in sexual offenders (pp. 59-84)*.*

Chichester, UK: Wiley-Blackwell. doi:10.1002/9780470747551.ch3

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*,

304-308. doi:10.1037/0021-843X.114.2.304

Gray, S. R., & Plaud, J. J. (2005). A comparison of the Abel Assessment for Sexual Interest and penile

plethysmography in an outpatient sample of sexual offenders. *Journal of Sexual Offender*

Commitment: Science and the Law, 1, 1-10. Retrieved from http://www.soccjournal.org/

Grieger, L., Hosser, D., & Schmidt, A. F. (2012). Predictive validity of self-reported self-control for

different forms of recidivism. *Journal of Criminal Psychology, 2,* 80-95.

doi:10.1108/20093821211264405

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in

implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology,*

74, 1464-1480. doi:10.1037//0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit

Association Test: I. An improved scoring algorithm. *Journal of Personality and Social*

Psychology, 85, 197-216. doi:10.1037/0022-3514.85.2.197

Gress, C. L. Z. (2005). Viewing time measures and sexual interest: Another piece of the puzzle. *Journal of Sexual Aggression, 11*, 117–125. doi:10.1080/13552600500063666

Gress, C. L. Z., & Laws, R. D. (2009). Measuring sexual deviance: Attention-based measures. In A. R. Beech, L. A. Craig, & K. D. Browne (Eds.), *Assessment and treatment of sex offenders: A Handbook* (pp. 109-128). New York, NY: Wiley-Blackwell.

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*, 389-394. doi:10.1016/0005-7967(95)00070-4

Hecker, J. E., King, M. W., & Scoular, R. J. (2009). The startle probe reflex: An alternative approach to the measurement of sexual interest. In D. Thornton & D. R. Laws (Eds.), *Cognitive approaches to the assessment of sexual interest in sexual offenders* (pp. 59-84)*.* Chichester, UK: Wiley-Blackwell. doi:10.1002/9780470747551.ch10

Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse: A Journal of Research and Treatment, 24,* 64-101. doi: 10.1177/1079063211409951

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,* 1275–1288. doi:10.1007/s10508-009-9595-2

Imhoff, R., Schmidt, A. F., Weiß, S., Young, A. W., & Banse, R. (2012). Vicarious Viewing Time: Prolonged response latencies for sexually attractive targets as a function of task- or stimulus-specific processing. *Archives of Sexual Behavior, 41,* 1389-1401. doi: 10.1007/s10508-011-9879-1

Kalmus, E., & Beech, A. R. (2005). Forensic assessment of sexual interest: A review. *Aggression and Violent Behavior, 10*, 193-218. doi:10.1016/j.avb.2003.12.002

Letourneau, E. J. (2002). A comparison of objective measures of sexual arousal and interest: Visual

reaction time and penile plethysmography. *Sexual Abuse: A Journal of Research and*

Treatment, 14, 207–223. doi:10.1177/107906320201400302

Mann, R. E., Hanson, K. R., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some

proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of*

Research and Treatment, 22, 191-217. doi:10.1177/1079063210366039

Mihailides, S., Devilly, G. J., & Ward, T. (2004). Implicit cognitive distortions and sexual offending.

Sexual Abuse: A Journal of Research and Treatment, 16, 333-350.

doi:10.1177/107906320401600406

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*,

1081-1090. doi:10.1007/s10508-009-9530-6

National Research Council, Committee to Review the Scientific Evidence on the Polygraph. (2003).

The polygraph and lie detection. Washington, DC: National Academy Press.

Nunes, K. L., Firestone, P., & Baldwin, M. W. (2007). Indirect assessment of cognitions of child sexual

abusers with the Implicit Association Test. *Criminal Justice and Behavior, 34*, 454-475.

doi:10.1177/0093854806291703

Ó Ciardha, C. (2011). A theoretical framework for understanding deviant sexual interest and cognitive

distortions as overlapping constructs contributing to sexual offending against children.

Aggression and Violent Behavior, 16, 493-502*.* doi:10.1016/j.avb.2011.05.001

Ó Ciardha, C., & Gormley, M. (2012). Using a pictorial-modified Stroop Task to explore the sexual

interests of sexual offenders against children. *Sexual Abuse: A Journal of Research and*

Treatment, 24, 175-197. doi:10.1177/1079063211407079

Price, S. A., Beech, A. R., Mitchell, I. J., Humphreys, G. W. (in press). The promises and perils of the

emotional Stroop task: A general review and considerations for use with forensic samples.

