Birds of a feather flock together: Evidence of prominent correlations within but not between self-report, behavioral, and electrophysiological measures of impulsivity

Indy Bernoster[a,b,1], Kristel De Groot[a,b,d,1,*], Matthias J. Wieser[d], Roy Thurik[a,b,c], Ingmar H.A. Franken[b,d]

[a] Erasmus School of Economics, Erasmus University Rotterdam, the Netherlands
[b] Erasmus University Rotterdam Institute for Behavior and Biology (EURIBEB), the Netherlands
[c] Montpellier Business School, France
[d] Department of Psychology, Erasmus University Rotterdam, the Netherlands

ABSTRACT

Despite many studies examining a combination of self-report, behavioral, and neurophysiological measures, only few address whether these different levels of measurement indeed reflect one construct. The present study aids in filling this gap by exploring the association between self-report, behavioral, and electrophysiological measures of impulsivity and related constructs such as sensation seeking, reward responsiveness, and ADHD symptoms. Individuals across two large samples (n = 133 and n = 142) completed questionnaires and performed behavioral tasks (the Eriksen Flanker task, the Go/No-Go task, the Reward task, and the Balloon Analogue Risk Task) during which brain activity was measured using electroencephalography (EEG). The resulting data showed that even though the correlations within each level of measurement were prominent, there was no evidence of significant correlations across the three measurement levels. These findings contradict the outcomes of some previous, smaller studies, which did report significant associations between self-reported impulsivity(-related) measures and behavior and/or electrophysiology. Therefore, we suggest using sufficiently large samples when investigating associations between different levels of measurement.

1. Introduction

Impulsivity is defined as "a predisposition toward rapid, unplanned reactions to internal or external stimuli without regard to the negative consequences of these reactions to the impulsive individual or to others" (Moeller, Barratt, Dougherty, Schmitz, & Swann, 2001, p. 1784Moeller et al., 2001Moeller, Barratt, Dougherty, Schmitz, & Swann, 2001, p. 1784). It is a normal aspect of behavior which is often functional, but can also be dysfunctional. Impulsivity is a multi-dimensional construct (Gerbring, Ahadi, & Patton, 1987; Khadka et al., 2017; Meda et al., 2009), and is closely related to other constructs such as the Behavioral Activation System (BAS; Carver & White, 1994). BAS is in turn associated with reward responsiveness (Carver & White, 1994), which consists of reward sensitivity and rash impulsiveness (Dawe, Gullo, & Loxton, 2004), of which particularly the latter is closely associated with impulsivity (Franken & Muris, 2006). Impulsivity is

also closely related to sensation seeking (Whiteside & Lynam, 2001; Zuckerman & Neeb, 1979), "the seeking of varied, novel, complex, and intense sensations and experiences, and the willingness to take physical, social, legal, and financial risks for sake of such experience" (Zuckerman, 1994, p. 27). Both sensation seeking and impulsivity are related to risk taking (Jones & Lejuez, 2005; Lejuez et al., 2002; Romer, 2010; Steinberg, 2008), although they may differ in timing and neural underpinnings (Steinberg et al., 2008). Furthermore, impulsivity is a hallmark symptom of several mental disorders (Chamberlain, Stochl, Redden, & Grant, 2018): various facets of impulsivity such as urgency and a lack of premediation and perseverance characterize for example Attention Deficit/Hyperactivity Disorder (ADHD; Lopez, Dauvilliers, Jaussent, Billieux, & Bayard, 2015).

Impulsivity and related constructs such as sensation seeking, reward responsiveness, and ADHD symptoms can be investigated using self-report measures (e.g. Lopez et al., 2015), behavioral measures (e.g.

Sharma, Markon, & Clark, 2014), and neurophysiological measures such as electroencephalography (EEG; e.g. Taylor, Visser, Fueggle, Bellgrove, & Fox, 2018). However, only few studies address whether these different levels of measurement (i.e. self-report, behavior, and neurophysiology) indeed reflect one construct. Research in other areas has already demonstrated that this is not necessarily the case by showing that single constructs measured on different levels are only weakly connected. For instance, Dittmar, Krehl, and Lautenbacher (2011) investigated the association between these three levels of measurement for pain-related information processing. After correcting for multiple testing, they found no significant associations between the electrophysiological measures (recorded during processing pain-related words) and the behavioral measures (acquired from the dot-probe task), nor between electrophysiology and self-reports (obtained from the Pain Catastrophizing Scale, Pain Anxiety Symptoms Scale, and Pain Hypervigilance and Awareness Questionnaire). With respect to behavior and self-report, only one (out of nine) associations was significant. Another study examining different levels of measurement focused on anxiety and depression, in specific defensive reactivity and cognitive control in young children (Moser, Durbin, Patrick, & Schmidt, 2015). Self-report measures consisted of two parental reports: the Child Behavior Questionnaire and Child Behavior Checklist. Further, children performed 15 behavioral tasks designed to probe defensive reactivity and cognitive control. Neurophysiological measures included the Fear-Potentiated Startle, resting-state EEG asymmetry, and EEG Event-Related Potentials (ERPs). The findings showed that only 2 out of the 11 correlations between different measurement levels were significant: the combined behavioral score correlated with the ERP, and one of the questionnaire scores correlated with EEG asymmetry. None of the questionnaire scores was significantly related to the behavioral measures. These findings again indicate that single constructs measured on different levels are only weakly related.

The present study contributes to this small body of literature on the associations between different measurement levels by providing a comprehensive overview of the associations between self-reports, behavior, and electrophysiology in the broad domain of impulsivity. Subsets of these associations have already been examined by previous studies. For example, self-reported impulsivity has been related to several behavioral outcomes, such as decreased behavioral inhibition in a Go/No-Go task (Littel et al., 2012), increased uncertain decision-making in the Balloon Analogue Risk Task (BART; Lauriola, Panno, Levin, & Lejuez, 2014; Lejuez et al., 2002), and slower stopping reaction times in a stop-signal task (Logan, Schachar, & Tannock, 1997). Self-reports have also been related to electrophysiology: individuals who score high on impulsivity were shown to have reduced error-related negativity (ERN) amplitudes in response to incorrect trials on the Go/No-Go task (Littel et al., 2012), and on punishment (Potts, George, Martin, & Barratt, 2006; Potts, Martin, Burton, & Montague, 2006) or incorrect (Luijten, Van Meel, & Franken, 2011) trials of the Eriksen Flanker task, all implying poor error processing. Results concerning other ERPs are more equivocal. For example, some studies related increased impulsivity to smaller P3 amplitudes on the stop-signal task (Shen, Lee, & Chen, 2014), the continuous performance task (Kam, Dominelli, & Carlson, 2012), and a gambling task (Gao et al., 2016), whereas others reported larger stop P3s for high-impulsive individuals using again the stop-signal task (Lansbergen, Böcker, Bekker, & Kenemans, 2007). In a similar fashion, some report a clear relationship between high impulsivity and decreased N2 amplitudes (Gao et al., 2016), whereas others find no significant association (Zhou, Yuan, Yao, Li, & Cheng, 2010) or find that the direction of the association depends on the specific impulsivity domain being examined (Kam et al., 2012).

In addition to studies examining impulsivity, some studies employed self-report measures of related constructs, such as sensation seeking, reward responsiveness, and ADHD. For example, self-reported sensation seeking has been associated with increased uncertain decisions in the BART (Lauriola et al., 2014; Lejuez, Aklin, Zvolensky, &

Pedulla, 2003), reduced ERN amplitudes in the Eriksen Flanker task (Zheng, Sheng, Xu, & Zhang, 2014), and reduced P3 amplitudes in a passive oddball paradigm (Wang & Wang, 2001). Furthermore, self-reported reward responsiveness has been related to shorter reaction times on the Go/No-Go task (De Pascalis, Varriale, & D'Antuono, 2010) and to P3 amplitudes, although literature is inconsistent as to whether this latter relationship is negative (De Pascalis et al., 2010) or positive (Van den Berg, Franken, & Muris, 2011). Finally, ADHD symptoms have been related to more errors on response inhibition tasks such as the Eriksen Flanker and Go/No-Go task (Geburek, Rist, Gediga, Stroux, & Pedersen, 2013; Jonkman, Van Melis, Kemner, & Markus, 2007; Van Meel, Heslenfeld, Oosterlaan, & Sergeant, 2007; Wiersema, Van der Meere, & Roeyers, 2005), and to attenuated P3 (Liotti, Pliszka, Perez, Kothmann, & Woldorff, 2005; Shen et al., 2014; Sutcubasi et al., 2018), Pe (Groen et al., 2008; Herrmann et al., 2009; Jonkman et al., 2007; Wiersema et al., 2005; Wiersema, Van der Meere, & Roeyers, 2009), and ERN (Geburek et al., 2013; Groen et al., 2008; Liotti et al., 2005) amplitudes, although several studies were unable to confirm this latter relationship (Herrmann et al., 2009; Jonkman et al., 2007; Wiersema et al., 2005; Wiersema et al., 2009).

Although these studies have revealed important insights and are excellent starting points for further inquiries, they have some limitations with regard to (1) the consistency of the findings, (2) the number of investigated measurement levels and constructs, and (3) sample size. The present study is a first attempt to overcome these limitations. *First*, the present study adds value to the current body of literature by extending the knowledge on the role of behavioral and electrophysiological measures of impulsivity. The studies described above provide much insight but are far from conclusive. Examples of such inconsistent findings have already been discussed, such as whether the P3 amplitude is larger or smaller in relation to impulsivity and reward responsiveness, and whether or not the N2 and ERN are impacted by respectively impulsivity and ADHD. These and other inconsistencies throughout the impulsivity literature confirm that the field has not (yet) reached consensus, especially when it comes to associations between measures originating from multiple levels.

