Kristel de Groot

Challenges in the Measurement and Interpretation of Risk-Related Behaviour and Brain Activity in Lab Studies

Challenges in the Measurement and Interpretation of Risk-Related Behaviour and Brain Activity in Lab Studies

Challenges in the Measurement and Interpretation of Risk-Related Behaviour and Brain Activity in Lab Studies

Uitdagingen bij het meten en interpreteren van risicogerelateerd gedrag en hersenactiviteit in laboratoriumstudies

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr.ir. A.J. Schuit

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on
Wednesday 3 December 2025 at 13.00 hrs

by

Kristel de Groot
born in Gouda, the Netherlands.

Erasmus University Rotterdam

Doctoral Committee:

Promotor:	Prof.dr. C.A. Rietveld
Copromotors:	Dr. O. Lindemann Prof.dr. J.W. van Strien
Other members:	Prof.dr. J.L. Ellis Prof.dr. K.I.M. Rohde Dr. F.M. van der Veen

Erasmus Research Institute of Management – ERIM
The joint research institute of the Rotterdam School of Management (RSM)
and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam
Internet: www.erim.eur.nl

ERIM PhD Series in Research in Management, 610
ERIM reference number: EPS-2025-610-S&E
ISBN 978-90-5892-751-4

Cover: Herman Jan Couwenberg / Djanko, www.djanko.nl
Design: Canon
Print: OBT bv, www.obt.eu

This publication (cover and interior) is printed on certified FSC® paper.

FSC www.fsc.org
MIX
Papier | Ondersteunt verantwoord bosbeheer
FSC® C022586

Schelluinen, 28 juni 2012

Beste Kristel,
Van harte met jouw VWO-diploma.
Het cadeautje is een aanmoediging,
hoewel het op een ander vakgebied
voor jou zal zijn, waarschijnlijk.
Als het je lukt om een proefschrift
te schrijven, hetgeen ik uiteraard
verwacht: laat het even weten!

Een veel plezierige tijdens
jouw studententijd -
ten
Hoewel soms moeilijk, ben
ik trots en verheugd dat ik
met je het mogen werken
De grote, enge, verre man. Bert
ZOZ →

Acknowledgements

A popular quote by Dr. John Mola states: “Finishing a PhD is like finishing a group project where your partner made a ton of mistakes at the beginning of the assignment. Except your partner is just you 4 years ago”. I relate to this wholeheartedly. Looking at the final product of my PhD journey – this book – I notice flaws: missed opportunities to make stronger arguments, suboptimal design choices… I see things I could have done much better. This may sound negative, but I believe it is not. As a researcher, especially early in your career, it is natural to look at prior work and not be satisfied. In fact, this signals something good: that you have become better. Noticing weak spots in my work shows that I have grown – arguably one of the main objectives of doing a PhD in the first place.

Importantly, growth does not happen in a vacuum. It involves the people around you, and I am grateful to those who helped me grow over these past years.

First, I would like to thank my supervisors. Jan, you were an excellent mentor during my master’s thesis and I am glad I could continue working with you during my PhD. Even after your retirement, I could always reach out with questions on EEG analysis, and you would draw on your many years of experience to help me find a solution. Oliver, thank you for your continued belief in me, for giving me the freedom to explore a wide range of topics, and for introducing me to new interests, like the project on statistics anxiety. Niels, your door was always open (literally and figuratively), which I appreciated more than you may know. You always say that I am a worrier, and you are right about that. All the more reason for me to be grateful for your encouragement; your consistently swift attention to work I shared; and your kind, capable, and no-nonsense guidance.

Second, I want to thank others who provided mentorship. Liselotte, I am grateful I had the opportunity to teach alongside you. You showed me the kind of teacher I aspire to be: approachable, enthusiastic, and holistic. Roy and Jan (van Ours), thank you both for guiding me through part of my PhD journey. To my co-authors: much is learned from writing papers, even more so when doing it with others. I thoroughly enjoyed our collaborations and hope to work together again. To my committee members: thank you for taking the time to read my work and helping me reflect on it further during the defence. Importantly, I want to acknowledge my secondary school teachers who encouraged me to attend university: Boersma, De Jong, Posthuma, and Van Wijk (†). It was not an obvious choice for me to pursue higher education. You helped me find my way.

Third, a big thank-you to my colleagues. Too many to chronicle, but I want to name a few. Indy and Nienke, I feel fortunate to have met you; I enjoyed our collaborations and still enjoy our cups of tea. Eric and Romina, thank you for being my paranymphs. You have become friends more than colleagues, supporting me more than I could have wished for. Work (and beyond) is more exciting with you around. Matthew, Miranda, Vahid, thank you for our chats and your valuable advice. A special thanks to the professional support staff: everyone at the secretariats, the doctoral office, and the Erasmus Behavioural Lab. Thank you for your practical and moral support. Ylaise, we go way back to the H-building. Thank you for your companionship and for always knowing how to cheer me up.

Fourth, I want to acknowledge my students. I have enjoyed creating PowerPoints and workbooks, watching your presentations, and supervising your theses. Above all, I appreciated your curious and critical questions, which motivated me to dive deeper into the material and hone my teaching skills. Seeing you grow has been a pleasure, and I am grateful to you for helping me grow.

Fifth, I want to thank the people closest to me. My grandparents Jan (†), Janny, Hans, and Tiny, who provided a loving and supportive environment in which I could safely learn and grow. My in-laws, who welcomed me warmly and always show interest in my research. My longtime friends Robin and Diewertje, who are there for me without exception. And my partner Sander, who deserves the most thanks of all. Sander, not only are you my best friend and a brilliant sounding board on any topic, but you also stood by me when I became ill. I could not have come back to finish this PhD without your unwavering support and am grateful for your understanding and kindness over these years. We have both grown, and I look forward to growing further together.

Table of Contents

Chapter 1

General Introduction

1.1 Research Motivation and Aims

More so than the natural sciences and especially in stark contrast to the formal sciences, the social sciences rest fundamentally on operationalisation. Whereas the thermal velocity of helium at room temperature, the mass of a planet like Jupiter, or the speed with which an acorn falls towards the forest floor can be measured directly, that is, through universally accepted laws and derivations, "fuzzy" concepts such as ambiguity aversion, impulsivity, or self-efficacy are not readily measurable, and must first be *defined*. This process of changing a conceptual variable into an operational variable (i.e., operationalisation) is integral to good scientific practice, as it makes hypotheses falsifiable and enables researchers to replicate findings.

Operationalisation, however, is not an easy feat. As Landis and Cortina (2015) stress: "the fact that we must define and measure unobservable constructs adds an element of difficulty to our sciences that does not exist in others" (p. 15). Most commonly, concepts in the social sciences can be defined and measured in multiple ways, without one clearly surpassing the others. If two researchers were to measure a participant's height, major disagreement would be unlikely to occur. Compare this to, for example, assessing anhedonia in participants with depression, for which one review alone discussed nine possible self-report scales and seven types of behavioural measures (Rizvi et al., 2016); or to the measurement of mental fatigue, for which a recent paper reviewed no fewer than five self-report scales, two behavioural measures, and three types of (neuro)physiological measurements (Kunasegaran et al., 2023). These are not extreme examples; rather, they provide a typical representation of the large number of options that are generally available for operationalising concepts in the social sciences.

When looking at the choices researchers make when deciding on operationalisations, recent years have shown an increase in the popularity of behavioural and (neuro)physiological measures over introspective, self-report ones in various disciplines, such as psychology (Haslam et al., 2022; Schwartz et al., 2016; Wieczorek et al., 2021), management (Alsharif & Mohd Isa, 2024; Di Stefano & Gutierrez; Suomala, 2018a, 2018b), and economics (Borawska, 2017; Gunessee & Lane, 2023; Nermend & Łatuszyńska, 2016). In part, this can be attributed to these measures becoming increasingly accessible. Software for behavioural tasks or for physiological measurement systems now often comes equipped with a graphical user interface, and no longer requires advanced programming skills. Moreover, open-source software is becoming more common (Mathôt et al., 2012), reducing cost and fostering the proliferation of public repositories featuring readily-available tasks. Furthermore, the options for hosting behavioural tasks online instead of in a laboratory have

rapidly expanded (Rodd, 2024), enabling researchers to collect this data on a larger scale and in more diverse populations. Similar trends can be observed for neurophysiological measures, with increasingly more low-cost and portable technologies being available, especially for electroencephalography (EEG) and eye-tracking (Funke et al., 2016; Krigolson et al., 2017).

Another reason for the popularity of behavioural and neurophysiological measures, however, are the qualities these measures are presumed to possess – or rather, their ability to circumvent certain downsides of self-report measures (Corneille & Gawronski, 2024). Since the early days of psychological assessment, self-reports – in which research participants are asked directly about their mental content – have received criticism, among others regarding social desirability bias (S. B. Eysenck et al., 1966) and the limited introspective abilities participants have (Nisbett & Wilson, 1977). What participants say in self-reports may not be accurate, as they may not wish to share their thoughts or feelings, or have difficulty accessing them. For these reasons, self-reports are frequently referred to as "subjective" measurements of the concepts one tries to research, whereas behavioural and neurophysiological measures are often said to "objectively" assess these same concepts (Alsharif & Mohd Isa, 2024; Rizvi et al., 2016). A popular notion seems to be that by using behavioural and neurophysiological measures we *inherently come closer* to the concepts we intend to measure, thus providing us with a better operationalisation.

What appears to be overlooked, however, is that behavioural and neurophysiological measures are *still operationalisations*, and therefore come with their own challenges and trade-offs. Suppose we aim to study impulsivity, a fuzzy concept that we cannot directly measure. One way in which we could operationalise this is with a self-report scale, like the Barratt Impulsiveness Scale 11 (Patton et al., 1995). Another way could be through impulsive action as demonstrated by errors made on a behavioural task like the Go/No-Go task (Weafer et al., 2013), or through measurement of brain activity at the time participants make these erroneous decisions (Ruchsow et al., 2008). Can we confidently say that the latter two methods bring us closer to the concept of interest and are thus a better choice? True, when interpreting participants' error rate or brain activity, there is no need to take into account answering bias or limited introspection. But the number of errors made by a participant could easily be impacted by their ability to pay continuous attention, or even by their visual acuity, which would render it a poor operationalisation nonetheless. Likewise, brain activity recorded during the decision could be confounded by activity related to the motor response the task requires. Moreover, before interpretation of brain data is possible, much cleaning of the raw signals is needed, which could also introduce bias. As such, it is evident that behavioural and neurophysiological measures are not inherently valid. Just like

with self-reports, their value is conditional on their characteristics and on how well these fit the concept one intends to measure.

For this reason, the present dissertation critically evaluates the characteristics of various behavioural and neurophysiological measures. In particular, it focuses on measures that aim to operationalise *risk-taking propensity (RTP)* and related concepts like reward responsiveness, as these play an important role across several scientific disciplines and are now frequently examined using behavioural and neurophysiological measures, in addition to self-reports. RTP is a trait-like attribute that can be defined as someone's overall willingness or inclination to take risks. It is generally higher in men than in women, and for both genders peaks in adolescence and young adulthood, and declines with increased age (Byrnes et al., 1999; Collado et al., 2014; Josef et al., 2016). Individuals' risk-taking propensity has the potential to greatly impact life outcomes, including through health behaviours like substance use or unsafe driving (Anderson & Mellor, 2008; C. Wang et al., 2025); occupational behaviours like the decision to start a business, and the success of that venture (Danso et al., 2016; Nieß & Biemann, 2014); and mobility and migration behaviours (Balaz & Williams, 2011; Jaeger et al., 2010). Consequently, the measurement of RTP takes a central position in various social sciences, such as psychology, economics, and management.

The dissertation is divided into two parts. Part I is dedicated to challenges related to measuring risk-taking propensity using *behavioural tasks*, with a special focus on the Balloon Analogue Risk Task (BART; Lejuez et al., 2002), a popular computer task that is often employed in psychological literature to measure RTP and related concepts like impulsivity and reward responsiveness. In addition, a more recently developed measure, the Columbia Card Task (CCT; Figner et al., 2009), is explored, which also originated in the psychological literature but has great potential for use in management and economics. Part II of the dissertation is dedicated to challenges in measuring risk-taking propensity with *brain activity*, in particular electroencephalography (EEG) as recorded during the BART, CCT, and other behavioural tasks. Here, extra consideration is given to the problem of small sample sizes, which is common in EEG research (Clayson et al., 2019). Across Parts I and II, the overarching question that is addressed is to what extent behaviour and brain activity recorded during computer tasks in a laboratory setting are successful in operationalising concepts such as risk-taking propensity; or, in more generic terms: what conclusions can we draw from findings obtained from these behavioural and brain measures?