Journal of Sexual Aggression, 17, 1-16. doi:10.1080/13552600.2010.545149

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

Perugini, M. & Banse, R. (2007). Editorial: Personality, implicit self-concept and automaticity. *European Journal of Personality, 21*, 257-261. doi:10.1002/per.637

Perugini, M., Richetin, J., & Zogmaister, C. (2010). Prediction of behavior. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of social cognition – Measurement, theory, and applications* (p. 255-277). New York, NY: Guilford.

Petruschke, P., Imhoff, R., Banse, R., & Weber (in preparation). Bottom-up versus top-down responding to sexually preferred stimuli: an fMRI study. Manuscript in preparation.

Pöppl, T. A., Nitschke, J., Dombert, B., Santtila, P., Greenlee, M. W., Osterheider, M., & Mokros, A. (2011). Functional cortical and subcortical abnormalities in pedophilia: A combined study using a choice reaction time task and fMRI. *Journal of Sexual Medicine, 8*, 1660-1674. doi:10.1111/j.1743-6109.2011.02248.x

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

Raymond, J. E., Shapiro, K. L., & Arnell, K. A. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18,* 849–860. doi:10.1037//0096-1523.18.3.849

Rosenzweig, S. (1942). The photoscope as an objective device for evaluating sexual interest. *Psychosomatic Medicine, 4,* 150–157.

Sachsenmaier, S. J., & Gress, C. L. Z. (2009). The Abel Assessment for Sexual Interests-2: A critical review. In D. Thornton & D. R. Laws (Eds.), *Cognitive approaches to the assessment of sexual interest in sexual offenders* (pp. 31–57). Chichester, UK: Wiley-Blackwell. doi:10.1002/9780470747551.ch2

Schmidt, A. F., Gykiere, K., Vanhoeck, K., Mann, R. E., & Banse, R. (in press). Direct and indirect

measures of sexual maturity preferences differentiate subtypes of child sexual abusers. *Sexual*

Abuse: A Journal of Research and Treatment.

Seto, M. C. (2008). *Pedophilia and sexual offending against children: Theory, assessment and*

intervention. Washington, DC: APA. doi:10.1037/11639-000

Seto, M. C., Harris, G. T, Rice, M. E., & Barbaree, H. E. (2004). The Screening Scale for Pedophilic

Interests predicts recidivism among adult sex offenders with child victims. *Archives of Sexual*

Behavior, 33, 455-466. doi:10.1023/B:ASEB.0000037426.55935.9c

Seto, M. C., & Lalumière, M. L. (2001). A brief screening scale to identify pedophilic interests among

child molesters. *Sexual Abuse: A Journal of Research and Treatment, 13,* 15-25.

doi:10.1177/107906320101300103

Snowden, R. J., Craig, R. L., Gray, N. S. (2011). Indirect behavioral measures of cognition among

sexual offenders. *Journal of Sex Research, 48*, 192-217. doi:10.1080/00224499.2011.557750

Snowden, R. J., & Gray, N., S. (2010). Implicit social cognition in forensic settings. In B. Gawronski, &

B. K. Payne (Eds.), *Handbook of implicit social cognition – measurement, theory, and*

applications (pp. 522-534).

Smith, P., & Waterman, M. (2004). Processing bias for sexual material: The Emotional Stroop and

sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 16*, 163–171.

doi:10.1177/107906320401600206

Stinson, J. D., & Becker, J. V. (2008). Assessing sexual deviance: A comparison of physiological,

historical, and self-report measures. *Journal of Psychiatric Practice, 14*, 379-388.

doi:10.1097/01.pra.0000341892.51124.85

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental*

Psychology, 18, 643-662. doi:10.1037/h0054651

Teige-Mocigemba, S., Klauer, K. C., & Sherman, J. W. (2010). A practical guide to Implicit Association

Tests and related tasks. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of social cognition –*

Measurement, theory, and applications (p. 117-139). New York, NY: Guilford.

Thornton, D., & Laws, D. R. (2009a). *Cognitive approaches to the assessment of sexual interest in*

sexual offenders. Chichester, UK: Wiley-Blackwell. doi:10.1002/9780470747551

Thornton, D., & Laws, D. R. (2009b). Postscript: Steps towards effective assessment of sexual interest.

In D. Thornton & D. R. Laws (Eds.), *Cognitive approaches to the assessment of sexual interest in*

sexual offenders (pp. 59-84)*.* Chichester, UK: Wiley-Blackwell.

doi:10.1002/9780470747551.ch11

Van Leeuwen, M., van Baaren, R., Chakhssi, F., Loonen, M., Lippman, M., & Dijksterhuis, A. (2012).

Detecting implicit paedophilic preferences: Improving predictability. Manuscript submitted for

publication.