Second, the present study deals with *multiple* constructs (i.e. impulsivity, sensation seeking, reward responsiveness, and ADHD) and *multiple* levels of measurement (i.e. self-report, behavior, and electrophysiology). Most studies investigate the association between a *single* self-reported construct and *either* behavioral *or* electrophysiological measures (e.g. Logan et al., 1997; Potts, George et al., 2006). This fits with the primary aim of these studies, but makes that they do not fully take into account the complexity of associations between multiple constructs and multiple levels of measurement. A small number of studies examines *multiple* constructs on *multiple* levels, but limit their examination to two measurement levels (Meda et al., 2009; Reynolds, Ortengren, Richards, & De Wit, 2006).

Third, we use *two* relatively *large* samples. Most papers cited in the present study that involve electrophysiology use relatively small samples consisting of 20–40 participants. This is consistent with the broader field of EEG research: the average size of the 81 samples discussed in a recent systematic review on ERPs in relation to risk-taking (Chandrakumar, Feuerriegel, Bode, Grech, & Keage, 2018) was a mere 29.01 (SD = 18.54). The key problem regarding small samples is that they lead to low statistical power and thus have a lower chance that discovered effects are genuinely true (Button et al., 2013; Forstmeier, Wagenmakers, & Parker, 2017; Ioannidis, 2005). Moser et al. (2015) also recommend the use of larger samples, specifically in EEG research, to establish reliability. Therefore, the present study explores two large non-clinical samples, with sample sizes of 133 and 142 participants.

In sum, the present study aims to investigate the associations between self-report measures, behavioral measures, and electrophysiological measures for impulsivity and related constructs. As discussed, the associations between these three different levels of measurement have already been examined for pain-related

information-processing (Dittmar et al., 2011) and defensive reactivity and cognitive control (Moser et al., 2015). For impulsivity, however, no such large-scale study exists, despite the construct being central to the field of (neuro)psychology. Not only does impulsivity impact daily life (ranging from recreational activities to education and employment), aberrant displays of it are present in several major diseases, such as dementia (Arvanitakis, 2010), Huntington's chorea (Kalkhoven, Sennef, Peeters, & Van Den Bos, 2014), and Parkinson's disease (Chaudhuri, Odin, Antonini, & Martinez-Martin, 2011), as well as in addiction and pathological gambling (Limbrick-Oldfield, Van Holst, & Clark, 2013). Furthermore, impulsivity is a rather well-suited construct for examining the associations between self-reports, behavior, and electrophysiology for the simple reason that many well-validated measures for the construct exist on all three levels.

For the present study, the following such measures were selected: self-report measures included the ImpSS-8 scale (Webster & Crysel, 2012), the Brief Sensation Seeking Scale (BSSS; Hoyle, Stephenson, Palmgreen, Lorch, & Donohew, 2002), the Reward Responsiveness (RR) scale (Van den Berg, Franken, & Muris, 2010), and the ADHD Self-Report Scale (ASRS-6; Kessler et al., 2005). To obtain behavioral and electrophysiological measures, participants performed the Eriksen Flanker task (Eriksen & Eriksen, 1974; Marhe, Van de Wetering, & Franken, 2013), the Go/No-Go task (Donders, 1969; Littel et al., 2012), the Reward task (Franken, Van den Berg, & Van Strien, 2010; Potts, George et al., 2006; Potts, Martin et al., 2006), and the BART (Euser, Van Meel, Snelleman, & Franken, 2011; Lejuez et al., 2002; Pleskac, Wallsten, Wang, & Lejuez, 2008). These measures all reflect constructs often associated with impulsivity, and indeed roughly match those used in previous studies examining associations between different levels of measurement. Since measures can focus on several different aspects of impulsivity, a broad range was included: measures originating from different contexts, such as non-clinical (reward responsiveness) vs. clinical (ADHD symptoms); measures using different kinds of feedback stimuli, such as financial tokens (Reward task, BART) vs. abstract ones (Eriksen Flanker, Go/No-Go); and measures tapping different domains (Bechara, Damasio, & Damasio, 2000), such as the motor domain (Eriksen Flanker, Go/No-Go) vs. cognition (BART).

In the present study these particular measures as well as the impulsivity construct in general are subservient to the overarching aim of examining the associations between different levels of measurement, namely self-reports, behavioral measures, and electrophysiology. Our focus is therefore not on any individual association but on the overall pattern of associations present in the data. However, since most previous studies do focus on individual associations, our hypotheses are based on these findings, which mostly show significant relationships. Taking into account the fact that our impulsivity-related constructs do not fully overlap, we expect our self-report measures, behavioral measures, and electrophysiological measures to show only small (but significant) correlations.

2. Data and method

The present section describes the two samples (Sample 1 and Sample 2) and the methods used to analyze these samples. The available data and the exact methods used differ between the two samples because both were collected and processed by different researchers at different times. These differences in fact support the ecological validity of the present study by showing that the found results do not depend on the idiosyncrasies of data collection and processing.

2.1. Sample 1

2.1.1. Participants

The first sample consists of third- and fourth-year university students (N = 169) and was collected between September 2013 and May

2014. Incomplete observations were excluded[2] resulting in a final sample of $n = 133$ (average age of 22.23 ($SD = 2.46$) and 39 percent women).

2.1.2. Procedure

At least two days before the lab session, participants received an email asking them to not drink coffee or smoke cigarettes in the 90 min before the lab session to prevent acute caffeine/nicotine effects. This email also contained a link to the web-based questionnaire including the self-report measures. Further, it was communicated that the six best-performing (highest accuracy in both lab tasks) participants would receive a financial reward of 100 euros.

Upon arrival in the lab, the participant was informed about the procedure and provided written informed consent. Then, the participant was seated in a comfortable chair in a light- and sound-attenuated EEG room. Participants were wired to the EEG and performed two behavioral tasks, a Go/No-Go task (Donders, 1969; Littel et al., 2012) and an Eriksen Flanker task (Eriksen & Eriksen, 1974; Marhe et al., 2013), during which EEG was recorded. The total lab session lasted approximately two hours. All tasks were programmed using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). Session design was approved by the local institutional review board. Part of the data is reported in a previous study (Rietdijk, Franken, & Thurik, 2014) that addresses the internal consistency of the electrophysiological measures.

2.1.3. Measures
2.1.3.1. Self-report measures. The online questionnaire included self-report measures on *Impulsivity*, *Sensation Seeking*, and *ADHD symptoms*, as well as two control variables: age and gender (1 = female). *Impulsivity* and *Sensation Seeking* were measured using the ImpSS-8 scale (Webster & Crysel, 2012), which incorporates the best items from the larger ImpSS-19 scale (Zuckerman, Kuhlman, Joireman, Teta, & Kraft, 1993). *Impulsivity* was measured by four items ("I usually think about what I am doing before doing it" (reverse-scored), "I often do things on impulse", "I very seldom spend much time on the details of planning ahead", "I often get so carried away by new and exciting things and ideas that I never think of possible complications"), and *Sensation Seeking* by another four ("I enjoy getting into new situations where you cannot predict how things will turn out", "I like doing things just for the thrill of it", "I sometimes do 'crazy' things just of fun", "I like to explore a strange city or section of town by myself, even if it means getting lost"). Items were rated on a 7-point scale ranging from completely disagree to completely agree. Cronbach's alpha was .50 for *Impulsivity* and .71 for *Sensation Seeking*.

ADHD symptoms were measured using the ASRS-6 (Kessler et al., 2005), which includes the following items: "How often do you have trouble wrapping up the fine details of a project, once the challenging parts have been done?", "How often do you have difficulty getting things in order when you have to do a task that requires organization?", "When you have a task that requires a lot of thought, how often do you avoid or delay getting started?", "How often do you have problems remembering appointments or obligations?", "How often do you fidget or squirm with your hands or your feet when you have to sit down for a long time?", and "How often do you feel overly active and compelled to do things, like you were driven by a motor?". Response options

---

[2] None of the participants reported head surgeries, pregnancy, or any history of psychiatric illness (these exclusion criteria were checked the day before data recording). Nine participants were excluded because of errors during data recording, and one participant was excluded for reporting an age of 0. A number of 12 participants were removed due to too many artefacts (e.g. movement, noise) or too few (< 20) correct No-Go trials on the Go/No-Go task. A number of 16 participants were removed due to too many artefacts (e.g. movement, noise) or too few (< 5) error trials on the Eriksen Flanker task. Two participants fit two exclusion criteria, resulting in a total sample of 133 (169 – 9 – 1 – 12 – 16 + 2).

included "never", "rarely", "sometimes", "often", and "very often". Cronbach's alpha equaled .52.

2.1.3.2. Behavioral measures. Participants completed two behavioral tasks: the Go/No-Go task and the Eriksen Flanker task. The Go/No-Go task (Donders, 1969; Littel et al., 2012) consisted of 500 trials (of which 125 were No-Go trials), including 30 practice trials. In each trial, a vowel (A, E, I, O, or U) was shown. When the vowel differed from the previously shown vowel, participants had to indicate a 'Go' by pressing a button with their right index finger as fast as possible. In case of the vowel being equal, participants had to indicate a 'No-Go' by withholding a response. Vowels were visible for 200 ms, and between consecutive vowels the screen was empty for a randomly varying duration between 1020 and 1220 ms. Vowels were presented in white on a black background. Four behavioral measures were obtained from the Go/No-Go task: (1) the number of incorrect No-Go trials (*GNG Number Incorrect No Go*), indicating impulsive pressing; (2) the number of incorrect Go trials (*GNG Number Incorrect Go*), which can be used as a benchmark measure; (3) the number of times individuals had two incorrect trials in a row (post-incorrect incorrect trials; *GNG Number Post-Incorrect Incorrect*), which is an indicator of extreme impulsiveness; and (4) the average response time on the correct Go trials and incorrect No-Go trials (*GNG Average Response Time*), for which lower response times indicate impulsivity (note that response times for incorrect Go trials and correct No-Go trials do not exist since by definition participants do not press in these instances).

