

Supplementary Materials

Procedure

All participants were assessed individually on two consecutive days, by first checking eligibility based on the inclusion/exclusion criteria, followed by obtaining demographic data, job status, psychological state, and medical history through semi-structured interviews and self-report questionnaires. Subsequently, a comprehensive neuropsychological assessment was conducted. On the second day of assessment, the neurophysiological and indirect test procedures, including a virtual reality (VR) exposure, were conducted. Following the VR exposure, the actual study aims were disclosed, complemented by the assessment of the used NRP picture set, questionnaires on experienced childhood trauma / abuse (Childhood Trauma Questionnaire [CTQ], Bernstein & Fink, 1998), and on psychosexual characteristics (Multiphasic Sex Inventory [MSI], Deegener, 1996). The order of tests was fixed (**Table S1**).

Indirect tests

IAT: The concept of maturation was represented by five Not-Real-People (NRP) pictures (Pacific Behavioral Assessment Corporation, 2004) of children and adults from either gender: FT1C-1, FT1C-2, FT1C-3, FT1C-4, FT1C-6, FT5C-3, FT5C-4, FT5C-5, FT5C-6, FT5C-7, MT1C-1, MT1C-2, MT1C-4, MT1C-6, MT1C-7, MT5C-1, MT5C-4, MT5C-6, MT5C-7, MT5C-8. The set of attributes was subdivided in five sexual ('erotisch' [erotic], 'erregend' [arousing], 'geil' [horny], 'sinnlich' [sensual], 'aufreizend' [slinky]) and five non-sexual (neutral) ones ('fade' [insipid], 'ausdruckslos' [blank], 'egal' [equal], 'lahm' [lame], 'langweilig' [boring]).

Semantic Misattribution Procedure (SMP): Subliminal and supraliminal prime trials were introduced to tackle differential modes of stimulus processing (Dehaene, Changeux,

Naccache, Sackur, & Sergent, 2006; Del Cul, Baillet, & Dehaene, 2007). Each trial had the following structure: A fixation cross was initially presented for 1,400 to 1,799 ms, followed by the prime shown for 7 ms. After the prime, the fixation cross and a random noise pattern as mask were presented for either 28 ms and 469 ms, respectively, (subliminal presentation) or for 72 ms and 425 ms (supraliminal presentation). Then, the Chinese ideograph appeared on the screen for 749 ms, followed by another mask. The NRP pictures FT1N-1, FT1N-3, FT3N-7, FT3N-8, FT5N-5, FT5N-7, MT1N-4, MT1N-6, MT3N-7, MT3N-8, MT5N-1, and MT5N-6 were used as primes (Pacific Behavioral Assessment Corporation, 2004). In addition, there were also trials with neutral primes (object pictures) and trials with no primes ('mask only trials'). There were 480 trials in the critical conditions (3 Tanner stages x 2 [male/female] x 2 [subliminal/supraliminal] x 40 trials), 48 trials with neutral primes, as well as 86 mask-only trials.

CRT task: Forty female and male NRP images of all 5 Tanner stages (T1 to T5), either in bathing suits or nude, were presented at the center of a screen. The following NRP pictures were used: FT1C-1, FT1C-2, FT1N-1, FT1N-2, FT2C-1, FT2C-2, FT2N-1, FT2N-2, FT3C-1, FT3C-2, FT3N-1, FT3N-2, FT4C-1, FT4C-2, FT4N-1, FT4N-2, FT5C-1, FT5C-2, FT5N-1, FT5N-2, MT1C-1, MT1C-2, MT1N-1, MT1N-2, MT2C-1, MT2C-2, MT2N-1, MT2N-2, MT3C-1, MT3C-2, MT3N-1, MT3N-2, MT4C-1, MT4C-2, MT4N-1, MT4N-2, MT5C-1, MT5C-2, MT5N-1, MT5N-2. Images were displayed until the participants pressed a response button. There was neither a response deadline nor response feedback. After a delay of 1,500 ms (blank screen), a new trial started. A total of five blocks included 40 trials each. Within blocks, images and target locations were presented in a balanced, pseudo-randomized order.