Verschuere, B., Ben-Shakar, G., & Meijer, E. (2011). *Memory Detection: Theory and Application of the*

Concealed Information Test. Cambridge, UK: Cambridge University Press.

Walters, G. D. (2006). Risk-appraisal versus self-report in the prediction of criminal justice outcomes.

Criminal Justice and Behavior, 33, 279-304. doi:10.1177/0093854805284409

Ward, T. (2000). Sexual offenders' cognitive distortions as implicit theories. *Aggression and Violent*

Behavior, *5,* 491–507. doi:10.1016/S1359-1789(98)00036-6

Ward, T., & Beech, T. (2006). An integrated theory of sexual offending. *Aggression and Violent*

Behavior, 11, 44-63. doi:10.1016/j.avb.2005.05.002

Worling, J. R. (2006). Assessing sexual arousal with adolescent males who have offended sexually:

Self-report and unobtrusively measured viewing time. *Sexual Abuse: A Journal of Research and*

Treatment, 18, 383–400. doi:10.1177/107906320601800406

Wright, L. W.,& Adams, H. E. (1994). Assessment of sexual preference using a choice reaction time

task. *Journal of Psychopathology and Behavioral Assessment, 16*, 221–231.

doi:10.1007/BF02229209

Table 1. Overview of psychometric results from studies on latency-based indirect measures of deviant sexual interest in children.

Measure	Categories	Reliability	Validity		
			Group Comparison (*n*)	Effect-size Reported	Cohen's *d* Equivalent
Viewing Time (VT)					
Harris, Rice, & Quinsey (1996)	Adult-Child	n/a	CSO vs. NOC	*d* = 1.00**	1.00
			Girls-only CSO vs. NOC	*r* = .60***	1.50
Abel, Huffmann, Warberg, & Holland (1998)	Adult Male	α = .88			
	Adolescent Male	α = .89			
	Young Male	α = .87			
	Adult Female	α = .86			
	Adolescent Female	α = .84			
	Young Female	α = .90			
Letourneau (2002)	Male Children (age 2-4)	α = .60			
	Male Children (age 8-10)	α = .75			
	Male Adolescents (age 14-17)	α = .90			
	Male Adults (age 22 and over)	α = .90			
	Male Children (age 0-10)	n/a	CSO with boy victims (10) vs. SO (47)	*r* = .69**	2.51
	Female Children (age 2-4)	α = .87			
	Female Children (age 8-10)	α = .86			
	Female Adolescents (age 14-17)	α = .85			
	Female Adults (age 22 and over)	α = .80			
	Female Children (0-10)	n/a	CSO with girl victims (34) vs. SO (23)	*r* = .08	0.16
Abel et al. (2004)	Children-Adults	n/a	Adolescent CSO (1170) vs. Adolescent AC (534)	*AUC = .64*[snr]	0.51
Gress (2005)	Children-Adults	n/a	CSO (19) vs. Rapists (7)	Frequency Table	1.08*
			CSO with male or mixed victims (9) vs. CSO with female victims (17)	Frequency Table	1.65*
Gray & Plaud (2005)	n/a	n/a	Dissimulating pedophilic CSO (11) vs. Pedophilic CSO (28)	Frequency Table	-2.13*
			VT (39) vs. PPG (39)	Frequency Table	0.43
Worling (2006)	Prepubescent/Postpubescent[b]	n/a	CSO (52) vs. Sexual offender s with peer or adolescent victims (26)	*AUC*[b] = .61	0.40
	Male Toddlers	α = .82			
	Male Preadolescents	α = .79	CSO with two or more victims vs. SO	*AUC*[b] = .60	0.36
	Male Adolescents	α = .62	CSO with any male victims vs. SO	*AUC*[b] = .69**	0.70
	Male Adults	α = .72	CSO with only male victims vs. SO	*AUC*[b] = .73**	0.87
	Female Toddlers	α = .73	CSO with any female victims vs. SO	*AUC*[b] = .42	-0.29
	Female Preadolescents	α = .82	CSO with only female victims vs. SO	*AUC*[b] = .43	-0.25
	Female Adolescents	α = .77			