The Eriksen Flanker Task (Eriksen & Eriksen, 1974; Marhe et al., 2013) consisted of 400 trials, including eight practice trials. In each trial, participants saw one out of four letter strings ('SSSSS', 'SSHSS', 'HHSHH', or 'HHHHH'). Letter strings appeared 100 times each in a completely random order. Participants were instructed to press a pre-defined button with their right index finger if the central letter was an 'H' and another button with their left index finger if the central letter was an 'S'. Half of the trials were congruent (i.e. 'SSSSS' or 'HHHHH') and the other half were incongruent (i.e. 'SSHSS' or 'HHSHH'). Trials started with a 150 ms cue ('^') pointing at the location of the central letter in the letter string. Then, the string appeared for 52 ms followed by a black screen for 648 ms, so that the total response time was 700 ms. Finally, a feedback symbol appeared for 500 ms indicating whether a response was correct ('ooo'), incorrect ('xxx'), or too late ('!'). Between trials there was a 100 ms break. Three behavioral measures were obtained from the Eriksen Flanker task: (1) the number of incorrect trials (*EF Number Incorrect*), indicating quick and imprecise responding; (2) the average response time for incongruent trials (*EF Average Response Time Incongruent*), which might indicate impulsivity as these trials require participants to 'take a step back' before responding; and (3) the difference between the average response time after incorrect trials and the average response time after correct trials (*EF Difference Average Response Time Post-Incorrect - Post-Correct*).

2.1.3.3. Electrophysiological measures. EEG was recorded during both the Go/No-Go task and Eriksen Flanker task using a Biosemi Active-Two amplifier system (Biosemi, Amsterdam, the Netherlands). A number of 32 active Ag/AgCl electrodes mounted in an elastic cap were placed on the scalp according to the 10–20 International System, with two extra electrodes at FCz and CPz. Additional electrodes were attached to the left and right mastoids (for referencing), the outer canthi of both eyes (for recording a horizontal electrooculogram), and the infraorbital and supraorbital region of the left eye (for recording a vertical electrooculogram). Signals were digitalized with a sample rate of 512 Hz and a 24-bit A/D conversion with a band pass of 0–134 Hz.

The recorded raw EEG signals were transformed offline using Brain Vision Analyzer 2.0 (Brain Products, Munich, Germany). Data were re-referenced to the computed mastoids. In addition, all signals were filtered with a band pass of 0.10–30 Hz (phase shift free Butterworth filters; 24 dB/octave slope). Ocular corrections were performed using

the Gratton, Coles, and Donchin (1983) algorithm. Topographical interpolation (Soong, Lind, Shaw, & Koles, 1993) was employed to calculate new values for bad channels, with a maximum of three channels per participant (data were excluded if more than three bad channels had to be interpolated). The data from the Go/No-Go task were segmented into epochs of 1000 ms (200 ms before to 800 ms after stimulus presentation); data from the Eriksen Flanker task were segmented into epochs of 700 ms (100 ms before to 600 ms after the response). The prestimulus period (respectively 200 ms and 100 ms) served as a baseline. Epochs including a signal that exceeded ± 100 μV were excluded. Ultimately, the average number of artefact-free segments on the Go/No-Go task was 70.95 for No-Go and 298.16 for Go trials. The average number of artefact-free segments on the Eriksen Flanker task was 22.17 for incorrect and 315.92 for correct trials.

The electrophysiological measures of interest in the Go/No-Go task are the N2 (representing mismatch detection) and the P3 (representing more elaborate appraisal of the stimuli). We opted for analyzing difference waves, which has the advantage of eradicating exogenous components, i.e. elements that are elicited in response to all stimuli and hence across all conditions (Miltner, Braun, & Coles, 1997). Furthermore, difference waves correct for individual differences in general wave amplitude, which is particularly useful for correlational studies since absolute waves may reflect a general tendency for smaller or larger amplitudes (instead of the underlying construct such as impulsivity). The N2 difference wave for the Go/No-Go task (*GNG N2*) was defined as the difference between the mean amplitude on No-Go trials vs. Go trials within the 175–250 ms time interval, averaged across midline electrodes (Fz, FCz, Cz, CPz, Pz) given that we were not interested in laterality effects. The P3 difference wave for the Go/No-Go task (*GNG P3*) was defined as the difference between the mean amplitude on No-Go trials vs. Go trials within the 300–500 ms time interval, again averaged across midline electrodes.

The electrophysiological measures of interest in the Eriksen Flanker task are the ERN (representing early error processing) and the Pe (representing conscious error processing). Again, the analyses focused on difference scores and used the averaged activity across the midline electrodes. The ERN difference wave for the Eriksen Flanker task (*EF ERN*) was defined as the difference between the mean amplitude on incorrect vs. correct trials within the 25–75 ms time interval. The Pe difference wave for the Eriksen Flanker task (*EF Pe*) was defined as the difference between the mean amplitude on incorrect vs. correct trials within the 200–400 ms time interval.

For both tasks the selection of the ERPs and the time windows chosen for calculating the average amplitudes were similar to those examined in previous studies (Littel et al., 2012; Marhe et al., 2013; Rietdijk et al., 2014), and were compatible with visual inspection of the present grand averaged waveforms (see Figs. 1 and 2).

2.2. Sample 2

2.2.1. Sample

The second sample again consists of university students (*N* = 181) and was collected between May 2015 and April 2016. Incomplete observations were excluded[3] resulting in a final sample of *n* = 142 (average age of 20.63 (*SD* = 2.04) and 54 percent women).

2.2.2. Procedure

After signing up for the study, participants received an e-mail asking

---

[3] Incomplete observations included 16 no-shows for the lab session, 6 participants with incorrect electrophysiological measurements on only the BART, 10 participants with incorrect electrophysiological measurements on only the Reward task, and 7 participants who had incorrect electrophysiological measurements on both the BART and the Reward task. Here, incorrect refers to not having enough trials to obtain a reliable electrophysiological measurement. These exclusions resulted in a final sample of 142 (181 – 16 – 6 – 10 – 7).
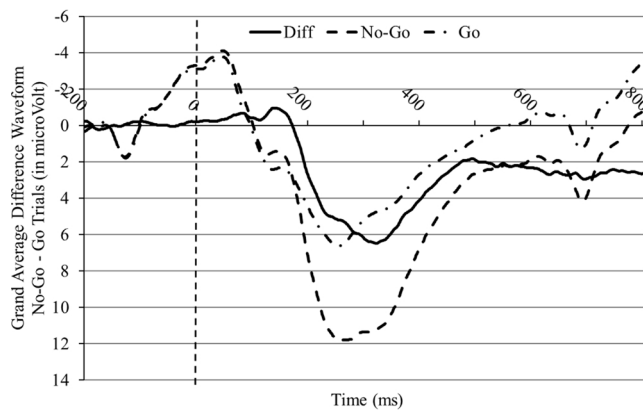
Fig. 1. Grand averaged difference and absolute waveforms for the Go/No-Go task, averaged over all midline electrodes. The differece waveform is similar to Rietdijk et al. (2014), where it was used for a different purpose.

Fig. 2. Grand averaged difference and absolute waveforms for the Eriksen Flanker task, averaged over all midline electrodes. The differece waveform is similar to Rietdijk et al. (2014), where it was used for a different purpose.
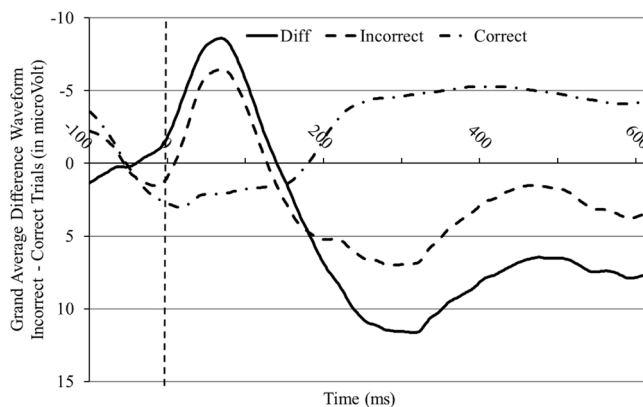
them to not drink coffee and/or energy drinks on the day of the experiment. The email also contained a link to the web-based questionnaire including the self-report measures, and explained the procedure and the reward system: participants received a show-up fee of five euros[4] and could earn an additional 7.50 euros by performing well on the tasks. One day before the lab session, participants received a reminder e-mail with a summary of the most important information.

Upon arrival in the lab, the participant was informed about the procedure and provided written informed consent. Then, the participant was seated in a comfortable chair in a light- and sound-attenuated EEG room. Participants were wired to the EEG and performed two behavioral tasks, a Reward task (Franken et al., 2010; Potts, Martin et al., 2006) and an automatic BART (Euser et al., 2011; Lejuez et al., 2002; Pleskac et al., 2008), during which EEG was recorded. The total lab session lasted approximately two hours. All tasks were programmed using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). Session design was approved by the local institutional review board.

2.2.3. Measures

2.2.3.1. Self-report measures. The online questionnaire included self-

report measures on *Sensation Seeking*, *Reward Responsiveness*, and *ADHD symptoms*, as well as two control variables: age and gender (1 = female). *Sensation Seeking* was measured using the Brief Sensation Seeking Scale (BSSS; Hoyle et al., 2002), which consists of eight items: "I would like to explore strange places", "I get restless when I spend too much time at home", "I like to do frightening things", "I like wild parties", "I would like to take off on a trip with no pre-planned routes or timetables", "I prefer friends who are excitingly unpredictable", "I would like to try bungee jumping", and "I would love to have new and exciting experiences, even if they are illegal". The items were rated on a 5-point scale ranging from strongly disagree to strongly agree. Cronbach's alpha was .78.

Reward Responsiveness was measured using the 8-item RR scale (Van den Berg et al., 2010). Four items of this scale are original: "I am someone who goes all-out", "If I discover something new I like, I usually continue doing it for a while", "I would do anything to achieve my goals", and "When I am successful at something, I continue doing it". The remaining four items are revised BAS scale (Carver & White, 1994) items: "When I go after something I use a 'no holds barred' approach", "When I see an opportunity of something I like, I get excited right away", "When I'm doing well at something, I love to keep at it", and "If I see a chance of something I want, I move on it right away". Items were rated on a 4-point scale. Response options included "strong disagreement", "mild disagreement", "mild agreement", and "strong agreement". Cronbach's alpha equaled .78.