Results

Reliability of the indirect tests

Data on the reliability of indirect tests (IAT, VTs, CRT task) were recently published by Dombert et al. (2017) and Welsch, Schmidt, Turner, & Rettenberger (2020). The reliabilities of the indirect tests as used in the current study were assessed by calculating split-half reliability estimates, which were obtained by creating two equivalent subsets of trials and calculating correlations (Spearman's rho) of the derived measures between these two halves. Spearman-Brown correction was applied to estimate the reliability of the full length original test. The split-half reliability estimates were $\rho = .77$ for viewing times, $\rho = .92$ and $.98$ for the reaction times in the CRT task and IAT, as well as $\rho = .94$ and $.96$ for the response ratios in SMP of the subliminal and supraliminal condition respectively (all $p < 0.001$, $n = 62$). On the level of test outcome measures, the split-half reliability estimates were $\rho = .44$ ($p = .07$) for d_{VT} , $\rho = .21$ ($p = .33$) for d_{CRT} , $\rho = .95$ ($n = 62$, $p < .001$) for d_{IAT} , as well as $\rho = .61$ ($p < .001$) and $.70$ ($p < .001$) for d_{SMP} of the subliminal and supraliminal condition respectively. The relatively low split-half reliability estimates of d_{VT} and d_{CRT} are likely due to the small number of stimulus presentations, forming the empirical basis for calculating these estimates, as well as to the calculation of difference values.

Table S1*Overview of the conducted procedures**

Day 1	Day 2
<i>Clinical examination</i>	<i>Neurophysiology</i>
<ul style="list-style-type: none"> - Demographic data - Family anamnesis - Case history - Drugs / alcohol / smoking - Sexual orientation - Critical life events - Personality disorders - Psychopathy 	<ul style="list-style-type: none"> - Go/NoGo paradigm
	<i>Indirect Tests</i>
	<ul style="list-style-type: none"> - Implicit Association Test (IAT) - Semantic misattribution procedure (SMP) - Viewing Time (VT) - Choice reaction time (CRT) task
	<i>Virtual Reality (VR) exposure ('beach walk')</i>
	<i>Disclosure of the study aims</i>
<i>Neuropsychological testing</i>	<i>Explicit tests/questionnaires</i>
<ul style="list-style-type: none"> - Attention - Working memory - Episodic memory - Executive functions - Risk taking behavior - Intelligence - Motor behavior 	<ul style="list-style-type: none"> - Childhood Trauma Questionnaire (CTQ) - Multiphasic Sex Inventory (MSI) - Screening Scale for Pedophilic Interests (SSPI-2)

* The procedures are shown in their chronological order, except that the VR exposure took place between the SMP and VT.

Table S2*Summary of the neuropsychological tests administered*

Task/assessment	Tested domain	Extracted variables
Attention Network Test, ANT, (Fan, McCandliss, Fossella, Flombaum, & Posner, 2005)	Attention	Total errors, alertness, orientation, conflict
California Verbal Learning Test, CVLT (Niemann, Sturm, Thöne-Otto, & Willmes-von-Hinkeldey, 2008)	Episodic memory	Learning curve intercept and slope, interference, short delay (SD) and long delay (LD) free recall, SD and LD cued recall, SD and LD savings, LD recognition hits and hit rate, false alarms and false alarms rate
Trail Making Test, TMT, A/B (Reitan, 1958)	Executive functions (cognitive flexibility)	TMT-A-time: processing speed, TMT-B-time: cognitive flexibility
Wisconsin Card Sorting Test, WCST (Heaton, Chelune, Curtiss, Kay, & Talley, 1993)	Executive functions (cognitive flexibility)	Perseveration errors, concept failures, number of correct concepts, average time per move
Go/Nogo, (Rosvold, Mirsky, Sarason, Bransome Jr, & Beck, 1956)	Executive functions (motor inhibition)	False alarm (FA) rate, reaction time [RT] to hits and FA
Reversal Learning, RL (Rolls, 1999)	Executive functions (learning)	RL capacity
Tower of Hanoi, ToH (Kotovskiy, Hayes, & Simon, 1985)	Executive functions (problem solving)	number of moves, completion time
Stroop Task (Stroop, 1935)	Executive functions (response conflict)	Interference
Wechsler Adult Intelligence Scale, WAIS, IV, Similarities, (Wechsler, 2008)	Executive functions (verbal abstraction)	Score
Verbal Fluency (Lezak, 1995)	Executive functions (verbal production)	Sum score animals and “S”- words
Leistungsprüfsystem LPS (Horn, 1983)	Intelligence (abstract reasoning)	Scale 3, sum score
Multiple Choice Vocabulary Test MWT-B (Lehrl, 1977)	Intelligence (verbal)	Score
Motor Tapping Test, MTT (Reitan & Wolfson, 1985)	Motor behavior	Taps per second
Cambridge Gambling Task, CGT (Rogers et al., 1999)	Risk taking	Total score, impulsivity, risk height
Iowa Gambling Task (Bechara, Damasio, Damasio, & Anderson, 1994)	Risk taking	Total gain
n-back task (Kirchner, 1958)	Working memory	Mean score, standard deviation, omission errors, commission errors
Corsi Block-Tapping Test CBTT	Working memory (spatial)	Span forward, n of correct forward, Span backward, n of correct backward

(Corsi, 1972)