Measure	Categories	Reliability	Validity		
			Group Comparison (*n*)	Effect-size Reported	Cohen's *d* Equivalent
Glasgow (2009)	Female Adults	α = .77			
	Male Child	α = .93			
	Male Preadolescent	α = .8			
	Male Adolescent	α = .9			
	Male Adult	α = .89			
	Female Child	α = .92			
	Female Preadolescent	α = .87			
	Female Adolescent	α = .89			
	Female Adult	α = .93			
	Adult-Child	n/a	CSO (31) vs. NOC (31)	AUC = .87[snr]	1.59
Banse, Schmidt, & Clarbour (2010)	Postpubescent Males	α = .85	CS0 (38) vs. NSOC (37)/NOC (38)	AUC = .82*	1.29
	Postpubescent Females	α = .86		AUC = .56	0.21
	Prepubescent Males	α = .85		AUC = .80*	1.19
	Prepubescent Females	α = .77		AUC = .76*	1.00
	Children-Adults	n/a		AUC = .51	0.04
	Postpubescent Males		CSO with boy victims only (14) vs.	AUC = .89*	1.73
	Postpubescent Females		NSOC (37)	AUC = .63	0.47
	Prepubescent Males			AUC = .90*	1.81
	Prepubescent Females			AUC = .81*	1.24
	Children-Adults			AUC = .33	-0.62
	Postpubescent Males		CSO with girl victims only (16) vs.	AUC = .78*	1.09
	Postpubescent Females		NSOC (37)	AUC = .74*	0.91
	Prepubescent Males			AUC = .86*	1.53
	Prepubescent Females			AUC = .73*	0.87
	Children-Adults			AUC = .46	-0.14
Schmidt, Gykiere, Vanhoeck, Mann, & Banse (2012)	Postpubescent Males	α = .90	CSO with male or mixed victims (18) vs. Girls-only CS0 (36)	r = .47**	1.13
	Postpubescent Females	α = .90		r= -.37**	-0.84
	Prepubescent Males	α = .95		r = .47*	1.13
	Prepubescent Females	α = .93		r = .00	0.00
	Children-Adults	n/a		r =.42**	0.98
Babchishin, Nunes, & Kessous (2012)	Child-Adult	n/a	CSO (35) vs. NSOC (21)	d = 1.15*	1.15
			Admitting CSO (20) vs. NSOC (20)	d = 1.32*	1.32
			Denying CSO (12) vs. NSOC (20)	d = 1.22*	1.22
Fromberger et al. (in press)	Children-Adults	n/a	Pedophilic CSO (19) vs. AC (7)/NOC (48)	AUC = 0.76***	1.00

Measure	Categories	Reliability	Group Comparison (n)	Effect-size Reported	Cohen's d Equivalent
Implicit Association Test (IAT)					
Mihailides, Devilly, & Ward (2004)	Children-Not children/Sexual-Not sexual	n/a	CSO (25) vs. NSOC (25)	$t = 3.15$**	0.63
			CSO (25) vs. NOC (25)	$t = 4.76$***	0.95
Gray, Brown, MacCulloch, Smith, & Snowden (2005)	Children-Adults/Sex-Non-sex	n/a	CSO (18) vs. AC (60)	$d = 0.84$*	0.84
Nunes, Firestone, & Baldwin (2007)	Children-Adults/Sexy-Not sexy	n/a	CSO (24) vs. NOC (29)	$r = .33$*	0.70
	Children-Adults/Pleasant-Unpleasant	n/a	CSO (27) vs. NOC (29)	$r = .21$	0.43
Brown, Gray, & Snowden (2009)	Children-Adults/Sex-Non-sex	$\alpha = .80$ $r_{tt} = .63$	CSO with victims < 12 years of age (54) vs. CSO with victims > 12 years of age (21)	$d = 0.77$**	0.77
			CSO (54) vs. NSOC (49)	$d = 0.92$***	0.92
			Admitting CSO (20) vs. NSOC (49)	$d = 0.75$*	0.75
			Denying CSO (55) vs. NSOC (49)	$d = 1.01$**	1.01
			Denying (55) vs. Admitting CSO (20)	$d = 0.27$	0.27
Banse, Schmidt, & Clarbour (2010)	1. Boys-Men/Sexually exciting-Sexually unexciting	$\alpha = .65$	CS0 (38) vs. NSOC (37)/NOC (38)	$AUC = .62$*	0.43
	2. Girls-Women/Sexually exciting-Sexually unexciting	$\alpha = .79$		$AUC = .72$*	0.82
	3 .Children-Adults/Sexually exciting-Sexually unexciting	n/a		$AUC = .71$*	0.78
	1.		CSO with boy victims only (14) vs.	$AUC = .60$	0.36
	2.		NSOC (37)	$AUC = .67$	0.62
	3.			$AUC = .71$*	0.78
	1.		CSO with girl victims only (16) vs.	$AUC = .57$	0.25
	2.		NSOC (37)	$AUC = .56$	0.21
	3.			$AUC = .60$	0.36
Babchishin, Nunes, & Kessous (2012)	Children-Adults/Sexy-Not sexy	n/a	CSO (35) vs. NSOC (21)	$d = 0.44$	0.44
			Admitting CSO (22) vs. NSOC (21)	$d = 0.35$	0.35
			Denying CSO (13) vs. NSOC (21)	$d = 0.51$	0.51
Schmidt, Gykiere, Vanhoeck, Mann, & Banse (2012)	Boys-Men/Sexually exciting-Sexually unexciting	$\alpha = .61$	CSO with male or mixed victims (18) vs. Girls-only CS0 (36)	$r = .29$*	0.64
	Girls-Women/Sexually exciting-Sexually unexciting	$\alpha = .82$		$r = .23$	0.50
	Children-Adults/Sexually exciting-Sexually unexciting	n/a		$r = .32$*	0.72
Van Leeuwen et al. (2012)	Children-Adults/Sex-related-Neutral	n/a	SCP (20) vs. NOC (20)	$AUC = .89$[snr]	1.73