ADHD symptoms were measured using the ASRS-6 (Kessler et al., 2005), which is explained in more detail in the description of Sample 1. For Sample 2, Cronbach's alpha was .50.

2.2.3.2. Behavioral measures. Participants completed two behavioral tasks: the passive Reward task and the automatic BART. The Reward task (Franken et al., 2010; Potts, Martin et al., 2006) consisted of 240 trials and eight additional practice trials. On each trial, participants were shown two consecutive stimuli that could be a picture of a lemon or a picture of a golden bar. Stimulus one predicted similarity of stimulus two in 80 percent of the trials. For example, if the first picture of a given trial was a lemon, there was an 80 percent chance that the second picture was a lemon as well and a 20 percent chance that the second picture was a golden bar. The second picture indicated a gain or a no gain. The task started with a white fixation cross ('+') on a black screen for 300 ms. Then, the first stimulus was shown for a period of 500 ms, after which the black screen with a fixation cross appeared again (300 ms) followed by the second stimulus (500 ms). A final black screen with a fixation mark (300 ms) was shown before the score screen (600 ms), which indicated a gain ('+1') or a no-gain ('+0'). For counter-balancing purposes, half of the participants were shown the golden bar as gain picture, whereas for the other half the lemon was indicative of a gain.[5] In case of a gain, the total number of points increased, which translated linearly to receiving more money. Since the Reward task is passive, no behavioral measures were obtained.

The automatic BART (Euser et al., 2011; Lejuez et al., 2002; Pleskac et al., 2008) consisted of 60 trials. On each trial, a picture of a balloon was shown. Participants had to inflate the balloon by selecting a number of pumps (between 1 and 128) and then clicking a predefined button labeled 'P' to start pumping. If the number of pumps was too high, the balloon could burst after pumping, which was indicated by a picture of a burst balloon accompanied by a red cross. In these cases, participants did not earn points. If the balloon did not burst, participants were shown a green dollar sign, and received points equal to the number of pumps. For each trial, the balloon had a predefined bursting point, determined by a random draw of 60 (trials) from an interval

---

[4] Psychology students received a start-up fee of two participant hours (i.e. hours contributing to the mandatory number of hours they need to fulfil as a research participant).

[5] It was examined whether condition influenced our results. Although average brain potentials differed between conditions, the findings for the correlations and associations remained similar.

distribution between 1 and 128. The bursting points were the same for each participant, but unknown to them. Hence, decisions were made under conditions of uncertainty (De Groot & Thurik, 2018). As for the Reward task, earned points were linearly translated to the amount of money participants received. Two behavioral measures were obtained from the BART: (1) the average number of pumps (*BART Average Pumps*), indicating a more uncertain choice; and (2) the average response time (*BART Average Response Time*), i.e. the time it took participants to choose a number between 1 and 128 and to press the 'P'.

2.2.3.3. Electrophysiological measures. EEG was recorded using the same settings as reported for Sample 1. The recorded raw EEG signals were transformed offline using Brain Vision Analyzer 2.1 (Brain Products, Munich, Germany). Data were re-referenced to the computed mastoids. In addition, all signals were filtered with a band pass of 0.10–30 Hz for the N2, P2, and P3 of the Reward task and for the P3 of the BART, and 2–12 Hz for the Feedback-Related Negativity (FRN) of the BART (phase shift free Butterworth filters; 24 dB/octave slope). Topographical interpolation (Soong et al., 1993) was employed to calculate new values for bad channels, with a maximum of three channels per participant (data were excluded if more than three bad channels had to be interpolated). Data were segmented into epochs of 1000 ms (200 ms before to 800 ms after stimulus presentation for the Reward task; and 200 ms before to 800 ms after feedback, i.e. the actual burst or gain, in the BART). Then, ocular corrections were performed using the Gratton et al. (1983) algorithm. The pre-stimulus period (200 ms for both tasks) served as a baseline. Epochs including a signal that exceeded ± 75 μV were excluded. Ultimately, the average number of artefact-free segments on the Reward task was 22.56 for unexpected gain and 22.43 for unexpected loss trials. The average number of artefact-free segments on the BART was, with regard to the FRN, 27.71 for loss and 32.15 for gain trails, and, with regard to the P3, 25.70 for loss and 29.41 for gain trials.

The electrophysiological measures of interest in the Reward task are the N2 (representing mismatch detection), the P2 (representing attention to (deviating) stimuli), and the P3 (representing elaborate stimulus appraisal). The analyses employed difference scores obtained from midline electrodes (justifications for these choices can be found in the description of Sample 1). The Reward task difference scores were defined as the difference between the mean amplitude on the unexpected gain trials vs. unexpected loss trials within the 200–300 ms time interval (for the N2; *REWARD N2*), the 150–230 ms time interval (for the P2; *REWARD P2*), and the 300–400 ms time interval (for the P3; *REWARD P3*).

The electrophysiological measures of interest in the BART are the FRN (representing error processing), and the P3 (representing elaborate stimulus appraisal). The BART difference scores were defined as the difference between the mean amplitude on the loss trials vs. gain trails within the 200–275 ms time interval (for the FRN; *BART FRN*) and within the 250–400 ms time interval (for the P3; *BART P3*).

As for Sample 1, the selection of the ERPs and the time windows chosen for calculating the average amplitudes were similar to those examined in previous studies (Euser et al., 2011; Salim, Van der Veen, Van Dongen, & Franken, 2015; Warren & Holroyd, 2012), and were compatible with visual inspection of the present grand averaged waveforms (see Figs. 3 and 4).

2.3. Analyses

First, we performed psychometric checks relevant to our planned analyses: (1) a check for common method bias to examine whether variance in the data could be attributed to the employed measurement method and thus alter correlations; and (2) a check on the variance
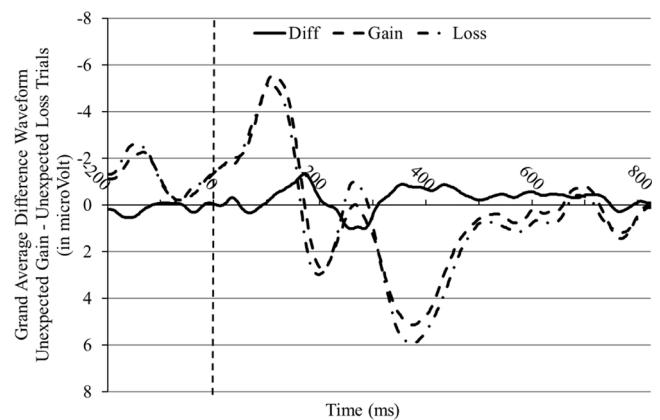

Fig. 3. Grand averaged difference and absolute waveforms for the Reward task, averaged over all midline electrodes.
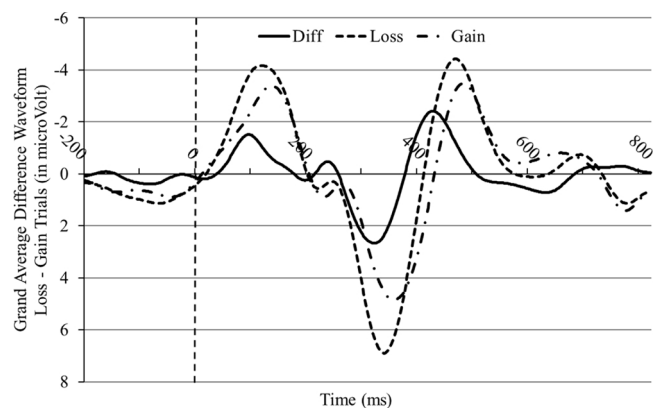

Fig. 4. Grand averaged difference and absolute waveforms for the BART, averaged over all midline electrodes.

inflation factors (VIFs), which indicate the level of multicollinearity[6], high correlations in independent variables which can lead to inaccurate estimates for the regression coefficients.

Second, we calculated the mean, standard deviation (SD), minimum (Min), maximum (Max), Cronbach's alpha, and correlations. Detailed analyses on the correlations then examined the number of correlations within each measurement level, and the number of correlations between measurement levels.

Third, we used linear regression models to further investigate whether behavioral and electrophysiological measures jointly contribute to the understanding of impulsivity(-related) constructs, given that the combined predictive value of these measures may be more salient compared to when they are related to self-reports individually. For each self-reported construct, we analyzed three multiple regression models: the first model only included behavioral predictors, the second only included electrophysiological predictors, and the third included both behavior and electrophysiology. The coefficients of the regression models were estimated using Ordinary Least Squares (OLS). To allow for comparison between the models, coefficients were standardized.

Finally, we used bootstrapping to obtain an overview of the number of significant correlations and associations we would have found if we had used smaller samples. By using large samples, the present study

---

[6] A VIF of, for instance, 4 indicates that in a regression including all variables of the analysis the standard error of the coefficient of this specific variable is two times (the square root of 4 is 2) as large as it would be if the variable was uncorrelated with the other variables. If the VIF is smaller than the suggested threshold of 10.00 (Diamantopoulos, Riefler, & Roth, 2008; Hair, Anderson, Babin, & Black, 2010), there is no indication of multicollinearity.

reduced the chance of identified effects being false. However, many studies investigating electrophysiology employ smaller samples of 20 to 40 participants. Therefore, we used the present data to bootstrap smaller samples (sized 20, 30, and 40) from our full sample (1000 iterations) to obtain the results we would have found if we had used a sample size more equal to that used in previous studies.

3. Results

3.1. Psychometric checks

Our data could be at risk of common method bias, which could lead to inflated or deflated correlations and hence to type I or II errors (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Therefore, we examined the possible common method bias using Harman's single factor test. The first principal component explained 11.94 percent of the variance in Sample 1, and 14.76 percent in Sample 2. Since this is below the threshold of 50.00 percent, the risk of common method bias in our data is small. The VIFs are reported in Tables 1 and 2 for respectively Samples 1 and 2. The highest VIF in Table 1 is 3.34 (for *GNG Number Post-Incorrect Incorrect*), and that in Table 2 is 4.55 (for *REWARD N2*). Hence, there is no indication of multicollinearity.