1.1.1 Part I: Behavioural Tasks

Chapters 2, 3, and 4 form Part I of the dissertation, which focuses on challenges related to measuring risk-taking propensity using behavioural tasks. First, Chapter 2 dives deeper into the definition of RTP. In particular, it explains how decision-making can take place under conditions of risk or under conditions of uncertainty, and how this distinction is often not addressed in psychological literature despite a great deal of empirical research showing that participants' behaviour and brain activity differ between these contexts. Furthermore, Chapter 2 explains how risk-taking behaviour in the Balloon Analogue Risk Task (BART; Lejuez et al., 2002), the most frequently used instrument for measuring risk-taking propensity in psychology (C. Wang et al., 2025), can best be characterised as decision-making under uncertainty, or as a shift across trials from decision-making under uncertainty to decision-making under risk, and discusses the implications this has for its use as an operationalisation of RTP.

Chapter 3 further concentrates on the BART and provides an in-depth overview of four methodological shortcomings of the task, including the shift from decision-making under uncertainty to decision-making under risk that was introduced in Chapter 2. Collectively, the shortcomings that are reviewed suggest that the BART is not suited as a measure of risk-taking propensity – and possibly not of related concepts like impulsivity and reward responsiveness either. Chapter 3 offers various concrete solutions for researchers to mitigate these shortcomings in order to measure RTP more aptly, denouncing a one-size-fits-all approach and emphasising the importance of carefully matching the operationalisation to the specific research question a study aims to answer.

Finally, Chapter 4 implements several of the solutions that are proposed in Chapter 3 with the aim of measuring risk-taking propensity in relation to entrepreneurial intentions. Previous studies on this relationship were criticised for using behavioural tasks with low ecological validity. Chapter 4 employs a task with high ecological validity: the "warm" version of the Columbia Card Task (CCT; Figner et al., 2009), which is commended in Chapter 3 for circumventing several of the methodological shortcomings that characterise the BART. It also uses a statistical model to address methodological shortcomings of the CCT itself. In addition, Chapter 4 explains and showcases how features from the CCT, which originates from the field of psychology, are beneficial for research in economics as well.

1.1.2 Part II: Brain Activity

Chapters 5, 6, and 7 form Part II of the dissertation, which looks into the challenges of measuring risk-taking propensity through brain activity, in particular using

electroencephalography (EEG). This is a relatively inexpensive technique that is often used in psychological research and that can detect changes in the brain's electrical activity at the millisecond level. First, Chapter 5 revisits the Columbia Card Task (CCT) that was used in Chapter 4, now examining changes in the EEG signal in response to winning and losing points in the task, under the assumption that these changes reflect RTP or related concepts like reward responsiveness. Pre-empting the findings, robust EEG signals are observed, but only weak correlations between these signals and self-reported and behavioural measures of the same constructs (including participants' behaviour on the CCT) are found – despite the use of a well-powered sample. Chapter 5 further identifies methodological shortcomings of the "hot" version of the CCT and offers solutions.

Chapter 6 can be seen as a conceptual replication of Chapter 5, following up on its null-findings by examining the (lack of) associations between self-reported, behavioural, and EEG operationalisations of impulsivity and related concepts (Study 1) and reward responsiveness and related concepts (Study 2) in similarly large, independent samples. In addition, given that research involving EEG is often characterised by much smaller sample sizes than those used in Chapters 5 and 6, Chapter 6 explores whether running the same analyses with smaller – more "authentic" – samples results in different findings. Taken together, weak correlations are observed in Chapter 6, particularly between different types of measurement (such as self-reports and EEG), in line with the findings from Chapter 5.

Finally, in a further attempt to understand the findings from Chapters 5 and 6, Chapter 7 examines selective reporting and statistical power in the EEG literature, contending that these factors may explain why existing research often reports strong and statistically significant findings, whereas Chapters 5 and 6 – using large samples and a broad range of measurements – do not. It does so through a large-scale literature review and meta-analysis focused on EEG research on addiction, which often reports substantial differences in the EEG signal of persons with versus without addiction as recorded during various behavioural tasks. Whereas Chapters 5 and 6 ultimately call into question the value of operationalisations using EEG, Chapter 7 does find support for them, though also showing that most EEG research is severely underpowered.

1.2 Contributions of the Dissertation to Science and Society

The contributions of this dissertation can be summarised by the following three proverbs and adages:

The proof of the pudding is in the eating. The value or quality of something cannot be judged from appearances and promises, but requires empirical testing and subsequent evaluation (i.e., one needs to eat the pudding before deciding whether it is any good). This resonates strongly with the scientific method, yet surprisingly, the behavioural and neurophysiological measures we employ to operationalise the constructs at the centre of our research are themselves often not thoroughly investigated. Several chapters in this dissertation address this omission. Specifically, Chapters 2 and particularly 3 do so by critically evaluating behaviour in the BART as an operationalisation of risk-taking propensity. This is vital as the BART is the most frequently used instrument for measuring risk-taking propensity in psychology (C. Wang et al., 2025). Chapters 4 and 5 contribute by looking at the qualities of the CCT: Chapter 4 applies advanced models to the CCT's data to solve shortcomings of behavioural tasks identified in Chapter 3; and Chapter 5 tests the suitability of the CCT for EEG research. Together, these chapters emphasise the need to critically evaluate the characteristics of measurement instruments for their intended purpose, before using them for that purpose on a larger scale. Finally, Chapter 7 "eats the pudding" on a different level: not that of individual instruments, but that of the published literature as a whole. At first sight, the EEG literature on addiction seems solid: many studies report statistically significant findings. Chapter 7 examines the evidential value and underlying power of these studies to see if their surface-level strength is legitimate.

A man with one watch knows what time it is; a man with two watches is never sure. With only one piece of information, no conflicts arise, but this creates a false sense of security: the information could be inaccurate, but there is no way to know whether it is. If one has multiple sources of information – like several different measurements of risk-taking propensity from the same participants – this can lead to confusion and doubt, especially if the measurements do not align, but it makes one aware of the potential for error and the importance of evaluating which measure (i.e., which watch) is most accurate. Chapters 5 and 6 of this dissertation help researchers be mindful of this dilemma. Chapter 5 examines whether RTP as operationalised with EEG correlates with behavioural and self-report operationalisations. On a larger scale, Chapter 6 compares self-report, behavioural, and EEG operationalisations of impulsivity (Study 1) and reward responsiveness (Study 2). Pre-empting the findings, most of the different measurements correlate poorly, supporting the idea that operationalisation is more complex than one might think. Which instrument should researchers choose if multiple instruments claim to measure the same concept, yet produce very different results? Chapters 5 and 6 try to interpret where these differences come from to help researchers make an informed decision about their choice of measurement(s), and to encourage researchers to be transparent about their motivation for specific choices.

Measure twice, cut once. Careful consideration and preliminary testing before commencing a project can help prevent costly mistakes. Once a carpenter has cut the wood, there is no undoing it. Similarly, a researcher's suboptimal choice in operationalisation cannot be undone after data collection has finished. In an era of publish-or-perish, it may not seem appealing to spend additional time pondering the selection of one's measurement instruments. However, the costs of not doing so can be large. The research on risk-taking propensity and related concepts often has strong societal implications. As illustrated in Section 1.1, it may help, for example, to identify who is more and who is less likely to develop addiction, to choose a certain career path, or to move away from where they are born. If researchers want to say something useful about such relevant relations, it is essential that RTP is measured well. After all, the conclusions we can draw from our research depend in large part on how well the concepts of interest are measured, with better measurement leading to more accurate conclusions. The chapters included in this dissertation encourage and help researchers to think critically about operationalisation, the measures they use, and the conclusions they draw. With that, this dissertation is an ode to fundamental research, but with clear implications for practice.

1.3 Author Contributions and Publication Status

The research presented in these chapters was predominantly conducted and written up independently by myself. However, valuable collaborative input from others was used as well for the majority of the chapters, which is acknowledged in the sections below. Separate from the content of this dissertation, I would like to acknowledge the Dutch Research Council for funding this research.

As for Part I of the dissertation, in Chapter 2 (published as De Groot & Thurik, 2018) Prof.dr. Roy Thurik participated in the conceptualisation of the study, and in addition provided feedback on the text. Data collection, data curation, and writing of the manuscript were done independently by myself. Prof.dr. Han Bleichrodt, Prof.dr. Kirsten Rohde, and Prof.dr. Peter Wakker provided feedback on an early version of this paper. Chapter 3 (published as De Groot, 2020) was single-authored, but benefitted from feedback by Prof.dr. Marco Lauriola, Prof.dr.ir. Jan van Ours, and Sander Wieman MSc. Finally, Chapter 4 (published as Dijkstra et al., 2022) was written in collaboration with Prof.dr. Niels Rietveld and Dr. Nienke Dijkstra. The research question was contrived by the authors together. The data were collected by master students (Denisa Debrecka MSc, Eline Hagenberg MSc, Marin Baelde MSc) under my supervision. I also cleaned and curated the data, and was responsible for the majority of the Methods section. Dr. Dijkstra wrote the Statistical Methodology

section, and was responsible for programming the analyses. Prof.dr. Rietveld directed the writing of the Introduction and Discussion sections. All authors contributed to the text in the Introduction, Results, and Discussion section.

With regard to Part II of this dissertation, Chapter 5 (published as De Groot & Van Strien, 2019) was written in collaboration with Prof.dr. Jan van Strien, who was involved in the conceptualisation of the study and who provided feedback on the text. Data collection, data curation, analysis, and writing of the manuscript were done independently by myself. Dr. Indy Bernoster and Sander Wieman MSc provided feedback on an early version of this paper. Chapter 6 (published as Bernoster et al., 2019) is the result of a collaboration with Dr. Indy Bernoster, using data collected by Dr. Wim Rietdijk (sample 1) and by Dr. Plato Leung, Dr. Indy Bernoster, and Marwan Aboul Magd MSc (sample 2). In this paper I was responsible for writing, while Dr. Bernoster performed the analyses. Prof.dr. Ingmar Franken, Prof.dr Matthias Wieser, and Prof.dr. Roy Thurik played a supervisory role. Dr. Maartje Luijten, Dr. Reshmi Marhe, and Dr. Ronald de Vlaming provided feedback on an early version of this paper. Finally, Chapter 7 was written in collaboration with Dr. Eric Slob, Dr. Oliver Lindemann, and Prof.dr. Jan van Strien. The conceptualisation of this study and the execution of the review and meta-analysis were performed by myself, including the literature search, data extraction, data curation, analysis, visualisation, and write-up. Dr. Slob reviewed the manuscript multiple times, providing valuable feedback, and aided in the calibration of the data extraction process. Dr. Lindemann and Prof.dr. Van Strien supervised the writing of this Chapter. Library liaison Judith Gulpers MA helped with creating the search queries and with locating difficult-to-find articles. At the time of writing, Chapter 7 is being prepared for journal submission.

The general introduction (Chapter 1) and discussion (Chapter 8) were written independently by myself, with feedback from Prof.dr. Niels Rietveld and Sander Wieman MSc. The dissertation was written without the use of (generative) AI.

PART I

Behavioural Tasks

Chapter 2

Disentangling Risk and Uncertainty: When Risk-Taking Measures are not About Risk

Abstract

Many studies claim to measure decision-making under risk by employing the Domain-Specific Risk-Taking (DOSPERT) scale, a self-report measure, or the Balloon Analogue Risk Task (BART), a behavioural task. However, these tasks do not measure decision-making under risk but decision-making under uncertainty, a related but distinct concept. The present commentary discusses both the theoretical and empirical basis of the distinction between uncertainty and risk from the viewpoint of several scientific disciplines and reports how many studies wrongfully employ the DOSPERT scale and BART as risk-taking measures. Importantly, we call for proper distinguishing between (tasks measuring) decision-making under uncertainty and decision-making under risk in psychology, and related fields. We believe this is vital as research has shown that people's attitudes, behaviour, and brain activity differ between both concepts, indicating that confusing the concepts may lead researchers to erroneous conclusions.

This chapter has been published as: De Groot, K., & Thurik, R. (2018). Disentangling risk and uncertainty: When risk-taking measures are not about risk. *Frontiers in Psychology, 9*, Article 2194. https://doi.org/10.3389/fpsyg.2018.02194

2.1 Introduction

Many studies claim to examine decision-making under risk, using a broad range of measures to indicate an individual's level of risk-taking. Well-known examples of these measures include the Domain-Specific Risk-Taking (DOSPERT, Blais & Weber, 2006) scale, a self-report measure, and the Balloon Analogue Risk Task (BART, Lejuez et al., 2002), a behavioural measure. However, despite including the word "risk" in their names, the DOSPERT scale and BART in fact do not deal with attitude towards risk. The concept they deal with is attitude towards uncertainty, a related but distinct phenomenon that is often contaminated with risk attitude in psychological literature.