Wechsler Memory Scale , WMS, IV (Wechsler, 2009)	Working memory (verbal)	Digit span forward (DSF) and backward (DSB)
Edinburgh Handedness Inventory, EHS, (Oldfield, 1971)	Handedness	Laterality quotient of the weighted sum of preferred handedness in 10 unimanual tasks

Table S3*Logistic regression models*

CTL vs CSO	R²	d_{VT}	d_{IAT}	d_{CRT}	Orienting	WM errors	IQ_C
Indirect tests	0.36	-0.322	-0.315	-0.35			
Neuropsychology	0.20				0.24	0.28	0.37
Combined	0.47	-0.308	-0.307	-0.32	0.27	0.23	
ncCSO vs cCSO	R²	d_{CRT}	Risk taking				
Indirect tests	0.03	-0.17					
Neuropsychology	0.12		0.35				

Note: R² and standardized regression coefficients for all described logistic regression models. (Pseudo-) R² was calculated using the method of McFadden (1974), and the standardized coefficients were derived following the variance-based approach of Menard (1995).

References

- Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50(1), 7–15.
- Bernstein, D. P., & Fink, L. (1998). Childhood trauma questionnaire: a retrospective self-report. Orlando: Psychological Corporation.
- Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. Montreal: McGill University.
- Deegener, G. (1996). *Multiphasic sex inventory (MSI): Fragebogen zur Erfassung psychosexueller Merkmale bei Sexualtättern: Handbuch*. Hogrefe, Verlag für Psychologie.
- Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences*, 10(5), 204–211.
- Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol*, 5(10), e260. <https://doi.org/10.1371/journal.pbio.0050260>.
- Dombert, B., Antfolk, J., Kallvik, L., Zappalà, A., Osterheider, M., Mokros, A., & Santtila, P. (2017). Identifying pedophilic interest in sex offenders against children with the indirect choice reaction time task. *European Journal of Psychological Assessment* 33, 345-351.
- Fan, J., McCandliss, B. D., Fossella, J., Flombaum, J. I., & Posner, M. I. (2005). The activation of attentional networks. *Neuroimage*, 26(2), 471–479.
- Heaton, R. K., Chelune, G. J., Curtiss, G., Kay, G. G., & Talley, J. L. (1993). Wisconsin card sorting test. Odessa, FL: Psychological Assessment Resources.
- Horn, W. (1983). Leistungsprüfsystem LPS (2nd edition). Göttingen: Hogrefe.
- Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, 55(4), 352-358.
- Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, 17(2), 248–294.
- Lehrl, S. (1977). Mehrfachwahl-Wortschatz-Test (MWT-B). Erlangen: Straube.
- Lezak, M. (1995). Verbal fluency. New York: Oxford University Press.
- McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, 3(4), 303–328. [https://doi.org/10.1016/0047-2727\(74\)90003-6](https://doi.org/10.1016/0047-2727(74)90003-6)
- Menard, S. W. (1995). Applied logistic regression analysis. ThousandOaks, CA: Sage.

- Niemann, H., Sturm, W., Thöne-Otto, A., & Willmes-von-Hinkeldey, K. (2008). *California Verbal Learning Test. German adaptation*. Frankfurt: Pearson Assessment.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113.
- Pacific Behavioral Assessment Corporation. (2004). *The NRP (Not Real People) Visual Stimulus Set*.
- Reitan, R.M. (1958) Validity of the trail making test as an indicator of organic brain damage. *Perceptual and Motor Skills*, 8, 271-276.
- Reitan, R. M., & Wolfson, D. (1985). The Halstead-Reitan neuropsychological test battery. Tucson, AZ: Neuropsychology Press.
- Rogers, R. D., Everitt, B. J., Baldacchino, A., Blackshaw, A. J., Swainson, R., Wynne, K., ... others. (1999). Dissociable deficits in the decision-making cognition of chronic amphetamine abusers, opiate abusers, patients with focal damage to prefrontal cortex, and tryptophan-depleted normal volunteers: evidence for monoaminergic mechanisms. *Neuropsychopharmacology*, 20(4), 322–339.
- Rolls, E. T. (1999). The functions of the orbitofrontal cortex. *Neurocase*, 5(4), 301–312.
- Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome Jr, E. D., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology*, 20(5), 343-350.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643-662.
- Wechsler, D. (2008). Wechsler Adult Intelligence Scale-IV (WAIS-IV). Administration and Scoring Manual. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2009). Wechsler Memory Scale IV (WMS-IV). San Antonio, TX: The Psychological Corporation.
- Welsch, R., Schmidt, A. F., Turner, D., & Rettenberger, M. (2020). Test-retest reliability and temporal agreement of direct and indirect sexual interest measures. *Sexual Abuse*. Advance online publication. doi:10.1177/1079063220904354