3.2. Correlation analyses

Tables 1 and 2 show the descriptive statistics for the variables in Sample 1 and Sample 2, respectively. For Sample 1, 100.00 percent of the correlations within the impulsivity(-related) self-report measures, 57.14 percent of the correlations within behavioral measures, and 50.00 percent of the correlations within the electrophysiological measures was significant. However, only 19.05 percent of correlations between behavioral and self-reported measures, 8.33 percent of correlations between electrophysiological and self-reported measures, and 17.86 percent of correlations between behavioral and electrophysiological measures reached significance.

With respect to the correlations of Sample 2, 66.67 percent of the correlations within impulsivity(-related) self-report measures, 100.00 percent of the correlations within behavioral measures, and 30.00 percent of the correlations within the electrophysiological measures was significant. However, none of the correlations between behavioral and self-reported measures, only 6.67 percent of correlations between electrophysiological and self-reported measures, and 10.00 percent of correlations between behavioral and electrophysiological measures reached significance.

3.3. Regression analyses

Tables 3 and 4 show the results of the OLS regressions investigating whether the joint behavioral measures, the joint electrophysiological measures, or all behavioral and electrophysiological measures combined contribute to the prediction of self-reported impulsivity(-related) constructs in respectively Sample 1 and Sample 2. For these regressions, relevant associations are those including behavioral and electrophysiological measures, i.e. excluding those with age and gender. For Sample 1, the models including only behavior (Models 1) and the models only including electrophysiology (Models 2) together have a total of 33 relevant associations. As we allow a five percent chance at a Type I error, we may expect 1.65 of the associations to be wrongly marked as 'significant'. Hence, the one significant association (between *GNG P3* and *Impulsivity*) that we find cannot be interpreted. Furthermore, the *F*-values for Models 3, in which both behavior and electrophysiology are included, are not significant. This means that all variables together do not significantly explain the variance in the self-reported constructs *Impulsivity*, *Sensation Seeking*, and *ADHD symptoms* better than just the intercept does.

For Sample 2, Models 1 and 2 have 21 relevant associations, meaning that we can expect 1.05 significant associations as a result of Type I error. In fact, none of the associations in our data is significant,

Table 1

Descriptive statistics (mean, standard deviation (SD), minimum (Min), maximum (Max), variance inflation factor (VIF), Cronbach's alpha (on the diagonal), and correlations) for the variables of Sample 1 (n = 133).

	Mean	SD	Min	Max	VIF	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
						Correlations and Cronbach's alpha														
1. Impulsivity (self-report)	3.55	0.91	1.25	5.50	1.53	0.50														
2. Sensation Seeking (self-report)	5.12	1.03	2.00	7.00	1.29	0.39***	0.71													
3. ADHD symptoms (self-report)	2.79	0.52	1.67	5.00	1.32	0.39***	0.18*	0.52												
4. Age	22.23	2.46	17.00	32.00		−0.05	0.08	0.03	–											
5. Gender	0.39	0.49	0.00	1.00	1.62	0.04	−0.11	−0.19*	−0.11	–										
6. GNG Number Incorrect NoGo (behavior)	40.81	17.69	9.00	97.00	1.62	0.07	−0.06	0.00	0.03	0.04	–									
7. GNG Number Incorrect Go (behavior)	8.74	17.33	0.00	186.00	2.75	−0.17*	−0.13	−0.19*	0.15	0.01	0.08	–								
8. GNG Number Post-Incorrect Incorrect (behavior)	4.35	5.59	0.00	39.00	3.34	−0.09	−0.19*	−0.12	0.08	0.07	0.34***	0.76***	–							
9. GNG Average Response Time (behavior)	348.91	50.13	230.69	489.65	1.55	−0.08	−0.18*	−0.14	0.01	0.11	−0.17*	0.26**	0.25**	–						
10. EF Number Incorrect (behavior)	35.86	25.43	2.00	148.00	1.59	0.08	−0.12	−0.10	−0.09	0.13	0.34***	0.13	0.31***	0.11	–					
11. EF Average Response Time Incongruent (behavior)	429.81	41.26	320.14	543.44	1.37	−0.04	0.05	−0.12	0.05	0.18*	−0.10	−0.03	−0.13	0.22*	−0.22*	–				
12. EF Difference Average Response Time Post-Incorrect - Post-Correct (behavior)	19.05	28.53	−60.46	131.24	1.37	0.05	0.16	0.08	0.17	−0.02	−0.17*	0.12	0.03	0.06	−0.32***	0.36***	–			
13. GNG N2 (electrophysiology)[a]	−0.70	2.18	−5.13	5.20	1.65	0.13	0.03	−0.07	0.14	0.08	−0.27**	0.15	0.05	0.13	0.06	0.06	0.09	–		
14. GNG P3 (electrophysiology)[a]	5.00	4.37	−5.70	19.40	1.90	0.19*	0.16	0.12	−0.02	−0.03	0.01	−0.13	−0.21*	−0.40***	−0.03	−0.07	0.03	0.38***	–	
15. EF ERN (electrophysiology)[a]	−7.74	5.19	−20.62	5.00	1.32	−0.09	−0.14	0.02	−0.07	0.11	0.08	0.09	0.11	0.12	0.25**	0.10	−0.08	0.13	−0.07	–
16. EF Pe (electrophysiology)[a]	10.17	6.18	−3.82	30.52	1.35	−0.09	0.05	0.08	−0.23**	0.12	−0.08	−0.04	−0.10	−0.12	−0.24**	0.06	0.11	0.01	0.29***	0.20*

Note: ***: p < .001, **: p < .01, and *: p < .05, GNG = Go/No-Go, EF = Eriksen Flanker, [a]: difference score.

Table 2
Descriptive statistics (mean, standard deviation (SD), minimum (Min), maximum (Max), variance inflation factor (VIF), Cronbach's alpha (on the diagonal), and correlations) for the variables of Sample 2 (n = 142).

	Mean	SD	Min	Max	VIF	Correlations and Cronbach's alpha										
						1	2	3	4	5	6	7	8	9	10	11
1. Reward Responsiveness (self-report)	3.24	0.38	2.25	4.00	1.14	0.78										
2. Sensation Seeking (self-report)	3.20	0.71	1.25	4.75	1.20	0.19*	0.78									
3. ADHD symptoms (self-report)	2.75	0.54	1.67	4.00	1.14	−0.06	0.27**	0.50								
4. Age	20.63	2.04	18.00	30.00		0.09	0.20*	0.25**	–							
5. Gender	0.54	0.50	0.00	1.00		0.13	−0.07	0.09	−0.02	–						
6. BART Average Pumps (behavior)	61.86	10.09	24.87	90.83	1.18	−0.09	0.03	0.14	0.14	−0.22**	–					
7. BART Average Response Time (behavior)	6457.59	29574.15	1853.38	355985.00	1.18	0.11	−0.08	−0.15	−0.11	0.07	−0.31***	–				
8. REWARD N2 (electrophysiology)[a]	0.27	4.94	−13.21	16.32	4.55	0.17*	0.01	−0.05	0.01	0.05	−0.05	−0.02	–			
9. REWARD P2 (electrophysiology)[a]	−0.68	4.47	−13.12	10.06	3.50	0.11	0.05	−0.08	0.05	−0.03	−0.03	0.04	0.83***	–		
10. REWARD P3 (electrophysiology)[a]	−0.90	5.93	−14.51	14.87	2.81	0.09	−0.04	−0.04	0.01	0.05	−0.04	−0.03	0.79***	0.71***	–	
11. BART FRN (electrophysiology)[a]	0.26	2.46	−7.32	5.56	1.04	−0.04	0.13	0.01	0.08	−0.03	0.01	−0.06	0.01	0.00	0.05	–
12. BART P3 (electrophysiology)[a]	4.09	4.58	−8.39	21.15	1.09	−0.15	0.01	0.01	0.03	0.09	−0.17*	−0.06	−0.07	−0.01	−0.08	0.08

Note: ***: $p < .001$, **: $p < .01$, and *: $p < .05$, [a],: difference score.

and hence none of the *F*-values of Models 3 reaches significance. Therefore, neither the models in Sample 1 nor Sample 2 provide evidence for an association between self-reported *Impulsivity, Sensation Seeking, Reward Responsiveness*, and *ADHD symptoms* on the one hand, and behavioral and electrophysiological measures on the other.

3.4. Bootstrapping

The reported correlations and associations are based on two relatively large samples. However, many studies employ smaller samples, which reduces the chance that discovered effects are genuinely true. Therefore, we used bootstrapping (1000 iterations) to randomly select subsamples sized 20, 30 and 40 from our full sample to create an overview of the percentage of significant correlations and associations (based on a five percent significance level) we would have found if we had used such small samples. The results of this bootstrapping analysis are summarized in Table 5. With respect to the correlations, we cannot provide clear evidence that using smaller samples would have led to a higher percentage of significant values. However, compared to analyzing the full sample, analyzing smaller subsets (sized 20, 30 and 40) does increase the percentage of significant associations as found in the regression analyses for both Sample 1 (from 3.03 to 5.48–6.13) and Sample 2 (from 0.00 to 4.11–4.88). Hence, had our sample been smaller, we would have found more significant associations (using the same five percent significance level).

4. Discussion

The present paper examined the association between self-report measures, behavioral measures, and electrophysiological measures for the construct of impulsivity and related constructs such as sensation seeking, reward responsiveness, and ADHD symptoms. Although some previous studies report significant associations between self-reports, behavior, and electrophysiology, the present data were unable to confirm this. Using two large independent samples, we showed a high number of significant correlations *within* measurement levels, but only few significant correlations *between* different measurement levels. Regression analyses supported our correlational findings and showed no evidence of (joint) associations between behavior or electrophysiology, and self-reports. The few significant associations found between these measurement levels could not be interpreted as we adopted a five percent significance level. Bootstrap analyses showed that if we had used smaller sample sizes, like the ones used in many previous studies, the number of significant associations in our regression analyses would have been higher.