2.2 The Theory Behind Decision-Making Under Uncertainty Versus Risk

In economics, the distinction between uncertainty and risk proposed by F. H. Knight (1921) has become classic and has been hardly contested. In the case of risk, the outcome is unknown, but the probability distribution governing that outcome is known. Uncertainty, on the other hand, is characterised by both an unknown outcome and an unknown probability distribution. In both cases, preferences are defined across chance distributions of outcomes. For risk, these chances are taken to be objective, whereas for uncertainty, they are subjective. Consider betting with a friend by rolling a die. If one rolls at least a four, one wins 30 Euros (or Pounds, Dollars, Yen, Republic Dataries, Bitcoins, etc.). If one rolls lower, one loses. If the die is unbiased, one's decision to accept the bet is taken with the knowledge that one has a 50 percent chance of winning and losing. This situation is characterised by risk. However, if the die has an unknown bias, the situation is characterised by uncertainty. The latter applies to all situations in which one knows that there is a chance of winning and losing but has no information on the exact distribution of these chances.

When laypersons talk about risk, they generally mean uncertainty, as the outcome probabilities are seldom known in everyday situations. In contrast to laypersons, scientists cannot afford to confound the concepts of risk and uncertainty. Contaminating these two concepts and hence not adhering to the uncertainty/risk (U/R) distinction is problematic, as this distinction has been supported by various studies showing that it is not only conceptually but also empirically valid. These studies primarily come from three scientific disciplines: economics, psychology, and neurobiology.

2.3 Empirical Support

2.3.1 Findings from Economics

Behavioural economic literature has supported the validity of the U/R distinction by showing that individuals are less sensitive to likelihood information in the case of uncertainty compared to risk: likelihood insensitivity decreases with more information (Abdellaoui et al., 2005; Abdellaoui et al., 2011; Baillon et al., 2012; Baillon et al., 2013, 2017; Kahn & Sarin, 1988; Kahneman & Tversky, 1979; Kilka & Weber, 2001). This phenomenon can be illustrated with the use of a fictional lottery. In the lottery, the difference between a win probability of 0 and 0.1 is substantial, for it reflects the difference between no chance versus a chance. The difference between a probability of 0.9 and 1 is also large, for the latter means a certain win. By contrast, the difference between 0.3 and 0.4, or between 0.6 and 0.7 seems small. If one is not even sure whether these 0.3 and 0.4 (or 0.6 and 0.7) probabilities are accurate (uncertainty), it is plausible that one will treat them as equivalent. Hence, the less information individuals have (uncertainty vs. risk), the more they fail to sufficiently discriminate between different levels of likelihood (Baillon et al., 2012; Baillon, 2015; Tversky & Kahneman, 1992).

In addition, the behavioural economic literature shows aversion towards uncertain compared to risky choices (something referred to as ambiguity aversion): individuals prefer known probabilities over unknown probabilities, even if the known probability is low and the unknown probability could be a guaranteed win (Ellsberg, 1961). In the original experiments, this phenomenon is illustrated with urns. Imagine there are two urns: the "known" urn, with 50 red and 50 black balls, and the "unknown" urn, which contains 100 balls that are red or black in an unknown proportion. Winning is achieved by drawing a red ball. Which urn do people want to draw from? When asked this question, most people opt for the known urn. However, if they win when drawing a black ball, they also opt for the known urn. This decision contradicts the notion of probability: people act as if the chance of drawing a red ball from the unknown urn is less than 50 percent, but also as if the chance of drawing a black ball from that same urn is less than 50 percent. This so-called Ellsberg paradox illustrates our initial statement: when asked to choose, individuals prefer risk over uncertainty.

2.3.2 Findings from Psychology

Next to behavioural economic literature, psychological literature also supports the empirical distinction between uncertainty and risk. Buckert et al. (2014), for example, show that the cortisol response to stress impacts decision-making under risk, but not under uncertainty. In addition, several studies show how risk and uncertainty are differentially impaired in a broad range of (neuro)psychological disorders. For example, decision-making under uncertainty but not under risk has shown to be impaired in patients who have undergone unilateral temporal lobe surgery (Bonatti et al., 2009), patients with gambling problems (Brevers et al., 2012), breast cancer patients receiving adjuvant chemotherapy (X. Chen, Zhu, et al., 2013), patients with obsessive-compulsive disorder (H. W. Kim et al., 2015; Starcke et al., 2009, 2010; L. Zhang, Dong, Ji, Tao, et al., 2015; L. Zhang, Dong, Ji, Zhu, et al., 2015), and patients with pathological buying issues (Trotzke et al., 2015). The same dissociation holds for normal ageing (Zamarian et al., 2008). The situation, however, is reversed for patients with Parkinson's disease, who are differentially impaired in decision-making under risk, but not in decision-making under uncertainty (Euteneuer et al., 2009), which could be explained by the notion that decision-making under risk depends more on executive functioning than does decision-making under uncertainty (Brand et al., 2006). These dissociations provide further support for the empirical distinction between uncertainty and risk.

2.3.3 Findings from Neurobiology

Extending the psychological literature, several studies from the field of neurobiology indicate that risk and uncertainty are differentially coded in the brain. Currently, two hypotheses exist outlining how dealing with uncertainty versus risk differs neurobiologically (Schultz et al., 2008). First, both situations may recruit different brain systems, resulting in double dissociations. This hypothesis is supported by studies showing that risk recruits the orbitofrontal cortex, striatum, insula, and (posterior) parietal cortex, whereas uncertainty recruits the amygdala and parts of the frontal cortex such as the inferior frontal gyrus, and the (dorsal) lateral prefrontal cortex (Bach et al., 2009; Huettel et al., 2006; Krain et al., 2006; Platt & Huettel, 2008; Schultz et al., 2008). Second, risk and uncertainty may recruit a common brain mechanism but to different degrees, showing stronger responses to either ambiguous or risky choices. This idea is backed up by studies showing that the activity in some brain structures, such as the orbitofrontal cortex and amygdala, is positively related to the level of uncertainty in a task, whereas activity in the striatal system is negatively correlated (Hsu et al., 2005; Levy et al., 2010; Platt & Huettel, 2008; Schultz et al., 2008).

These findings support a graded rather than an all-or-nothing difference between how uncertainty and risk are neurobiologically coded. There is no conclusive evidence yet on whether uncertainty and risk are mutually exclusive or graded represented in the brain. However, regardless of which hypothesis is supported, uncertainty and risk can be said to differ from each other even at a biological level, which is an important signal of the essential difference between uncertainty and risk.

2.4 The DOSPERT Scale and BART

The discussed behavioural economic, psychological, and neurobiological studies demonstrate how uncertainty and risk differ not only on a theoretical basis, but also empirically. This emphasises the need to properly distinguish between the two concepts in research. However, as discussed, two paradigms that claim to measure decision-making under risk in fact deal with uncertainty: the DOSPERT scale and the BART. For the DOSPERT scale, items include "going camping in the wilderness", "drinking heavily at a social function", and "investing 10 percent of your annual income in a new business venture". For all these items and the others, the outcome distribution is unclear. There is a level of uncertainty, but since the probability distribution of the described situations is not known, this cannot be qualified as risk in the Knightian sense. The same is true for the BART. In the BART, individuals pump up a balloon that can explode at any time. Since the probability distribution of explosions is unknown to the participant ("participants were given no detailed information about the probability of an explosion", Lejuez et al., 2002, p. 77), there is again a level of uncertainty, which cannot be qualified as risk. For the DOSPERT scale, the uncertainty is bi-directional: both the researcher and the participant are ignorant of the probability distribution. For the BART, the uncertainty is one-directional: from the viewpoint of the researcher, the risk is known, as he or she knows how the probability distribution of the task has been programmed. The participant, on the other hand, is not given any information about this.

Although the DOSPERT scale and BART are both clear examples of uncertainty measures, studies applying them do generally not acknowledge that the measures deal with decision-making under uncertainty instead of risk. To examine how pervasive this mistake is, a literature review was conducted using Scopus across November, 2017. The first search parameter concerned the full name of both tasks as mentioned in the article title, abstract, or keywords as indexed in Scopus. This resulted in 17 articles on the DOSPERT scale, and 289 articles on the BART. These 306 articles were all read and classified into four categories. An article was categorised as recognising the uncertainty nature of the DOSPERT scale or

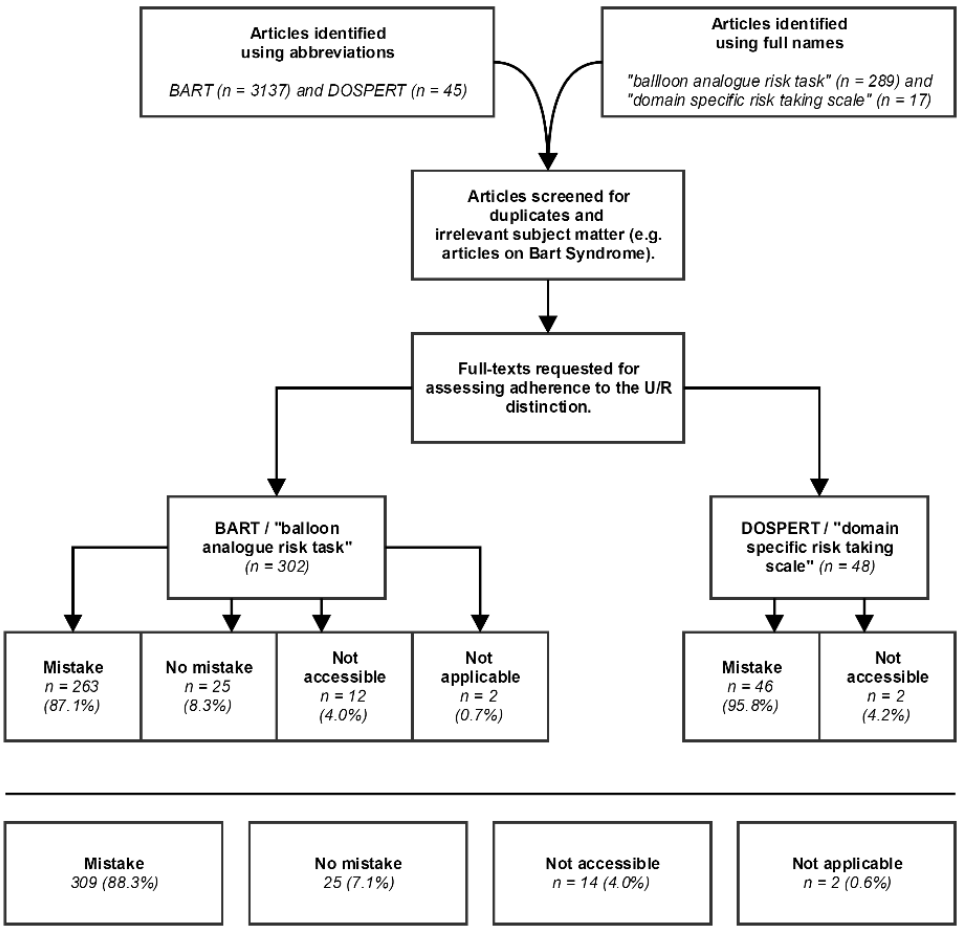
BART if the article explicitly mentioned that the DOSPERT scale or BART measures decision-making under uncertainty (instead of risk) because the probabilities relevant for the task are unknown to the participant; or if the article seemed to implicitly understand the difference between decision-making under uncertainty and risk, for example by discussing the conceptual difference between known versus unknown probabilities in relation to the experimental paradigms. However, articles were not included in this category if they only stated that probabilities in the task were unknown (which is simply a task characteristic), without relating this to uncertainty or risk. Articles were also not included if they correctly stated that the task measured uncertainty but mentioned invalid reasons for this (for example because "the outcome is unknown", which is a characteristic of both risk and uncertainty). Articles were classified as not recognising the U/R distinction if they did not seem to be aware of the distinction between the two concepts in relation to the DOSPERT scale or BART. This was for example reflected by consistently using the concepts intertwined without discussing the difference; explicitly stating that uncertainty and risk are equal; discussing uncertainty and risk but not relating this to the DOSPERT scale or BART; or stating that the DOSPERT scale or BART measures decision-making under risk, while not mentioning uncertainty (or related concepts such as ambiguity, or unknown probabilities) at all. The two remaining categories concerned "not accessible" (if the full text of an article could not be accessed) and "not applicable" (if the article did not say anything concerning either uncertainty or risk, for instance because the DOSPERT scale or BART was used for measuring a different construct, such as impulsivity). The second search parameter concerned the abbreviation of both measures, again as mentioned in the article title, abstract, or keywords as indexed in Scopus. This resulted in 45 articles on the DOSPERT scale, and 3137 on the BART. For these articles, the abstracts were examined, and irrelevant articles (e.g., articles discussing Bart Syndrome) were removed from the search findings. In addition, articles that were already included based on the first search parameters were also removed. The remaining articles were again all read and classified according to the abovementioned criteria. The final categorisation consisted of 48 articles on the DOSPERT scale, and 302 articles on the BART. The included articles were solely identified via the search parameters; no additional method of including articles was employed.

The results from the categorisation can be found in Figure 2.1. The findings were in line with our proposition that most studies do not adhere to the U/R distinction and do not correctly identify the DOSPERT scale and BART as uncertainty measures. This was also true for several studies published in this journal. Overall in the literature, only 7.1% of articles correctly adhered to the U/R distinction in relation to the DOSPERT scale and BART.