Measure	Categories	Reliability	Validity		
			Group Comparison (*n*)	Effect-size Reported	Cohen's *d* Equivalent
Implicit Relational Assessment Procedure (IRAP)					
Dawson, Barnes-Holmes, Gresswell, Hart, & Gore (2009)	Children-Adults	n/a	CSO (16) vs. NOC (16)	χ^2 = 5.489*	0.91
Eye Movement Tracking					
Fromberger et al. (in press)	Children-Adults (Initial fixation latency)	n/a	Pedophilic CSO (20) vs.	*AUC* = 0.90***	1.81
	Children-Adults (Relative fixaton time)	n/a	SO with adult victims (7)/NOC (48)	*AUC* = 0.83***	1.35
Choice Reaction Task (CRT)					
Mokros, Dombert, Osterheider, Zappalà, & Santilla (2010)	Infants-Adults	n/a	CSO (21) vs. NSOC (21)	*AUC* = 0.84**	1.41
Pöppl et al. (2011)	Infants-Adults	n/a	CSO (9) vs. NSOC (11)	*d* = 0.99*	0.99
Stroop Variants					
Smith & Waterman (2004; Emotional Stroop)	Sexual-Neutral	n/a	CSO (5) vs. Rapists (5)	*t* = 0.831	0.53
Price & Hanson (2007; Emotional Stroop)	Sexual-Neutral	n/a	CSO (15) vs. Rapists (15)	n/a	-0.45
			CSO (15) vs. Violent NSOC (15)	n/a	-0.03
			CSO (15) vs. Non-Violent NSOC (15)	n/a	0.14
			CSO (15) vs. NOC (15)	n/a	0.82*
	Child molesting-Neutral	n/a	CSO (15) vs. Rapists (15)	n/a	-0.28
			CSO (15) vs. Violent NSOC (15)	n/a	0.26
			CSO (15) vs. Non-Violent NSOC (15)	n/a	0.07
			CSO (15) vs. NOC (15)	n/a	0.58
	Rape-Neutral	n/a	CSO (15) vs. Rapists (15)	n/a	0.19
Ó Ciardha & Gormley (2012; Picture Stroop)	Children-Adults	n/a	CSO (24) vs. NOC (24)	*AUC* = .56	0.21
			Highly deviant CSO (15) vs. NOC (24)	*AUC* = .59	0.32
Van Leeuwen et al. (2012; Picture-Word Stroop)	Children-Adults	n/a	SCP (20) vs. NOC (20)	*AUC* = .84[snr]	1.41
Rapid Serial Visual Presentation (RSVP)					
Beech et al. (2008)	T_1 Child-T_1 Animal	n/a	Intrafamilial CSO (16) vs. NSOC (17)	*r* = .45**	1.00
			Extrafamilial CSO (18) vs NSOC (17)	*r* = .54***	1.28
Crooks, Rostill-Brookes, Beech, & Bickley (2009)	T_1 Child-T_1 Animal	n/a	Adolescent CSO (20) vs. Adolescent NSOC (26)	n/a	-

Note. All comparisons with male participants and based on uncorrected, raw data. n/a = not available; CSO = Child sexual offenders; NSOC = Non-sexual offenders; NOC = Non-offender controls; AC = Non-child sexual offending controls; SO = Sexual offender s with adult and/or child victims; SCP = Self-identified community pedophiles; PPG = Penile plethysmography, VT = Viewing time. [a] No differences reported for all discriminant analyses in Abel et al. (1998). [b] All comparisons/effect sizes reported in Worling (2006) are based on the *Prepubescent/Postpubescent Deviance Index.* [snr] = significance level not reported. * *p* < .05; ** *p* < .01; *** *p* < .001