Our present null results deviate from the majority of previous studies as discussed in the introduction that in fact did find significant associations between self-reported impulsivity(-related) constructs and behavior/electrophysiology. The discrepancy between our current null-findings and previous research possibly results from the limitations that characterize our study. First, some self-report measures showed low reliability. This lower consistency could have arisen from study design; participants were asked to fill out the questionnaires at home instead of in a lab, which can have provoked careless responding. Therefore, future studies may consider extending the lab session to also incorporate filling out the questionnaires. Second, although our samples are large, they are limited with regard to participant type and geographical distribution. Both samples consisted of students, who were recruited using participant databases of the same university. We therefore recommend replicating the present study in other research labs and with a broader range of participants. Third, the measures ought to represent impulsivity, but are not entirely similar to impulsivity, which possibly led to less consistent results. For example, we adopted reward responsiveness as an impulsivity-related construct, even though Franken and Muris (2006) showed that the original reward responsiveness dimension (Gray, 1987) consists of two separate dimensions of which especially one (rash impulsiveness) is related to impulsivity. Therefore,

Table 3
Coefficients of the regression analyses (standard errors in brackets) for Sample 1.

	Impulsivity (self-report)			Sensation Seeking (self-report)			ADHD symptoms (self-report)		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Age	−0.03 (0.09)	−0.10 (0.09)	−0.09 (0.09)	0.06 (0.09)	0.08 (0.09)	0.06 (0.09)	0.02 (0.09)	0.05 (0.09)	0.06 (0.09)
Gender	0.04 (0.09)	0.06 (0.09)	0.05 (0.09)	−0.08 (0.09)	−0.09 (0.09)	−0.09 (0.09)	−0.15 (0.09)	−0.19* (0.09)	−0.16 (0.09)
GNG Number Incorrect No-Go (behavior)	0.07 (0.11)		0.12 (0.11)	−0.01 (0.10)		0.00 (0.11)	0.02 (0.10)		−0.02 (0.11)
GNG Number Incorrect Go (behavior)	−0.19 (0.14)		−0.21 (0.14)	−0.01 (0.14)		−0.01 (0.14)	−0.26 (0.14)		−0.28 (0.14)
GNG Number Post-Incorrect Incorrect (behavior)	−0.02 (0.16)		0.01 (0.16)	−0.15 (0.15)		−0.14 (0.16)	0.08 (0.15)		0.12 (0.16)
GNG Average Response Time (behavior)	−0.02 (0.10)		0.02 (0.11)	−0.15 (0.10)		−0.12 (0.11)	−0.05 (0.10)		−0.00 (0.11)
EF Number Incorrect (behavior)	0.12 (0.10)		0.07 (0.11)	0.01 (0.10)		0.04 (0.11)	−0.06 (0.10)		−0.07 (0.11)
EF Average Response Time Incongruent (behavior)	−0.08 (0.10)		−0.07 (0.10)	0.02 (0.10)		0.04 (0.10)	−0.14 (0.10)		−0.16 (0.10)
EF Difference Average Response Time Post-Incorrect - Post-Correct (behavior)	0.16 (0.10)		0.15 (0.10)	0.16 (0.10)		0.14 (0.10)	0.14 (0.10)		0.14 (0.10)
GNG N2 (electrophysiology)[a]		0.07 (0.10)	0.14 (0.11)		−0.01 (0.10)	0.03 (0.11)		−0.13 (0.10)	−0.10 (0.11)
GNG P3 (electrophysiology)[a]		0.20* (0.10)	0.14 (0.12)		0.13 (0.10)	0.04 (0.12)		0.15 (0.10)	0.12 (0.12)
EF ERN (electrophysiology)[a]		−0.07 (0.09)	−0.08 (0.10)		−0.13 (0.09)	−0.11 (0.10)		0.06 (0.09)	0.11 (0.10)
EF Pe (electrophysiology)[a]		−0.17 (0.10)	−0.14 (0.10)		0.06 (0.10)	0.04 (0.11)		0.06 (0.10)	0.03 (0.10)
F-value	0.96	1.70	1.34	1.38	1.27	1.08	1.70	1.55	1.40
p-value	0.47	0.13	0.20	0.20	0.28	0.38	0.10	0.17	0.17
R-squared (adj.)	−0.00	0.03	0.03	0.03	0.01	0.01	0.05	0.03	0.04
n	133	133	133	133	133	133	133	133	133

Note: ***: $p < .001$, **: $p < .01$, and *: $p < .05$, GNG = Go/No-Go, EF = Eriksen Flanker, [a]: difference score.

future studies examining impulsivity could benefit from using well-defined models to operationalize the construct. An example of such a model is UPPS (Whiteside & Lynam, 2001), which proposes that impulsivity is composed of four dimensions: urgency, sensation seeking, lack of perseverance, and lack of premediation. Finally, we analyzed

EEG with the use of difference waves because this method eliminates the influence of exogenous components (Miltner et al., 1997) and corrects for individual differences in general wave amplitude. However, the use of difference waves is also associated with interpretation issues and lower between-subject variance (Meyer, Lerner, De Los Reyes,

Table 4
Coefficients of the regression analyses (standard errors in brackets) for Sample 2.

	Reward Responsiveness (self-report)			Sensation Seeking (self-report)			ADHD symptoms (self-report)		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Age	0.11 (0.09)	0.10 (0.08)	0.11 (0.08)	0.20* (0.08)	0.18* (0.08)	0.18* (0.09)	0.23** (0.08)	0.26** (0.08)	0.24** (0.08)
Gender	0.11 (0.09)	0.13 (0.08)	0.11 (0.09)	−0.07 (0.09)	−0.05 (0.09)	−0.05 (0.09)	0.12 (0.08)	0.08 (0.08)	0.11 (0.09)
BART Average Pumps (behavior)	−0.04 (0.09)		−0.06 (0.09)	−0.03 (0.09)		−0.04 (0.09)	0.10 (0.09)		0.10 (0.09)
BART Average Response Time (behavior)	0.10 (0.09)		0.09 (0.09)	−0.07 (0.09)		−0.08 (0.09)	−0.10 (0.09)		−0.09 (0.09)
REWARD N2 (electrophysiology)[a]		0.28 (0.17)	0.29 (0.18)		0.01 (0.18)	−0.01 (0.18)		0.05 (0.17)	0.04 (0.17)
REWARD P2 (electrophysiology)[a]		−0.04 (0.15)	−0.06 (0.16)		0.15 (0.15)	0.17 (0.16)		−0.15 (0.15)	−0.13 (0.15)
REWARD P3 (electrophysiology)[a]		−0.12 (0.14)	−0.12 (0.14)		−0.16 (0.14)	−0.17 (0.14)		0.02 (0.14)	0.02 (0.14)
BART FRN (electrophysiology)[a]		−0.03 (0.08)	−0.03 (0.08)		0.13 (0.08)	0.13 (0.09)		−0.01 (0.08)	−0.01 (0.08)
BART P3 (electrophysiology)[a]		−0.15 (0.08)	−0.16 (0.09)		−0.02 (0.09)	−0.03 (0.09)		−0.00 (0.08)	0.01 (0.09)
F-value	1.40	1.73	1.60	1.74	1.48	1.22	3.63	1.69	1.69
p-value	0.24	0.11	0.12	0.15	0.18	0.29	0.01	0.12	0.10
R-squared (adj.)	0.01	0.04	0.04	0.02	0.02	0.01	0.07	0.03	0.04
n	142	142	142	142	142	142	142	142	142

Note: ***: $p < .001$, **: $p < .01$, and *: $p < .05$, [a]: difference score.

Table 5
The bootstrapped mean percentage of significant correlations/associations (based on 1000 iterations).

	Size subsample	Sample 1			Sample 2		
		20	30	40	20	30	40
Correlations	**Behavior vs. Self-report**	7.94	8.49	9.00	5.30	4.72	4.63
	Electrophysiology vs. Self-report	6.19	6.38	6.70	5.82	5.39	5.07
	Behavior vs. Electrophysiology	8.42	9.98	11.87	5.06	4.68	4.69
Associations	**Behavior/Electrophysiology vs. Self-report (Models 1 and 2)**	5.48	5.86	6.13	4.88	4.42	4.11

Laird, & Hajcak, 2017), which possibly influenced our results. Re-running the main analyses using absolute instead of difference waves indicated that this was the case for one electrophysiological measure, the *GNG P3* in response to no-go trials, which showed more significant associations with self-reports and behavioral measures than did the difference wave. However, no notable discrepancies were observed for the other ERPs.

In addition to the limitations of our study, there are several more general explanations of why we did not find significant correlations/associations between the measurement levels. First, the time frames of behavioral/electrophysiological measures on the one hand and self-report measures on the other hand differ. Typically, behavioral and electrophysiological measures are in the range of (hundreds of) milliseconds, whereas self-report measures are commonly measured as a trait, hence over several years. In other words, behavioral and electrophysiological measures probe state impulsivity, whereas self-reports probe trait impulsivity. However, for the present data the correlations between the two state impulsivity measures (behavior and electrophysiology) did not outperform the correlations between the trait impulsivity measure (self-report) and either state impulsivity measure, indicating that this argument is (at least in itself) not sufficient to explain the lack of correlation between different measurement levels as found in the present study.

A second factor that may have contributed to the present results also focuses on the nature of the measurements. Behavior and electrophysiology are implicit measures because they largely operate outside awareness, whereas self-reports represent the more conscious processes and are therefore explicit measures (Dittmar et al., 2011; Eysenck, 1992). However, this discrepancy between implicit and explicit measures does not appear to be sufficient to explain the current findings because again our correlations between behavior and electrophysiology (both implicit) did not clearly outperform the correlations between either of these measures and the (explicit) self-reports.