The Knightian distinction between uncertainty and risk was not adhered to by 88.3% of articles.

Figure 2.1

Summary of the Categorisation Process and Outcomes



The classification of all examined articles was archived in an online repository and can be accessed via a weblink that is available from the authors upon request. In addition to the basic classification, the repository holds information indicating whether the authors thought a certain classificatory decision was up for debate. This was the case for 12.3% of articles. For these articles, the arguments on which the final decision was based are also reported.

The most common arguments reported are discussing the U/R distinction in relation to other tasks but not to the DOSPERT scale or BART; correctly stating that the DOSPERT scale or BART measures decision-making under uncertainty but reporting invalid arguments for this claim; and applying a Bayesian learning paradigm to the BART that quantifies the decision-maker's changing uncertainty about the chances of the balloon exploding, but that does not explicitly say anything regarding the U/R distinction. All articles characterised by these arguments were categorised as not adhering to the U/R distinction, which resulted in a relatively conservative classification. However, the percentage of articles not properly adhering to the U/R distinction in relation to the DOSPERT scale and BART remains high (namely 78.3%) if all articles now classified as "mistake" + "up for debate" were classified as properly adhering to the distinction.

2.5 The Importance of Distinguishing Between Decision-Making Under Uncertainty Versus Risk

From the literature review, we conclude that not properly adhering to the U/R distinction is a widespread problem. Most articles do not even mention the distinction, let alone correctly identify the DOSPERT scale and BART as measuring decision-making under uncertainty instead of risk. However, the present finding that 88.3% of articles does not adhere to the Knightian distinction between uncertainty and risk in relation to the DOSPERT scale or BART does not necessarily mean that 88.3% of authors are unaware of the distinction. In fact, we believe that most researchers understand the conceptual distinction between uncertainty and risk, but do not explicitly report on this in their articles. This absence of uncertainty/risk information in articles could simply be the result of common practises within the field. In psychology, terms such as "risk- taking" and "riskiness" often signify not only a known chance but also a directional effect: a high chance of loss. Uncertainty does not allow for a similar directional connotation, which may result in using the term less frequently. Furthermore, if only few studies explicitly distinguish between uncertainty and risk, this becomes the default within a field, leading others to also not report on this distinction even though they may have considered it when designing their study. This is reflected in the observation that articles that adhere to the U/R distinction are not only scarce but are also not consistently referred to in the literature. This contributes to the contamination of both concepts that currently dominates the literature, and makes research prone to confusion, especially when crossing disciplinary lines.

Being aware of the distinction between uncertainty and risk and applying this knowledge in scientific writings not only is of great importance for scientific coherence but

also has meaningful practical implications for government and business because the rules used for decision-making under risk differ from those used for decision-making under uncertainty. As an example, Angner (2012) discusses the regulation of new and unstudied chemical substances. There is little hard data on them, but there is some probability that they will turn out to be toxic. If a policy maker would argue that the decision at hand concerns uncertainty, he or she would have to decide that the new chemical should be banned or heavily regulated until its safety can be established. Speaking in behavioural economic terms, either the minimax (minimising the maximum amount of deaths) or the maximin (maximising the minimum amount of profit) criterion applies in this situation. However, if the policy maker argues that one can and must assign probabilities to all outcomes, he or she faces a choice under risk, and will probably permit the use of the new chemical because the probability that it will turn out to be truly dangerous is low (the expected utility, the alternative with the greatest amount of utility in the long run, is highest for permitting the use). This example shows that decision-making under uncertainty versus risk results in different responses. Therefore, whether a decision is treated as a choice under uncertainty or under risk can have real consequences.

2.6 What Should Researchers Do?

The aim of the present commentary is not to scold researchers from fields such as psychology for not using terminology and conventions used in economics. We do, however, encourage researchers to properly distinguish between uncertainty and risk. We believe that the majority of researchers are in fact already aware of this distinction, even though this is not always reflected in their writings. Moreover, the aim of our commentary is not to take credit for the idea that the DOSPERT scale and BART do not measure attitude towards risk but rather towards uncertainty. In fact, we mention several previous studies that explicitly contribute to this view by providing empirical support for the conceptual distinction. Furthermore, studies adhering to the U/R distinction in relation to the DOSPERT scale or BART can be found in the repository.

The aim of the present commentary is to unite previous research, and to make researchers explicitly aware of the distinction between uncertainty and risk. In addition, the aim is to advise researchers on what tasks (not) to use. For example, a self-report measure probing pure risk-taking should include clear indicators of the probability distribution underlying the outcomes of the described activities. Furthermore, the BART should not be used for measuring pure risk-taking. Instead, different behavioural tasks such as the Cambridge Gambling Task (CGT, Rogers et al., 1999), Game of Dice Task (GDT, Brand et

al., 2005), and Columbia Card Task (CCT, Figner et al., 2009) should be used when aiming to measure decision-making under risk. It should be noted though that in real life the chances are almost always unknown, which means that risk has more theoretical than practical importance. Therefore, there is certainly merit to using the DOSPERT scale and BART, especially considering their good external validity. In fact, we could even speculate that this good external validity can be explained by the fact that these tasks measure decision-making under uncertainty (and not risk), which corresponds well to the structure of decision-making in real life. Looking even closer at how decision-making in real life is accomplished, it appears that the distinction between uncertainty and risk is continuous rather than binary. In many cases, individuals have some estimate of the involved probabilities, which develops as they move further along in the decision-making process and receive feedback by sampling the environment. This development is mirrored, for example, in the Iowa Gambling Task (IGT, Bechara et al., 1994), in which participants learn the probabilities associated with card decks as they progress through the task. It could even be argued that the BART is characterised by a learning process as well, which is reflected by studies applying learning models to the task. However, regardless of what (version of a) task is used, it is important to be explicitly aware of what it is measuring: decision-making under uncertainty (BART and DOSPERT scale), decision-making under risk (CGT, GDT, and CCT), or a gradual shift from decision-making under uncertainty to decision-making under risk (IGT and possibly BART). This way, the used nomenclature can stay pure and help readers identify what concepts are examined in a particular study.

Scientists are expected to outperform laypersons in properly distinguishing between concepts. Considering the fluidity of interdisciplinary research, it is pertinent to employ a sole and clear-cut definition of concepts across fields. This is particularly important if concepts have been shown to differ both theoretically and empirically. The present commentary calls for distinguishing uncertainty and risk in the field of psychology and related fields where decision-making under uncertainty and decision-making under risk play an important role, such as neuroeconomics. This will help in using tasks that actually measure the concept one is interested in measuring, which will certainly aid in finding true relationships.

Chapter 3

Burst Beliefs – Methodological Problems in the Balloon Analogue Risk Task and Implications for its Use

Abstract

Studies in the field of psychology often employ (computerised) behavioural tasks, aimed at mimicking real-world situations that elicit certain actions in participants. Such tasks are for example used to study risk propensity, a trait-like tendency towards taking or avoiding risk. One of the most popular tasks for gauging risk propensity is the Balloon Analogue Risk Task (Lejuez et al., 2002), which has been shown to relate well to self-reported risk-taking and to real-world risk behaviours. However, despite its popularity and qualities, the BART has several methodological shortcomings, most of which have been reported before, but none of which are widely known. In the present paper, four such problems are explained and elaborated on: a lack of clarity as to whether decisions are characterised by uncertainty or risk; censoring of observations; confounding of risk and expected value; and poor decomposability into adaptive and maladaptive risk behaviour. Furthermore, for every problem, a range of possible solutions is discussed, which overall can be divided into three categories: using a different, more informative outcome index than the standard average pump score; modifying one or more task elements; or using a different task, either an alternative risk-taking task (sequential or otherwise), or a custom-made instrument. It is important to make use of these solutions, as applying the BART without accounting for its shortcomings may lead to interpretational problems, including false-positive and false-negative results. Depending on the research aims of a given study, certain shortcomings are more pressing than others, indicating the (type of) solutions most needed. By combining solutions and openly discussing shortcomings, researchers may be able to modify the BART in such a way that it can operationalise risk propensity without substantial methodological problems.

This chapter has been published as: De Groot, K. (2020). Burst beliefs – Methodological problems in the Balloon Analogue Risk Task and implications for its use. *Journal of Trial and Error, 1*, 43-51. https://doi.org/10.36850/mr1

3.1 Introduction

To a large extent, psychological science rests on the promises of operationalisation: defining fuzzy concepts as measurable variables, or in other words, changing conceptual variables into operational ones (Shuttleworth, 2008). This process is imperative because most concepts researchers hypothesise about are not straightforwardly quantifiable. By defining how a concept is measured, operationalisation allows hypotheses to take a falsifiable format and enables us to replicate findings. In a way, operationalisations are arbitrary, as concepts can be defined and thus measured in numerous ways – none of which are surely "right". Nonetheless, some measures may be more suitable than others.

A notable example of a concept that can be operationalised in various ways is risk-taking (Lauriola & Weller, 2018), which has an important place in clinical, cognitive, and developmental psychology, as well as in the fields of criminology, economics, and management. One way risk-taking is operationalised in these fields is through self-report measures, such as the Domain-Specific Risk-Taking (DOSPERT) scale (Blais & Weber, 2006) and the Financial Risk Tolerance assessment (Grable, 1999). Another way is through computerised behavioural tasks, like the Iowa Gambling Task (Bechara et al., 1994), the Cambridge Gambling Task (Rogers et al., 1999), the Game of Dice Task (Brand et al., 2005), the Balloon Analogue Risk Task (Lejuez et al., 2002), and the more recent but already widely used Columbia Card Task (Figner et al., 2009). Importantly, the quality of a study largely depends on the degree to which its operational measures reflect the underlying concept; in this case, one's disposition towards risk-taking. If a task is a poor proxy for a concept or is subject to methodological or interpretational problems, any data resulting from it are of limited value to our understanding of the concept. In this regard, several studies have challenged the operationalisation ability of the most-cited risk task, the Iowa Gambling Task (see e.g. Brand et al., 2006; Buelow & Suhr, 2009; Figner et al., 2009; Maia & McClelland, 2004). The Balloon Analogue Risk Task, which is the second-most cited, may yet suffer from even more severe issues, hindering its ability to operationalise risk-taking. While some individual issues have been reported in previous publications, no literature so far has discussed these collectively. The present commentary aspires to fill this gap.

3.2 The Balloon Analogue Risk Task

In the Balloon Analogue Risk Task, or BART for short, participants are presented with a computer screen showing a small balloon and a pump. They are told that every time they click the pump, the balloon expands, and a fixed amount of money (5 cents) is added to a

temporary bank. Every pump also increases the chance of the balloon exploding (marked by a "pop" sound from the computer), resulting in losing all money in the temporary bank for that particular balloon (trial). The point at which a balloon explodes varies across trials, ranging from the first pump to the point where the balloon fills the entire screen. Participants can decide to stop pumping the balloon at any point during a trial by clicking the "collect" button (left in Figure 3.1), which transfers the money accumulated in their temporary bank to their permanent one, while a slot machine sound is played. Once a balloon explodes or once participants cash a balloon's proceeds, the trial ends, and a new, uninflated, balloon appears.

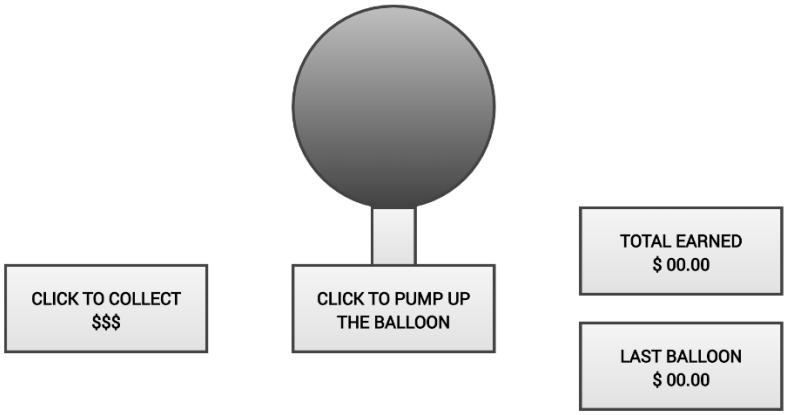
In the original study by Lejuez et al. (2002), participants were informed that they would complete 90 balloons: 30 orange, 30 yellow, and 30 blue ones. Unbeknownst to participants, differently coloured balloons had a different chance of exploding. The probability distribution governing their explosion points consisted of an array of n numbers from which on every pump a random number was drawn without replacement. If a 1 was drawn, the balloon exploded. Thus, the probability p of the balloon exploding on the first pump was $1/n$, and the probability of it exploding on pump i (given no prior explosion) was $p_i = \frac{1}{n-i+1}$. For orange balloons, the array ranged from 1 to 8 (hence $p_1 = \frac{1}{8-1+1} = 1/8$), for yellow balloons from 1 to 32 ($p_1 = \frac{1}{32-1+1} = 1/32$), and for blue ones from 1 to 128 ($p_1 = \frac{1}{128-1+1} = 1/128$). Their average explosion points were respectively 4, 16, and 64, with the same (randomly generated) sets of explosion points being used across all participants to limit extraneous variability. Neither the ranges nor the average explosion points were communicated to participants.

The BART's design is intended to reflect naturalistic decision-making, in which taking more risk generally increases the odds of encountering a loss. This sort of decision-making tends to be emotionally engaging, instigating a sense of increasing tension as the balloon increases in size (Schonberg et al., 2011). In support of the BART's validity, Lejuez et al. (2002) showed that the average number of times participants pumped the blue balloon significantly correlated with scores on risk-related constructs (sensation seeking, impulsivity) and with real-world risk behaviours, such as polydrug use, gambling, unsafe sex, and stealing. The orange and yellow pumps were originally not examined with respect to risk-related constructs, as their narrow ranges of outcome values (1-8 and 1-32) are less suited for capturing individual differences. Instead, their average pump numbers were analysed together with those of the blue balloons to show that the number of times participants choose to pump is sensitive to the probability of exploding. Overall, the data

showed the BART to have "particular promise as a behavioural index of risk-taking" (Lejuez et al., 2002, p. 82). As would be expected based on this conclusion, the BART (particularly its blue balloon) became a popular instrument for gauging individuals' propensity for risk-taking, with inconsistent findings being attributed to factors like sampling variability and inadequate statistical power (Lauriola et al., 2014), rather than problems inherent to the BART. However, several authors have argued that such problems exist (De Groot & Thurik, 2018; Gu et al., 2018; Schmidt et al., 2019; Schonberg et al., 2011), and that they limit the BART's ability to measure one's propensity for taking risk. The key problems that characterise the BART are 1) a lack of clarity as to whether decisions are characterised by uncertainty or risk, 2) censoring of observations, 3) confounding of risk and expected value, and 4) poor decomposability into adaptive and maladaptive risk behaviour.

Figure 3.1

Set-Up of the Original Balloon Analogue Risk Task as Described by Lejuez et al. (2002)



3.2.1 Risk or Uncertainty?

In economic theories of decision-making, a key distinction is that between uncertainty and risk, which is often accredited to F. H. Knight (1921), and was introduced to psychological thinking in a seminal paper by Edwards (1954) that lies at the origin of behavioural decision theory. When deciding under the condition of risk, the probabilities associated with the possible outcomes are *known*. When deciding under uncertainty (which some authors call ambiguity), this probability distribution is *unknown*.

For F. H. Knight (1921), this distinction was not only of theoretical but of practical importance as well. According to him, uncertainty – not risk – was the main driver of entrepreneurial success, as only people who recognise hidden opportunities can seize them and profit from them. Since then, the empirical relevance of the uncertainty-risk distinction has been confirmed in various fields of research. In economics, Ellsberg (1961) showed that individuals prefer risk over uncertainty, even if the known probabilities are unfavourable and the uncertain option could be a guaranteed win. In psychology, studies showed that uncertain and risky decisions involve different mental processes, as risk allows for statistical thinking (to optimise) but uncertainty involves heuristics (to satisfice) (Volz & Gigerenzer, 2012). In line with this, decision-making under risk is thought to depend more on executive function (such as categorisation and cognitive flexibility) for which the dorsolateral prefrontal cortex is important, whereas decision-making under uncertainty hinges on emotional processes (such as somatic feedback), which are more associated with the ventromedial prefrontal cortex and the amygdala (Brand et al., 2006). This may explain why patients with executive deficits, such as those with Parkinson's disease, have difficulty deciding under risk but have no trouble deciding under uncertainty (Euteneuer et al., 2009), whereas persons with obsessive-compulsive disorder, for example, show the opposite pattern (Starcke et al., 2009, 2010).

Given that uncertainty and risk differ theoretically and empirically, it is imperative for researchers to know the conditions under which participants decide. Unfortunately, despite the word "risk" in its name, these conditions are not straightforward in the BART. Since participants are never given "detailed information about the probability of an explosion" (Lejuez et al., 2002, p. 77), we can assume that at least during early trials, they decide under uncertainty (Bishara et al., 2009; De Groot & Thurik, 2018; Schonberg et al., 2011). As they move further along in the task and "sample the distribution" by pumping balloons and observing their outcomes, they get a better sense of the probabilities, which gradually moves their decisions in the direction of risk. Although not studied in the BART itself, such a shift has been shown for the Iowa Gambling Task, where performance in early trials does not correlate with that in later trials nor with executive function, indicating that people first decide under uncertainty and later under risk (Brand et al., 2006; Brand et al., 2007). While this effect may not be as strong in the BART, studies do show better performance in later compared to early trials, suggesting that participants indeed get a better grasp of the probability distribution over time (De Groot & Van Strien, 2019; Lejuez et al., 2002).[1]

[1] The relevant data collected by De Groot and Van Strien (2019) on per-block averages is not reported in the published report but will be shared upon request.

The BART's transition from uncertainty towards risk is problematic for several reasons. First, it is unclear *when* exactly this shift transpires, making it difficult to determine whether a decision in a given trial is made under uncertainty, risk, or something in between. Second, the point where decisions shift from uncertainty to risk is likely to differ between individuals, and is dependent on task characteristics (Brand et al., 2006; Brand et al., 2007). Third, the shift implies that the BART imposes learning demands, which could inadvertently impact participants' outcomes on the task, with those capable of updating their knowledge of the probabilities performing better than those who have difficulty doing so. Fourth, once participants manage to derive the task's probabilities, subsequent decisions are not characterised by what is usually considered risk. Contrary to decisions in which probabilities are explicitly described ("a priori" probabilities), probabilities in the BART are derived from experience. Since such probabilities depend on factors like sampling variability and one's memory of previous events, decision-makers treat experience-based probability differently, which is called the description-experience gap (Hau et al., 2008; Rakow & Newell, 2010). Most notably, when deciding based on experience, people do not act in accordance with prospect theory, but instead, underweight rare events and overweight common encounters. As people have more and more encounters (e.g., trials), their experiences will approach the precision of a priori probabilities, though in practice this is difficult to attain (F. H. Knight, 1921).

To address the inability of the BART to differentiate between complete uncertainty, experience-based risk, and description-based risk, several approaches may be used. One option is to apply a model to the BART's data that allows for participants learning through experience. An early example is a model by Wallsten et al. (2005) in which decision-makers update their probabilities from trial to trial, and continually re-evaluate their options. Alternatively, one could use a different task, in which decisions are either all characterised by uncertainty or risk, or which includes a well-understood shift between the two. Tasks that involve only uncertain decision-making are rather difficult to design, as they require participants to be ignorant of probability-related information and *remain* ignorant of that as well – automatically disqualifying tasks that have a learning curve. Tasks involving only decisions made under (a priori) risk are much more common and include the Cambridge Gambling Task, the Game of Dice Task, and the Columbia Card Task, the latter of which resembles the BART's dynamic, affective nature (Schonberg et al., 2011). Finally, a known shift from uncertainty to (experience-based) risk can be found in the Iowa Gambling Task. This task's shift, while not *fully* understood, has been studied more thoroughly than that in the BART.

3.2.2 Censored Observations

Statistical censoring refers to a condition in which the value of an observation is unknown because it is beyond a certain limit. This limit can exist by design, which is common in survival analysis. If a study on a surgical intervention follows patients for up to 10 years, the longevity scores of those who live past this term are censored, as their longevity is *at least* 10 (Young & McCoy, 2019). Censoring can also result from limits on what an instrument can reliably measure. For example, the full IQ score of the Wechsler Adult Intelligence Scale ranges from 40 to 160 (Sattler & Ryan, 2009), meaning that IQ scores of people performing either extremely poorly or extremely well are cut off at these boundaries and are thus censored.

In the BART, censoring (by design) occurs if a participant is stopped from taking more risk in a given trial, because the balloon they are pumping explodes, forcing the trial to end. Since such a trial ends prematurely, the number of times the participant pumped the balloon does not necessarily reflect the risk they were willing to take, meaning their risk propensity is censored. This is problematic for various reasons. First, including these censored trials biases the average number of pumps downwards (especially for high-risk takers), underestimating participants' willingness to take risks (Dijkstra et al., 2020; Pleskac et al., 2008). Likewise, the between-subjects variability across these averages is reduced (Lejuez et al., 2002). Overall, the (unadjusted) average number of pumps is an ill-suited operationalisation of risk propensity.

As censoring affects all sequential risk-taking tasks like the BART (involving multiple decisions per trial) and various other research paradigms, like survival analysis, several solutions have been proposed. In the paper introducing the BART, Lejuez et al. (2002) suggest computing an adjusted pump average using only trials in which participants stopped voluntarily, that is, in which the balloon did not burst. However, by omitting explosion trials, censored observations are essentially treated as randomly missing, which is inaccurate (Pleskac et al., 2008). The more risk someone takes, the more likely it is that the balloon bursts, and that the trial forcibly ends. The termination of trials is therefore not independent from participants' behaviour. As a result, Lejuez et al.'s adjusted score tends to discard trials in which participants take a lot of risk. This causes the average number of pumps to be biased downwards, similar to the unadjusted score, but to a lesser extent.

To circumvent the problem of censoring, Pleskac et al. (2008) developed an automatic response version of the BART. Contrary to the standard BART, in which participants inflate a balloon one pump at a time, the automatic BART lets them indicate their intended number of pumps beforehand. The balloon then inflates to the corresponding size, or until it bursts.

This procedure allows for an unbiased statistic of risk propensity, as the intended number of pumps is now observable in all trials (Pleskac et al., 2008). However, it increases the time between decision and outcome, which may make decisions less emotional (impulsive) and more cognitive (planned) (Pleskac et al., 2008), and may reduce the salience of the outcomes. These effects, in turn, can affect participants' risk-taking (Young & McCoy, 2019). In contrast, however, a study using the Bomb Risk Elicitation Task (Crosetto & Filippin, 2013), another risk task that uses delayed explosions to circumvent censoring, found that introducing such delays did not impact risk-taking.

Another solution to censoring is using a rigged task (Slovic, 1966). Participants are then told that failure can occur at any moment (in the BART, at any pump), but actually, it is set to occur at the last possible choice. Hence, participants can always stop voluntarily, and no scores are censored. To uphold credibility, "mock" trials are added, in which failure is set to occur early on. Deciding on the number and timing of mock trials, however, is a challenge. Since behaviour in a trial is affected by previous outcomes, experiencing (too) few failures could increase risk-taking (De Groot & Van Strien, 2019; Dijkstra et al., 2020). Therefore, rigged tasks should be designed such that they produce failure rates similar to non-rigged tasks and should take into account that failure rates differ between participants too. However, research on the Columbia Card Task, another sequential risk-taking task, shows that this is often not the case (De Groot & Van Strien, 2019).

A final remedy, which addresses the bias but leaves the BART unchanged, is to apply a statistical model to the resulting data that explicitly incorporates censored behaviour. Such models consider *all* observed data, using the censored trials as lower bounds in determining a participant's actual risk propensity. Some of them employ Bayesian (generalised) linear mixed-effects regression (Weller et al., 2019; Young & McCoy, 2019); others use maximum likelihood estimation, adding a cumulative distribution function to the likelihood function to account for censoring (Dijkstra et al., 2020; Tobin, 1958). Such models perform significantly better (i.e., have less biased predictions) than those that do not account for censoring. However, as is the case for all statistical models, their soundness hinges on the validity of their underlying assumptions (Schafer & Graham, 2002), such as that of normality, whose violation not all models are robust against (Powell, 1984).

3.2.3 Confounding and Decomposability

The BART was designed to resemble real-world risk situations, where taking modest risk is generally advantageous, but taking excessive risk is increasingly unfavourable (Lejuez et al., 2002; Wallsten et al., 2005). Within a trial, every successful pump earns

participants 5 cents, which are added to their temporary bank. As the amount accumulated in the bank grows, the relative gain of taking additional risk decreases, while the potential loss in case of an explosion increases. Additionally, the probability of the balloon exploding increases with every pump: from $1/128$ on the first to $1/127$ on the second, and so on.

This combination of characteristics makes that the task's structure entails a serious problem. Since both the balloon value (the amount collected in the temporary bank) and the explosion probability increase with every pump, the expected value of inflating the balloon – the product of the success chance and the reward, minus the product of the explosion chance and the balloon value – changes across a trial (Schmidt et al., 2019). This change is illustrated in Table 3.1. Early in a trial, the expected value of the pump is positive, so taking additional risk is advantageous. This prospect changes halfway when the expected value turns negative, making additional pumps unfavourable (Lejuez et al., 2002). Due to the expected value changing with each decision, it is *confounded* with risk (defined as the variability of the possible outcomes), which varies across decisions by design. Although such confounding can happen in real-life decision-making, it is not desirable in a controlled scientific environment: it makes it difficult to measure participants' risk propensity, as both risk and expected value may influence their decisions. The extent to which individuals are, for example, risk-seeking, can therefore not be determined, because this would require showing a preference for higher variance payoffs, holding expected value constant (Schonberg et al., 2011).