A third possible explanation for our lack of associations across measurement levels is that cognitive paradigms such as the ones used here may be unable to predict individual differences. Hedge, Powell, and Sumner (2017) state that cognitive paradigms have become well-established as a result of the low between-subject variability of their outcomes (e.g. reaction time, performance), but that this low between-subject variability causes low reliability for individual differences, making it difficult for tasks to consistently predict brain activity or self-report. Hedge et al. (2017) support their premise by showing that the intraclass correlations (ICCs) of seven classic tasks are relatively low. Other studies (focused on the dot-probe task) have supported the premise as well by showing that whereas ERPs in the task are internally reliable, reaction time differences are not (Kappenman, Farrens, Luck, & Proudfit, 2014; Reutter, Hewig, Wieser, & Osinsky, 2017). However, of the low ICCs reported by Hedge et al. (2017), the ones related to our tasks (i.e. the Eriksen Flanker task and the Go/No-Go task) were relatively favorable, ranging from moderate to excellent. Furthermore, the issue raised by Hedge et al. (2017) is limited to explaining the lack of correlations/associations between behavior and self-reports or electrophysiology, but cannot explain why self-reports and electrophysiology do not correlate with each other.

A final explanation for our present null-findings concerns a premise that we discussed in the introduction and that was partly supported by our own data: many previous studies employ small sample sizes, leading to low statistical power and a lower chance that findings are true. This explanation does not discard the other explanations we discussed, but can contrary to these other explanations explain both the current null-findings and the significant results reported in previous studies. The fact that most studies employing neurophysiology have a limited number of participants is understandable given that collecting such data requires a high investment of time and money. However, small samples can be considered 'unsafe' as they lead to low power (1 − β), the chance that effects are genuinely true (Button et al., 2013; Forstmeier et al., 2017; Ioannidis, 2005). Low-powered studies in turn have an increased chance at a Type II error (false negative: β), and have a lower positive predictive value (PPV), the probability that a positive finding is a true positive. Sample size does not directly impact the chance at a Type I error (false positive: α) since this is a fixed value chosen by the researcher. However, this chance can increase as a result of flexibility in methodological choices (Simmons, Nelson, & Simonsohn, 2011), which is particularly powerful when using small samples.

The problems related to low sample size are augmented by the file drawer problem (Rosenthal, 1979), the observation that null findings (such as the present ones) are often not distributed (Song et al., 2009) because journals are reluctant to publish null-findings and because scholars are hesitant to submit them in the first place (Ferguson & Heene, 2012). Together, small sample sizes and a bias towards publishing significant findings could explain the discrepancy between our current null-findings and the significant results reported in previous literature. To address these issues, it is important for future research to replicate small *n* studies. Replicating these studies in larger samples will not suddenly eradicate all positive findings. In fact, some studies examining multiple measurement levels for impulsivity did find significant associations using large samples. For example, Ait Oumeziane and Foti (2016) showed that lack of premeditation (a facet of impulsivity) is associated with decreased P3 amplitudes in individuals with low depression scores, but increased amplitudes in individuals who score high on depression. Furthermore, Hill, Samuel, and Foti (2016) reported that negative urgency, another facet of impulsivity, is associated with an increased Eriksen Flanker ERN in people who report low conscientiousness, whereas no association was observed for high conscientious people. The sample size of these studies was respectively *n* = 260 and *n* = 208. Carrying out such large-scale studies is imperative to provide results that are safe to interpret and that are hence truly informative regarding the relationship between different measurement levels. Unmistakably, this message is not confined to impulsivity research but applies to all constructs that can be measured on multiple levels of measurement.

Declaration of interest

None.

References

Ait Oumeziane, B., & Foti, D. (2016). Reward-related neural dysfunction across depression and impulsivity: A dimensional approach. *Psychophysiology, 53*(8), 1174–1184.

Arvanitakis, Z. (2010). Update on frontotemporal dementia. *The Neurologist, 16*(1), 16–22.

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex, 10*(3), 295–307.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., ... Munafo, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. Nature reviews. *Neuroscience, 14*, 365–376.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology, 67*(2), 319–333.

Chamberlain, S. R., Stochl, J., Redden, S. A., & Grant, J. E. (2018). Latent traits of impulsivity and compulsivity: Toward dimensional psychiatry. *Psychological Medicine, 48*(5), 810–821.

Chandrakumar, D., Feuerriegel, D., Bode, S., Grech, M., & Keage, H. A. (2018). Event-related potentials in relation to risk-taking: A systematic review. *Frontiers in Behavioral Neuroscience, 12*, 111.

Chaudhuri, K. R., Odin, P., Antonini, A., & Martinez-Martin, P. (2011). Parkinson's disease: The non-motor issues. *Parkinsonism & Related Disorders, 17*(10), 717–723.

Dawe, S., Gullo, M. J., & Loxton, N. J. (2004). Reward drive and rash impulsiveness as dimensions of impulsivity: Implications for substance misuse. *Addictive Behaviors, 29*(7), 1389–1405.

De Groot, K., & Thurik, R. (2018). Disentangling risk and uncertainty: When risk-taking measures are not about risk. *Frontiers in Psychology, 9*, 2194.

De Pascalis, V., Varriale, V., & D'Antuono, L. (2010). Event-related components of the punishment and reward sensitivity. *Clinical Electrophysiology, 121*(1), 60–76.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research, 61*(12), 1203–1218.

Dittmar, O., Krehl, R., & Lautenbacher, S. (2011). Association of self-report, behavioral and electrophysiological measures assessing pain-related information processing. *Pain Research & Management, 16*(1), 33–40.

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica, 30*, 412–431.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Attention, Perception & Psychophysics, 16*(1), 143–149.

Euser, A. S., Van Meel, C. S., Snelleman, M., & Franken, I. H. A. (2011). Acute effects of alcohol on feedback processing and outcome evaluation during risky decision-making: An ERP study. *Psychopharmacology, 217*(1), 111–125.

Eysenck, M. W. (1992). *Anxiety.* Hove: Lawrence Erlbaum Associates Ltd.

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science, 7*(6), 555–561.

Forstmeier, W., Wagenmakers, E., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings – A practical guide. *Biological Reviews, 92*, 1941–1968.

Franken, I. H. A., & Muris, P. (2006). Gray's impulsivity dimension: A distinction between reward sensitivity versus rash impulsiveness. *Personality and Individual Differences, 40*(7), 1337–1347.

Franken, I. H. A., Van den Berg, I., & Van Strien, J. W. (2010). Individual differences in alcohol drinking frequency are associated with electrophysiological responses to unexpected nonrewards. *Alcoholism, Clinical and Experimental Research, 34*(4), 702–707.

Gao, Y., Chen, H., Jia, H., Ming, Q., Yi, J., & Yao, S. (2016). Dysfunctional feedback processing in adolescent males with conduct disorder. *International Journal of Psychophysiology, 99*, 1–9.

Geburek, A. J., Rist, F., Gediga, F., Stroux, D., & Pedersen, A. (2013). Electrophysiological indices of error monitoring in juvenile and adult attention deficit hyperactivity disorder (ADHD) – A meta-analytic appraisal. *International Journal of Psychophysiology, 87*(3), 349–362.

Gerbring, D. W., Ahadi, S. A., & Patton, J. H. (1987). Toward a conceptualization of impulsivity: Components across the behavioral and self-report domains. *Multivariate Behavioral Research, 22*(3), 357–379.

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55*(4), 468–484.

Gray, J. A. (1987). *The psychology of fear and stress.* Cambridge: Cambridge University Press.

Groen, Y., Wijers, A. A., Mulder, L. J., Waggeveld, B., Minderaa, R. B., & Althaus, M.

(2008). Error and feedback processing in children with ADHD and children with Autistic Spectrum Disorder: An EEG event-related potential study. *Clinical Electrophysiology, 119*(11), 2476–2493.

Hair, J. F., Anderson, R. E., Babin, B. J., & Black, W. C. (2010). *Multivariate data analysis: A global perspective.* Upper Saddle River, New Jersey: Pearson.

Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavioral Research Methods,* 1–21.

Herrmann, M. J., Saathoff, C., Schreppel, T. J., Ehlis, A. C., Scheuerpflug, P., Pauli, P., ... Fallgatter, A. J. (2009). The effect of ADHD symptoms on performance monitoring in a non-clinical population. *Psychiatry Research, 169*(2), 144–148.

Hill, K. E., Samuel, D. B., & Foti, D. (2016). Contextualizing individual differences in error monitoring: Links with impulsivity, negative affect, and conscientiousness. *Psychophysiology, 53*(8), 1143–1153.

Hoyle, R. H., Stephenson, M. T., Palmgreen, P., Lorch, E. P., & Donohew, R. L. (2002). Reliability and validity of a brief measure of sensation seeking. *Personality and Individual Differences, 32*(3), 401–414.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), e124.

Jones, H. A., & Lejuez, C. W. (2005). Personality correlates of caffeine dependence: The role of sensation seeking, impulsivity, and risk taking. *Experimental and Clinical Psychopharmacology, 13*(3), 259–266.

Jonkman, L. M., Van Melis, J. J., Kemner, C., & Markus, C. R. (2007). Methylphenidate improves deficient error evaluation in children with ADHD: An event-related brain potential study. *Biological Psychology, 76*(3), 217–229.

Kalkhoven, C. E., Sennef, C., Peeters, A., & Van Den Bos, R. (2014). Risk-taking and pathological gambling behavior in Huntington's disease. *Frontiers in Behavioral Neuroscience, 8*, 103.

Kam, J. W., Dominelli, R., & Carlson, S. R. (2012). Differential relationships between sub-traits of BIS-11 impulsivity and executive processes: An ERP study. *International Journal of Psychophysiology, 85*(2), 174–187.

Kappenman, E. S., Farrens, J. L., Luck, S. J., & Proudfit, G. H. (2014). Behavioral and ERP measures of attentional bias to threat in the dot-probe task: Poor reliability and lack of correlation with anxiety. *Frontiers in Psychology, 5*, 1368.

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E. V. A., ... Ustun, T. B. (2005). The World Health Organization Adult ADHD Self-Report Scale (ASRS): A short screening scale for use in the general population. *Psychological Medicine, 35*(2), 245–256.