This confounding demonstrates that the BART's main observable outcome – the number of pumps participants press – cannot be interpreted as a straightforward indicator of risk propensity. Like many behavioural tasks, the BART supposedly gauges a single cognitive construct, but it manipulates various other, potentially confounding constructs as well (Schonberg et al., 2011). Expected value is an example of such a construct. As a result, the single score provided by the BART cannot easily be decomposed to identify the cognitive or neural mechanisms involved in the pump decisions. Studying the risk-taking process in isolation using the BART is therefore not possible.

Table 3.1

Changing Balloon Values, Explosion and Success Chances, and Expected Values Across Balloon Pumps

Pump Number	Balloon Value Before Pump	Balloon Value After Pump	Chance of Explosion	Chance of Success	Expected Value of Current Pump	Expected Value of All Remaining Pumps
(A)	(B)	(C)	(D)	(E)	(F)	(G)
1	€ -	€ 0.05	0.00781	0.99219	€ 0.04961	€ 1.60000
2	€ 0.05	€ 0.10	0.00787	0.99213	€ 0.04921	€ 1.56260
3	€ 0.10	€ 0.15	0.00794	0.99206	€ 0.04881	€ 1.52540
4	€ 0.15	€ 0.20	0.00800	0.99200	€ 0.04840	€ 1.48840
5	€ 0.20	€ 0.25	0.00806	0.99194	€ 0.04798	€ 1.45161
(…)						
62	€ 3.05	€ 3.10	0.01493	0.98507	€ 0.00373	€ 0.00672
63	€ 3.10	€ 3.15	0.01515	0.98485	€ 0.00227	€ 0.00303
64	€ 3.15	€ 3.20	0.01538	0.98462	€ 0.00077	€ 0.00077
65	€ 3.20	€ 3.25	0.01563	0.98438	€ -0.00078	€ -0.00078
66	€ 3.25	€ 3.30	0.01587	0.98413	€ -0.00238	€ -0.00238
(…)						
124	€ 6.15	€ 6.20	0.20000	0.80000	€ -1.19000	€ -1.19000
125	€ 6.20	€ 6.25	0.25000	0.75000	€ -1.51250	€ -1.51250
126	€ 6.25	€ 6.30	0.33333	0.66667	€ -2.05000	€ -2.05000
127	€ 6.30	€ 6.35	0.50000	0.50000	€ -3.12500	€ -3.12500
128	€ 6.35	€ 6.40	1.00000	0.00000	€ -6.35000	€ -6.35000

Note. The expected value of the current pump (F) is computed by multiplying the success chance (E) by 0.05, then subtracting the product of the explosion chance (D) and the balloon value before the pump (B) [$F = E * 0.05 - D * B$]. Alternatively, one can also take into account the expected value of any subsequent pumps, insofar as they are advantageous (G). This results in somewhat different values, but an identical tipping point at 64.

One approach for resolving the confounding and decomposability issues in the BART is to apply a computational model to its data that quantifies the cognitive mechanisms underlying the observed behaviour (Bishara et al., 2009). Such models were first proposed by Wallsten et al. (2005), inspired by an expectancy-valence model for decomposing behaviour in the Iowa Gambling Task (Busemeyer & Stout, 2002). Wallsten et al. explain decision variability using one parameter for risk-taking, one for response consistency, and two for learning. By applying these models, we can study risk-taking – and other aspects that determine BART behaviour – in isolation, by translating "what is observed but relatively uninformative to what is unobserved and relatively informative" (Van Ravenzwaaij et al., 2011, p. 95). However, data from the BART may not be rich enough to warrant the use of complicated decomposition models. For instance, a study on Wallsten et al.'s best performing model demonstrated that its learning parameters could not reliably be recovered (Van Ravenzwaaij et al., 2011). To allow for more extensive decomposition, one may need to resort to a different task, like the Iowa Gambling Task. Alternatively, one could use a task

that by design avoids confounding, such as the Columbia Card Task. Although dynamic and affective like the BART, this task orthogonally varies risk-related constructs, so that they can be decomposed into their underlying mechanisms – like sensitivity to gains, losses, and probabilities – without the use of a computational model (Dijkstra et al., 2020; Figner et al., 2009; Schonberg et al., 2011). Finally, researchers can choose to design a custom task to ensure that the constructs relevant to their hypotheses are not confounded. For example, a risk task presented in Schmidt et al. (2013) varies the level of risk but holds expected value constant. Solutions such as these should be considered carefully so that constructs crucial to a study's hypotheses can be isolated effectively.

3.2.4 The Normative Solution

The BART is designed in such a way that the balloons' average explosion point lies at 64, halfway the maximum number of pumps. This is achieved by randomly generating collections of explosion points until one produces an average of 64 over all trials, as well as within each set of 10 trials (Lejuez et al., 2002). Participants can then maximise their earnings by attempting to pump every balloon 64 times, which results in an explosion in about half of the trials, and an optimal overall expected value. Going back to Table 3.1, we can see exactly why this is the optimal, or *normative*, solution in the BART. Up to and including the 64th pump, the expected value of pumping the balloon is positive; after 64, the expected value is (increasingly) negative. It is, therefore, optimal to aim for 64 pumps on every balloon, and then stop. Choosing to pump more *or* fewer than 64 times will decrease expected earnings; and the farther one deviates from the optimum, the lower the expected earnings become (Lejuez et al., 2002; Pleskac et al., 2008; Wallsten et al., 2005). Remarkably, in most trials, participants stop pumping the balloon far before the optimal stopping point (Lejuez et al., 2002). In fact, the average adjusted pump score is typically between 26 and 35 (Pleskac et al., 2008). Real-world risk-avoiders and risk-takers alike rarely pump the balloon enough times to maximise their expected earnings. This is less of a problem in the automatic BART, although participants there still pump fewer than 64 times on average. For example, two recent studies reported averages of 61.9 (Bernoster et al., 2019) and 58.5 pumps (De Groot & Van Strien, 2019).

It is yet unknown exactly why participants often stop pumping before they reach the optimal point, but various factors may play a role. First, since the original BART requires participants to inflate balloons one pump at a time, it is plausible that they get tired of pumping after a while. Second, participants may want to limit their effort out of laziness or a desire to finish early (but see Young & McCoy, 2019). Third, they may become satiated:

due to diminishing marginal returns, adding 5 cents to a growing temporary bank may stop being an attractive prospect well before reaching pump 64. Fourth, participants may need time to learn which strategy results in maximal earnings (Lejuez et al., 2002). This conjecture is supported by the observation that participants in both the original and the automatic BART on average press closer to the normative solution in the final block of 10 trials than they do in previous blocks (De Groot & Van Strien, 2019; Lejuez et al., 2002).[1] It also corresponds with the presumed shift from deciding under uncertainty to deciding under risk. In the BART, learning the optimal solution is hard, as the range of possible explosion points is large (1-128), and individual explosions provide limited feedback. This is in line with findings by Lejuez et al. (2002), who show that larger explosion ranges result in larger relative deviations from the optimum.

The fact that participants in the BART often stop pumping before the optimal stopping point has serious implications for how the data can be interpreted. Up to 64 pumps, the risk they take can be characterised as *adaptive* or *functional*, as it results in higher earnings. After that point, it can be considered *maladaptive* or *dysfunctional*, as it reduces expected earnings. Since people generally pump fewer than 64 times, the BART cannot properly differentiate between adaptive and maladaptive risk behaviour, neither within nor between participants. A second, related problem is that experimental manipulations meant to increase risk-taking (such as adding time pressure or administering a certain drug) generally do not lead to lower earnings, as even the resulting higher pump numbers usually do not exceed 64 (Pleskac et al., 2008). For example, if a manipulation causes participants to take more risk and press 50 instead of 30 times, they are actually, on average, *better* off than before, the opposite of what one would expect in real life. In short, if participants mostly stay under 64 pumps, they simply never reach the point where taking more risk becomes disadvantageous, which limits the conclusions one can draw from the data.

The most straightforward way to mitigate these problems may be the modified BART developed by Pleskac et al. (2008), which differs from the original task in three ways. First, it involves an automatic response mode: participants indicate their intended number of pumps at the start of each trial, after which the balloon automatically inflates to the corresponding size (or until it bursts). Although meant to mitigate censoring, this adjustment may also prevent people from getting tired of pumping and from wanting to finish the task sooner. Second, the adjusted task provides explicit feedback about the explosion point of *every* balloon, not merely of those that actually explode. This may improve participants' learning across trials. Third, participants are (truthfully) informed that the range of pump numbers is 1-128 and that the best overall number of pumps is 64, further increasing the amount of information at their disposal.

These three modifications together successfully moved participants' behaviour closer to the normative solution of 64, with an average pump score of 57.7 for females and 63.7 for males (Pleskac et al., 2008). Part of this effect can be attributed to the automatic response mode, as these averages are higher than those from a manual BART with full feedback and strategy instructions added. Since this manual BART itself resulted in higher averages than the original BART, the feedback and instructions likely also contributed to the effect (Lejuez et al., 2002). Recent research, however, indicates that informing participants about the optimal strategy is not necessary, and even ill-advised. Two studies using an automatic BART with full feedback – but without strategy instructions – found equally high pump averages as did Pleskac and colleagues (Bernoster et al., 2019; De Groot & Van Strien, 2019). Additionally, these studies found that a subgroup of participants – often from a STEM background – seem to infer the optimal strategy without any help.[2] Their repeated 64-answers, therefore, reflect cognitive ability rather than risk propensity and reduce task variability. Informing participants about the optimal strategy can increase such problematic responses. Therefore, it seems best to add automatic responses and full feedback to the BART, but not strategy instructions. This will likely elicit sufficiently high pump averages, without compromising the validity of the task.

3.3 Discussion

Since it was first published in 2002, the BART has become one of the most popular tools in psychology to gauge individuals' propensity for risk-taking. Halfway through 2020, the original article describing the BART (Lejuez et al., 2002) had been cited over 1100 times in Scopus, most often in journals on decision research, addiction, and neuropsychology. This popularity is well-founded. The BART succeeds in recreating the "natural" feeling of exhilaration and tension people experience when taking risk, and thus has excellent *ecological validity*. Furthermore, it correlates well with self-reported risk-related constructs, such as impulsivity and sensation seeking, and with real-world risk behaviours, like polydrug use and unsafe sex, supporting its *convergent validity*. Lastly, it does not correlate with constructs like depression and anxiety, endorsing its *discriminant validity* (Lejuez et al., 2002). But despite these qualities, the BART suffers from methodological problems, most of which have been acknowledged in previous research as negatively impacting its rigor. The present paper is the first to give a comprehensive overview of these problems.

[2] The relevant data collected by Bernoster et al. (2019) and De Groot and Van Strien (2019) on individual answering patterns was not published but will be shared upon request.

The *first* problem concerns the lack of clarity as to whether decisions in the BART are made under uncertainty (where outcome probabilities are unknown) or risk (where they are known). Since participants are not given any information about the explosion probabilities, they first decide under uncertainty, which then gradually shifts towards risk as they learn more about the probabilities in the task. As it is unclear exactly when this shift takes place, it is difficult to determine whether a given decision is made under uncertainty, risk, or something in between. The *second* problem concerns statistical censoring, which occurs in trials where the balloon explodes, as participants are then prevented from taking additional risk. As a result, the average number of times participants pump the balloon underestimates their risk propensity. *Third*, the BART confounds risk with expected value. Since these constructs change simultaneously throughout a trial, participants' pump behaviour again does not reflect risk propensity, as decisions are influenced by both risk and expected value. This also means that the task is poorly decomposable, as it cannot disentangle the motives underlying a pump decision. A *final* problem concerns the task's normative solution. In the majority of trials, participants stop pumping before the point where expected earnings are maximised. Therefore, participants mostly take adaptive risk, which leads to higher earnings. Maladaptive risk-taking hardly occurs, even though one would expect to see such behaviour in certain cases.

Despite these problems, much of the research up to now has focused on the empirical findings produced by the BART, rather than on the task itself, with the majority of researchers using the task without critically reviewing whether its problems interfere with their aims. This can have undesirable consequences, such as when it leads to false positives or false negatives. For example, one may fail to show a relationship which only exists for decisions characterised by risk, as some trials in the BART are characterised by uncertainty instead. Conversely, a hypothesis may pertain to people's response to changing risk and be unjustly supported, as in the BART, risk and expected value simultaneously change and impact individuals' behaviour. Finding true positives and negatives hinges on several factors, an important one being the validity of the measurement instrument. Any data resulting from instruments that suffer from methodological or interpretational problems is of limited value to understanding the concepts they are supposed to operationalise.

For these reasons, it is imperative that researchers critically evaluate the "fit" between their research and the BART before deciding on using it. For many research aims, one will now see that the original BART does not suffice. Yet despite these "burst beliefs", there are three types of approaches one can take to account for its limitations. *First*, data from the original BART can be analysed using a different, more informative index than Lejuez et al.'s average adjusted pump score. For example, the models by Wallsten et al. (2005) break down

behaviour into risk-taking, response consistency, and learning. In addition, computational models can be used to take into account censoring and to provide an index of uncensored risk-taking in the BART (Dijkstra et al., 2020; Tobin, 1958; Weller et al., 2019; Young & McCoy, 2019). A *second* way of dealing with the BART's limitations is by modifying the task, for example by rigging it (Figner et al., 2009; Slovic, 1966), providing additional feedback, or automating the responses (Pleskac et al., 2008). *Third*, one may consider using a different task. This can be an existing (sequential) risk-taking task, like the Columbia Card Task (Figner et al., 2009), which performs better in terms of decomposability than the BART. Alternatively, researchers should consider creating a custom task that exactly suits their research, avoiding methodological flaws that could endanger the soundness of their conclusions. For instance, a task developed by Schmidt et al. (2013) involves decisions under conditions of explicit risk and does not confound risk with expected value. An important goal to keep in mind when designing such bespoke tasks is to combine strong ecological validity with methodological rigor (Schonberg et al., 2011).

Clearly, none of the solutions proposed can be considered a "universal" fix that solves all of the BART's problems. Depending on the aims of any given study, certain problems will be more pressing than others, indicating the (type of) solutions most needed. By combining solutions, researchers could work towards a task that can operationalise risk propensity without substantial methodological or interpretational problems. For example, an automatic BART with full feedback and explicit information on the probability distribution provides uncensored decisions made under clear risk that are at times risky enough to be maladaptive. If the resulting data from this adapted BART are then analysed using a model like that by Wallsten et al. (2005) or that by Van Ravenzwaaij et al. (2011), all problems reviewed in the current commentary would be addressed. However, this does not necessarily mean that this combination of solutions constitutes a universal fix after all, as the BART may face more problems than the ones discussed here. In all likelihood, the present review is not exhaustive. Researchers using the BART may know of additional problems, although this is unlikely to show in their work, as journals – and by extension researchers – do not consider "failure" a popular publishing theme (Ferguson & Heene, 2012; Song et al., 2009). Therefore, it is important for researchers to not only critically evaluate the instruments they use but to disclose these evaluations as well, so that any and all methodological shortcomings can be openly discussed and addressed, improving the quality of the measures used.

3.4 Conclusion

The present paper is the first to review the methodological shortcomings of the Balloon Analogue Risk Task, a highly popular risk-taking task in psychology. The main problems identified are the ambiguity between uncertainty and risk, censoring of observations, confounding of risk and expected value, and poor decomposability into adaptive and maladaptive risk-taking. In addition, the paper reviews solutions that mitigate these problems. By presenting this first-time inventory, the paper highlights earlier mentions of problems in the BART as well as proposed solutions. It calls for a critical attitude towards the BART and experimental tasks in general, as their design deserves at least as much attention as the findings they produce. It also sets the agenda for testing and comparing different tasks and task versions, to explore which designs result in the best usability, reliability, and validity, so that risk propensity can be measured in the most accurate way possible.

Chapter 4

Entrepreneurial Orientation and Decision-Making Under Risk and Uncertainty: Experimental Evidence From the Columbia Card Task

Abstract

We analyse the relationship between entrepreneurship and decision-making under risk and uncertainty using the Columbia Card Task: an experimental task eliciting affective decision-making under conditions of risk and uncertainty. In a sample of 127 university students, we find robust evidence that Individual Entrepreneurial Orientation (IEO) is negatively related to decision-making under risk and uncertainty. In addition, we show that distinguishing the subscales of IEO is key to understanding this relationship, since while the relationship is negative for the Proactiveness and Innovativeness subscales, it is positive for the Risk-Taking subscale. Moderation analyses show that heterogeneous sensitivity towards possible gains and losses explains the main relationship.

This chapter has been published as: Dijkstra, N. F., De Groot, K., & Rietveld, C. A. (2023). Entrepreneurial orientation and decision-making under risk and uncertainty: Experimental evidence from the Columbia Card Task. *Applied Psychology, 72*(4), 1577-1592. https://doi.org/10.1111/apps.12436

4.1 Introduction

Risk and uncertainty are key concepts in entrepreneurship theories (Van Praag, 1999), and a stream of literature explains how behavioural preferences make that some individuals are more willing than others to accept the risks and uncertainties that come with entrepreneurship (Åstebro et al., 2014; Rauch & Frese, 2007). Yet, empirical evidence on the relationship between risk and uncertainty preferences on the one hand and entrepreneurship on the other hand is rather mixed (Holm et al., 2013; Koudstaal et al., 2016). These mixed findings have been argued to originate from, among others, the use of self-report inventories instead of experimental measures (in which behaviour is actually observed) and the use of relatively simple scenarios to elicit individuals' behaviour that do not sufficiently resemble the complexity of real-world decision-making. Remedying these limitations, surprisingly, "lab-in-the-field" studies by Holm et al. (2013) and Koudstaal et al. (2016) find no differences between the behaviour of entrepreneurs and that of others using lottery tasks. The present study contributes to the literature by using an experimental paradigm from the field of psychology to examine risk and uncertainty preferences in relation to entrepreneurship. That is, we use the Columbia Card Task (CCT; Figner et al., 2009) to elicit affective decision-making under conditions of risk and uncertainty, and we relate observed experimental behaviour to the entrepreneurial orientation (EO) of individuals (Bolton & Lane, 2012) to seek empirical validation of a cornerstone of entrepreneurship theories (and layman definitions on what constitutes an entrepreneur).

4.2 Theoretical Background

4.2.1 Lab-in-the-Field Experiments

In light of the mixed findings on the relationship between risk and uncertainty preferences and entrepreneurship, Holm et al. (2013) and Koudstaal et al. (2016) re-examined this topic by carrying out lab-in-the-field experiments among entrepreneurs, managers, and employees, to derive risk preferences from actual choice behaviour in (incentivised) lottery tasks. In these experiments, participants are asked to choose between two options that both specify the chance of winning (or losing) a certain amount of money. The chance of winning (or losing), the direction of the outcome (gain or loss), as well as the outcome magnitude are systematically varied across trials. Based upon observed behaviour in these trials, one can describe the participants' behaviour using expected utility theory (Von Neumann & Morgenstern, 1944) or even more comprehensively using (cumulative) prospect

theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). That is, the design of these experiments enables the description of individual behaviour in terms of outcome sensitivity, aversion to loss, and probability weighting.

Both Holm et al. (2013) and Koudstaal et al. (2016) find that the behaviour of entrepreneurs does not differ from that of others in situations characterised by uncertainty and situations characterised by risk. Specifically, Holm et al. (2013) find that Chinese entrepreneurs are not more risk-taking than non-entrepreneurs from the same region nor do they differ in ambiguity aversion. Koudstaal et al. (2016) find that Dutch entrepreneurs are similar to managers when it comes to risk and ambiguity aversion (although they differ in aversion to loss). Furthermore, by also employing a measure of self-rated risk aversion, Koudstaal et al. (2016) find that while entrepreneurs and managers do not differ in actual choice behaviour under risk and uncertainty, they do differ in how risk-seeking they perceive (self-report) themselves to be.

4.2.2 Affective Engagement

Their own results support Holm et al.'s (2013) and Koudstaal et al.'s (2016) premise that previous mixed findings regarding the relationship between entrepreneurship and decision-making under risk and uncertainty may partly originate from the use of different research instruments with different levels of precision. Lottery tasks, as used by Holm et al. (2013) and Koudstaal et al. (2016), have, for instance, been criticised for being artificial, possibly impacting the external or ecological validity of the findings (Pedroni et al., 2018; Schonberg et al., 2011). Importantly, the absence of immediate feedback in lottery tasks makes that affective responses (senses of escalating tension and exhilaration) are not elicited. Affective responses are, however, often central to real-life risk-taking (Schonberg et al., 2011). Hence, in addition to systematic trial-to-trial variation and the use of incentives, risk-taking tasks should ideally employ a dynamic decision situation that elicits affective engagement.

An experimental task that is characterised by these requirements is the CCT (Figner et al., 2009), in which study subjects turn cards in a series of computer-based trials. Most of the cards are win cards. Turning these cards makes individuals earn points. However, a small number of cards are loss cards, and turning those makes participants lose points. In each trial, three pieces of information are supplied to participants: (1) The points they will gain when turning a win card, (2) the points they will lose when turning a loss card, and (3) the number of loss cards present in the trial. These three parameters are systematically varied over the trials and relate to the main components in cumulative prospect theory, namely

outcome sensitivity, sensitivity to losses, and probability weighting. Thus, despite the CCT originating from the field of psychology, it is also well-suited for use in economic research.

4.2.3 Present Research and Expectation

In the present study, we utilise the CCT to examine the relationship between risk and uncertainty preferences and individual-level aspects of entrepreneurship. Using a sample of university students, we relate study subjects' behaviour in the CCT to their individual entrepreneurial orientation (IEO), a well-known determinant of entrepreneurship. Based on the original firm-level scale for EO (Lumpkin & Dess, 1996; Lumpkin et al., 2009), IEO (Bolton & Lane, 2012) measures the traits and attitudes that are inherent in the original EO scale at the individual level. Given theoretical presumptions (Van Praag, 1999) and the ecological validity of the CCT (Pedroni et al., 2018; Schonberg et al., 2011), we expect behavioural risk and uncertainty preferences to be positively associated with IEO in our experiments.

The use of a student sample comes with the advantage that the reverse impact of occupational experiences on traits and attitudes is likely to be small (Bernoster et al., 2018). Moreover, the relevance of students' IEO within entrepreneurship research has been demonstrated repeatedly. For example, Bolton and Lane (2012) showed in a sample of students that IEO was positively associated with intentions for entrepreneurship, a finding that was supported by Koe (2016) and Sahoo and Panda (2019). Moreover, IEO has been shown to impact competitive strategy (Lechner & Gudmundsson, 2014) and entrepreneurial performance (Bolton, 2012). Nonetheless, we recognise that the two studies most directly related to the present study (Holm et al., 2013; Koudstaal et al., 2016) employ a measure for entrepreneurship related to actual occupational involvement. We discuss the potential impact of this difference in the Discussion section.

4.3 Method

4.3.1 Participants

A total of 130 university students from a variety of degree programmes participated in the study. Three of these participants did fill out the questionnaire (including the questions on IEO) but did not participate in the experiment (one for medical reasons and two did not show up). These participants were excluded from all analyses. Participants completed two versions of the CCT: one in which decisions were taken under risk and one in which

decisions were taken under conditions of uncertainty. Because of incorrect software settings, three participants were only able to complete part of the uncertainty CCT, which resulted in an analysis sample of n = 127 for decision-making under risk and n = 124 for decision-making under uncertainty. In the remainder of this article, we refer to these samples as the "Risk sample" and the "Uncertainty sample," respectively. The average age in both samples was 20.4 years, which corresponds to approximately third year bachelor students at the university where the experiments were conducted. A little more than half of the participants (55% in both samples) was female. The study was performed in accordance with the Helsinki declaration, and all procedures were approved by the institutional review board. All participants provided signed informed consent before participating. In return for their participation, psychology students received course credits, while the other students received a standard fee of 17.5 euros. In addition, all students could earn up to 7.5 euros extra based on task performance.

4.3.2 Procedure

Students voluntarily signed up for the study via one of three online participant systems running at the Dutch university the study was conducted at. One system was only accessible to psychology students; the other two were available to students from all disciplines. After signing up, participants received an e-mail in which they were asked to not consume any alcohol, coffee, or energy drinks the day of their appointment (to prevent these substances from impacting their behaviour). The e-mail also contained a link to a secured online survey including an informed consent form, questions on demographics, and the IEO questionnaire, which participants were asked to complete prior to their lab session and which took 10-15 minutes to complete. During their lab visit, students performed two versions of the CCT (one under uncertainty and one under risk) while being seated in a comfortable chair in a dimly lit sound-attenuated room. The tasks took approximately 30 minutes. The order in which the tasks were presented was counterbalanced. The tasks were programmed and presented using E-prime software. Before starting the tasks, participants received on-screen instructions, after which the experimenter confirmed that the participant had understood these. After finishing the tasks, participants were thanked for their participation and debriefed about the aims of the study.