Khadka, S., Stevens, M. C., Aslanzadeh, F., Narayanan, B., Hawkins, K. A., Austad, C. S., ... Potenza, M. N. (2017). Composite impulsivity-related domains in college students. *Journal of Psychiatric Research, 90*, 118–125.

Lansbergen, M. M., Böcker, K. B., Bekker, E. M., & Kenemans, J. L. (2007). Neural correlates of stopping and self-reported impulsivity. *Clinical Electrophysiology, 118*(9), 2089–2103.

Lauriola, M., Panno, A., Levin, I. P., & Lejuez, C. W. (2014). Individual differences in risky decision making: A meta-analysis of sensation seeking and impulsivity with the balloon analogue risk task. *Journal of Behavioral Decision Making, 27*(1), 20–36.

Lejuez, C. W., Aklin, W. M., Zvolensky, M. J., & Pedulla, C. M. (2003). Evaluation of the Balloon Analogue Risk Task (BART) as a predictor of adolescent real-world risk-taking behaviors. *Journal of Adolescence, 26*(4), 465–479.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., ... Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology Applied, 8*(2), 75–84.

Limbrick-Oldfield, E. H., Van Holst, R. J., & Clark, L. (2013). Fronto-striatal dysregulation in drug addiction and pathological gambling: Consistent inconsistencies? *NeuroImage Clinical, 2*, 385–393.

Liotti, M., Pliszka, S. R., Perez, R., Kothmann, D., & Woldorff, M. G. (2005). Abnormal brain activity related to performance monitoring and error detection in children with ADHD. *Cortex, 41*(3), 377–388.

Littel, M., Van den Berg, I., Luijten, M., Rooij, A. J., Keemink, L., & Franken, I. H. A. (2012). Error processing and response inhibition in excessive computer game players: An event-related potential study. *Addiction Biology, 17*(5), 934–947.

Logan, G. D., Schachar, R. J., & Tannock, R. (1997). Impulsivity and inhibitory control. *Psychological Science, 8*(1), 60–64.

Lopez, R., Dauvilliers, Y., Jaussent, I., Billieux, J., & Bayard, S. (2015). A multi-dimensional approach of impulsivity in adult attention deficit hyperactivity disorder. *Psychiatry Research, 227*(2-3), 290–295.

Luijten, M., Van Meel, C. S., & Franken, I. H. (2011). Diminished error processing in smokers during smoking cue exposure. *Pharmacology, Biochemistry, and Behavior, 97*(3), 514–520.

Marhe, R., Van de Wetering, B. J., & Franken, I. H. (2013). Error-related brain activity predicts cocaine use after treatment at 3-month follow-up. *Biological Psychiatry, 73*(8), 782–788.

Meda, S. A., Stevens, M. C., Potenza, M. N., Pittman, B., Gueorguieva, R., Andrews, M. M., ... Pearlson, G. D. (2009). Investigating the behavioral and self-report constructs of impulsivity domains using principal component analysis. *Behavioural Pharmacology, 20*, 390–399.

Meyer, A., Lerner, M. D., De Los Reyes, A., Laird, R. D., & Hajcak, G. (2017). Considering ERP difference scores as individual difference measures: Issues with subtraction and alternative approaches. *Psychophysiology, 54*(1), 114–122.

Miltner, W. H., Braun, C. H., & Coles, M. G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. *Journal of Cognitive Neuroscience, 9*(6), 788–798.

Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., & Swann, A. C. (2001). Psychiatric aspects of impulsivity. *The American Journal of Psychiatry, 158*(11), 1783–1793.

Moser, J. S., Durbin, C. E., Patrick, C. J., & Schmidt, N. B. (2015). Combining neural and behavioral indicators in the assessment of internalizing psychopathology in children and adolescents. *Journal of Clinical Child & Adolescent Psychology, 44*(2), 329–340.

Pleskac, T. J., Wallsten, T. S., Wang, P., & Lejuez, C. W. (2008). Development of an

automatic response mode to improve the clinical utility of sequential risk-taking tasks. *Experimental and Clinical Psychopharmacology, 16*(6), 555–564.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *The Journal of Applied Psychology, 88*(5), 879–903.

Potts, G. F., George, M. R. M., Martin, L. E., & Barratt, E. S. (2006). Reduced punishment sensitivity in neural systems of behavior monitoring in impulsive individuals. *Neuroscience Letters, 397*(1), 130–134.

Potts, G. F., Martin, L. E., Burton, P., & Montague, P. R. (2006). When things are better or worse than expected: The medial frontal cortex and the allocation of processing resources. *Journal of Cognitive Neuroscience, 18*(7), 1112–1119.

Reutter, M., Hewig, J., Wieser, M. J., & Osinsky, R. (2017). The N2pc component reliably captures attentional bias in social anxiety. *Psychophysiology, 54*(4), 519–527.

Reynolds, B., Ortengren, A., Richards, J. B., & De Wit, H. (2006). Dimensions of impulsive behavior: Personality and behavioral measures. *Personality and Individual Differences, 40*, 305–315.

Rietdijk, W. J., Franken, I. H., & Thurik, A. R. (2014). Internal consistency of event-related potentials associated with cognitive control: N2/P3 and ERN/Pe. *PloS One, 9*(7), e102672.

Romer, D. (2010). Adolescent risk taking, impulsivity, and brain development: Implications for prevention. *Developmental Psychobiology, 52*(3), 263–276.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641.

Salim, M. A. M., Van der Veen, F. M., Van Dongen, J. D., & Franken, I. H. (2015). Brain activity elicited by reward and reward omission in individuals with psychopathic traits: An ERP study. *Biological Psychology, 110*, 50–58.

Sharma, L., Markon, K. E., & Clark, L. A. (2014). Toward a theory of distinct types of "impulsive" behaviors: A meta-analysis of self-report and behavioral measures. *Psychological Bulletin, 140*(2), 374–408.

Shen, I. H., Lee, D. S., & Chen, C. L. (2014). The role of trait impulsivity in response inhibition: Event-related potentials in a stop-signal task. *International Journal of Psychophysiology, 91*(2), 80–87.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366.

Song, F., Parekh-Bhurke, S., Hooper, L., Loke, Y. K., Ryder, J. J., Sutton, A. J., ... Harvey, I. (2009). Extent of publication bias in different categories of research cohorts: A meta-analysis of empirical studies. *BMC Medical Research Methodology, 9*(1), 79.

Soong, A. C., Lind, J. C., Shaw, G. R., & Koles, Z. J. (1993). Systematic comparisons of interpolation techniques in topographic brain mapping. *Electroencephalography and Clinical Neurophysiology, 87*(4), 185–195.

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review, 28*(1), 78–106.

Steinberg, L., Albert, D., Cauffman, E., Banich, M., Graham, S., & Woolard, J. (2008). Age differences in sensation seeking and impulsivity as indexed by behavior and self-

report: Evidence for a dual systems model. *Developmental Psychology, 44*(6), 1764–1778.

Sutcubasi, B., Metin, B., Tas, C., Krzan, F. K., Sarı, B. A., Ozcimen, B., ... Tarhan, N. (2018). The relationship between responsiveness to social and monetary rewards and ADHD symptoms. *Cognitive, Affective & Behavioral Neuroscience, 18*(5), 857–868.

Taylor, J. B., Visser, T. A., Fueggle, S. M., Bellgrove, M. A., & Fox, A. M. (2018). The error-related negativity (ERN) is an electrophysiological marker of motor impulsiveness on the Barratt Impulsiveness Scale (BIS-11) during adolescence. *Developmental Cognitive Neuroscience, 30*, 77–86.

Van den Berg, I., Franken, I. H., & Muris, P. (2010). A new scale for measuring reward responsiveness. *Frontiers in Psychology, 1*, 239.

Van den Berg, I., Franken, I. H., & Muris, P. (2011). Individual differences in sensitivity to reward. *Journal of Psychophysiology, 25*(2), 81–86.

Van Meel, C. S., Heslenfeld, D. J., Oosterlaan, J., & Sergeant, J. A. (2007). Adaptive control deficits in attention-deficit/hyperactivity disorder (ADHD): The role of error processing. *Psychiatry Research, 151*(3), 211–220.

Wang, W., & Wang, Y. H. (2001). Sensation seeking correlates of passive auditory P3 to a single stimulus. *Neuropsychologica, 39*(11), 1188–1193.

Warren, C. M., & Holroyd, C. B. (2012). The impact of deliberative strategy dissociates ERP components related to conflict processing vs. reinforcement learning. *Frontiers in Neuroscience, 6*, 43.

Webster, G. D., & Crysel, L. C. (2012). 'Hit me, maybe, one More Time': Brief measures of impulsivity and sensation seeking and their prediction of blackjack bets and sexual promiscuity. *Journal of Research in Personality, 46*(5), 591–598.

Whiteside, S. P., & Lynam, D. R. (2001). The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences, 30*(4), 669–689.

Wiersema, J. R., Van der Meere, J. J., & Roeyers, H. (2005). ERP corerlates of impared error monitoring in children with ADHD. *Journal of Neural Transmission, 112*(10), 1417–1430.

Wiersema, J. R., Van der Meere, J. J., & Roeyers, H. (2009). ERP correlates of error monitoring in adult ADHD. *Journal of Neural Transmission, 116*(3), 371–379.

Zheng, Y., Sheng, W., Xu, J., & Zhang, Y. (2014). Sensation seeking and error processing. *Psychophysiology, 51*(9), 824–833.

Zhou, Z. H., Yuan, G. Z., Yao, J. J., Li, C., & Cheng, Z. H. (2010). An event-related potential investigation of deficient inhibitory control in individuals with pathological Internet use. *Acta Neuropsychiatrica, 22*(5), 228–236.

Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking.* New York, NY: Cambridge University Press.

Zuckerman, M., & Neeb, M. (1979). Sensation seeking and psychopathology. *Psychiatry Research, 1*(3), 255–264.

Zuckerman, M., Kuhlman, D. M., Joireman, J., Teta, P., & Kraft, M. (1993). A comparison of three structural models for personality: The Big three, the Big Five, and the Alternative five. *Journal of Personality and Social Psychology, 65*(4), 757–768.