4.3.3 Measures

4.3.3.1 Individual Entrepreneurial Orientation (IEO). The IEO questionnaire (Bolton & Lane, 2012) consists of 10 items, of which 3 form a Risk-Taking subscale, 4 an

Innovativeness subscale, and 3 a subscale on Proactiveness. These dimensions can vary independently from each other (Lumpkin & Dess, 1996) and are thus relevant to examine separately, in addition to the subjects' total IEO score. All 10 questionnaire items can be found in Table S4.1 in the Supplementary Materials. Items are answered on a 5-point Likert-scale (1 "Strongly disagree" to 5 "Strongly agree"), resulting in a total score range of 10-50, with higher scores indicating stronger EO. In our samples (n = 127 and n = 124), Cronbach's alpha was .70 for the full scale, .64 for Risk-Taking (.65 in the Uncertainty sample), .65 for Innovativeness, and .42 for Proactiveness. These lower reliabilities are likely partially driven by the small number of items per subscale (Pallant, 2007). However, for the Proactiveness subscale, a dimensionality problem was observed as well: a confirmatory factor analysis showed a relatively poor fit for the full model (χ^2 (32) = 77.74, $p < 0.05$, CFI = 0.80, RMSEA = 0.11, SRMR = 0.09), with the Risk-Taking and Innovativeness items loading significantly ($p < 0.001$) on their corresponding factors (standardised loadings 0.34-0.78) but the Proactiveness items loading poorly onto their factor. The results in the Uncertainty sample were virtually the same. This confirms that while findings related to the full IEO scale and the Risk and Innovativeness subscales can be interpreted without caution, those related to the Proactiveness subscale do require discretion. For this reason, as a robustness check, we additionally analysed a model specification in which the three Proactiveness items were added separately.

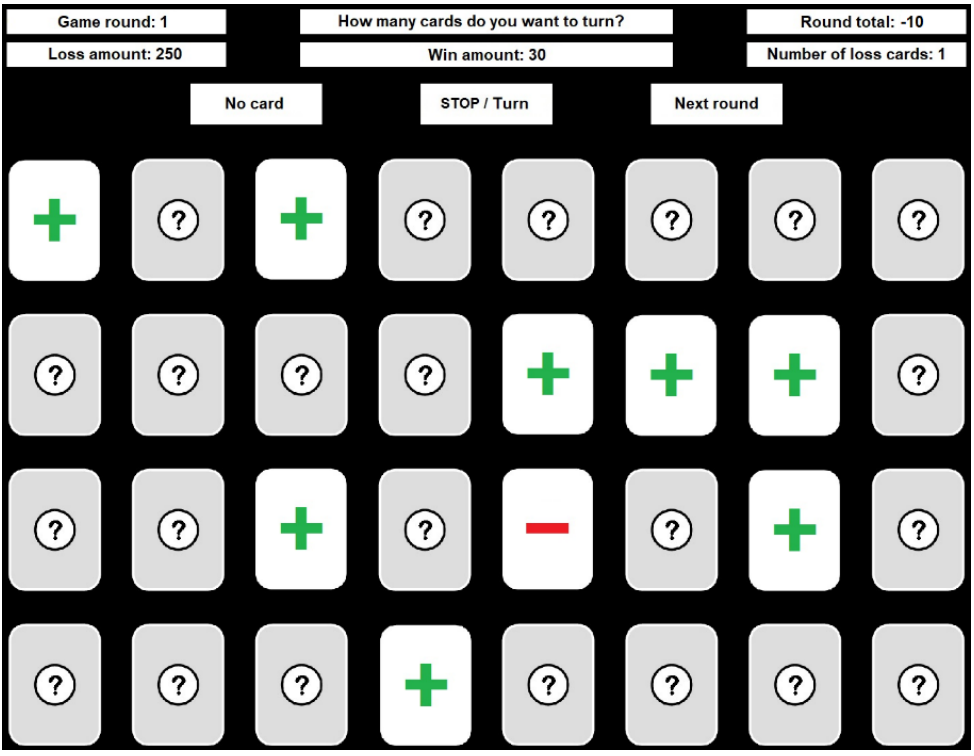
4.3.3.2 Columbia Card Task (CCT). Decision-making under risk and uncertainty was measured using the "warm"[3] CCT (Huang et al., 2013), which presents participants with a virtual array of 32 (4 × 8) cards that are positioned face down. Participants are informed that most of these cards are "win cards", which make them earn points that can be redeemed for money at the end of the task. However, there are also some "loss cards" for which points (and thus money) will be subtracted if turned. An example of a trial set-up is shown in Figure 4.1. The settings of the trial are presented at the top of the screen. Participants earn either 10 or 30 points per turned win card, lose either 250 or 750 points when turning a loss card, and either 1 or 3 loss cards are present among all cards. These parameters are systematically varied across trials. In this example, the win amount is 30, the loss amount 250, and only 1 loss card is present. At the start of a trial, all cards lie face down. Participants select the cards they want to turn (or click "no card" if they prefer to turn no cards). Then, they press the

[3] A "hot" and "cold" CCT also exist. In the hot CCT, cards turn directly after they are selected, which elicits affective engagement but also creates statistical censoring since in case of a loss card the trial forcedly ends and it is not known how many cards the person intended to turn. In the cold CCT, participants indicate at the start of a trial how many cards they want to turn, after which these are turned without the participant watching. Censoring is not an issue in this CCT version, but affective engagement is not elicited. The warm CCT does elicit affective engagement and is not censored.

"STOP/Turn" button, after which the cards automatically turn in the order they were selected in. If a loss card is encountered, the trial forcedly stops and the specified loss amount is subtracted from the already accumulated wins. In the example presented in Figure 4.1, the participant presses the STOP/Turn button after selecting nine cards, in which the ninth card is a loss card. In this scenario, earnings for this trial are $8 \times 30 - 250 = -10$ points. While originally the win and loss cards had respectively yellow smiley faces and red frowny faces (Figner et al., 2009; Huang et al., 2013), the present study adopted the design with a green plus sign and a red minus sign to reduce the possible impact of (abstracted) facial features on individuals' emotion and thereby decision-making.

Figure 4.1

Experimental Set-Up of the Warm Columbia Card Task (CCT)



The three information "parameters" (win points, loss points, and number of loss cards) were orthogonally varied across trials by means of a full factorial design, meaning that it took 2 × 2 × 2 = 8 trials for every possible combination of parameters to ensue. For the present study, all combinations were run six times (in random order), resulting in 6 × 8 = 48 trials per participant. In addition, the CCT was modified to reflect a situation of decision-making under uncertainty instead of risk. To this end, a second series of 48 trials was run in which information on the number of loss cards was withheld from participants. If participants inquired about the number of loss cards present in this "Uncertainty CCT" (which was still either 1 or 3), the experimenter answered that this number was random and could be anything between 0 and 32. For both the original ("Risk") and the modified ("Uncertainty") CCT, the main outcome variable was the number of cards selected, which is considered a measure of risk-taking behaviour as the likelihood of experiencing a loss increases with each card that is turned (Figner et al., 2009). A secondary outcome variable was the overall performance in terms of points earned across all 48 trials.

4.3.4 Statistical Methodology

To explain the number of cards selected in a trial in the CCT, we employ a model derived from the Censored Mixture Model (CMM), which was specifically designed for the CCT by Dijkstra et al. (2022). The CMM constitutes a Generalised Linear Model with some additional features. Because of the non-negative (0–32) and discrete (0, 1, 2, …) nature of the dependent variable, the model assumes the dependent variable to follow a negative binomial distribution. This distribution is preferred over the Poisson distribution, because it allows the mean and variance of the distribution of the dependent variable to be different. In the negative binomial distribution, the mean is specified through an inverse link function. To be able to interpret the regression coefficients in a linear fashion, we specify the inverse link function as follows:

$$h^{-1}(\eta_{it}) = log(exp(\eta_{it}) + 1),$$

with η_{it} being the linear combination of the explanatory variables included in the vector x'_{it}:

$$\eta_{it} = \alpha + x'_{it}\beta.$$

Importantly, the CMM offers flexibility regarding the probability assigned to certain outcomes. Because of the set-up of the CCT, a 4 × 8 array, participants often select multiples of four cards (e.g., they turn one complete row or column) or create a geometric pattern. Participants also frequently decide to not select any cards. Within the CMM, such excesses are categorised and assigned extra probability mass. Our first category consists of the outcome 0; the second category contains the outcomes 4, 8, 10, 12, 16, and 20. These categories are justified by the observed distribution of the number of cards selected (see Figure S4.1 in the Supplementary Materials).

Our main analyses constitute three different models which are similar in terms of the dependent variable (the number of selected cards) and control variables (Sex and Age). In the first model, the main explanatory variable is IEO. In the second model, the main explanatory variables are the subscales of IEO: Risk-Taking, Innovativeness, and Proactiveness. This model allows us to analyse which subcomponents of IEO are driving the results in the first model. In the third model, we extend the first model with interaction terms between IEO and the experimental parameters related to reward sensitivity (10 vs. 30 points per win card), loss sensitivity (–250 vs. –750 points in case of a loss card), and probability weighting (1 vs. 3 loss cards in the trial). Hence, the third model allows us to analyse whether heterogeneous sensitivity to possible gains, losses, or success chances explains the relationship between IEO and decision-making under risk and uncertainty in the first model. We estimate these three models using data from the Risk sample as well as data from the Uncertainty sample.

As control variables, we included dummy variables for the experimental conditions in a specific trial: 10 versus 30 points per win card, –250 versus –750 points in case of a loss card, and 1 versus 3 loss cards present. Moreover, we included a dummy variable indicating whether participants played the Risk CCT or Uncertainty CCT first (this order was counterbalanced to deal with possible sequence effects). Because of the experimental design and the relatively homogenous sample, we limited the further control variables to Sex (Female = 1, Male = 0) and Age (in years). To facilitate interpretation of effect sizes in the regression models, the continuous variables (i.e., IEO and its subscales and age) are standardised to have mean 0 and standard deviation 1 in the analysis samples.

4.4 Results

4.4.1 Descriptive Statistics

Table 4.1 contains the descriptive statistics of the two samples. It is important to note that in this table (as well as in Table 4.2), we look at the average number of selected cards by an individual (n = 127/124), whereas in the regressions, we model the number of selected cards by an individual in each specific trial ($n \times t$ = 6096/5952). Presumably because of the higher degree of uncertainty, the average number of selected cards is lower in the Uncertainty sample than in the Risk sample. The mean (unstandardised) values for IEO and its subscales are approximately in the middle of the respective scales. The slight differences between the descriptive statistics of the Risk sample and the Uncertainty sample are explained by the difference in sample size.

The pairwise correlations between the main variables in the analysis samples are shown in Table 4.2. The upper triangle reflects the correlations in the Risk sample and the lower triangle those in the Uncertainty sample. We find a significantly negative correlation between the average number of cards selected and IEO in the Risk sample. However, this correlation is insignificant in the Uncertainty sample. The table also shows positive correlations across the subscales of the IEO in both samples, which is in line with the results of Bolton and Lane (2012).

Table 4.1

Descriptive Statistics of the Analysis Samples

	Risk sample (n = 127)				Uncertainty sample (n = 124)			
	Mean	*SD*	Min.	Max.	Mean	*SD*	Min.	Max.
Average #cards selected	7.79	6.69	0	32	6.69	6.06	0	32
IEO	35.94	4.92	22	48	35.98	4.91	22	48
IEO Risk-Taking	10.47	2.17	5	15	10.47	2.15	5	15
IEO Innovativeness	14.32	2.57	9	20	14.35	2.57	9	20
IEO Proactiveness	11.15	1.99	6	15	11.15	1.97	6	15
Age (in years)	20.41	2.19	17	29	20.40	2.20	17	29
Sex (female)	0.55	0.50	0	1	0.55	0.50	0	1

Note. IEO = Individual Entrepreneurial Orientation; *SD* = Standard deviation; Min. = Minimum; Max. = Maximum.

Table 4.2

Pairwise Correlations of the Analysis Samples

		Risk sample (n = 127)						
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Uncertainty sample (n = 124)	(1) Average #cards selected	-	-0.27***	-0.11	-0.22**	-0.27***	-0.02	0.03
	(2) IEO	-0.02	-	0.75***	0.82***	0.59***	0.12	-0.18**
	(3) IEO Risk-Taking	0.06	0.76***	-	0.47***	0.17*	0.10	-0.30***
	(4) IEO Innovativeness	-0.03	0.82***	0.46***	-	0.22**	0.13	-0.11
	(5) IEO Proactiveness	-0.08	0.60***	0.18**	0.23**	-	0.01	0.02
	(6) Age	-0.05	0.13	0.09	0.13	0.04	-	0.08
	(7) Sex (female)	0.13	-0.20**	-0.31***	-0.11	-0.00	0.10	-

Note. IEO = Individual Entrepreneurial Orientation; *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

4.4.2 Regression Results

4.4.2.1 Main Variables. Tables 4.3 and 4.4 display the results of the regressions explaining the number of cards selected in a trial in the Risk sample and Uncertainty sample, respectively. Model 1 in both tables shows that students with a high EO exhibit more risk-averse behaviour. That is, they select fewer cards when the probabilities of a loss card are known (Table 4.3) as well as in the situation where these probabilities are unknown (Table 4.4). Nevertheless, the effect of IEO is considerably smaller in Table 4.4 than in Table 4.3.

Consistent with the results for the total IEO score, the effects of the IEO subscales (Model 2 in Tables 4.3 and 4.4) are larger in magnitude in the Risk sample than in the Uncertainty sample. Reassuringly, students who indicate to be risk-seeking select more cards in the CCT than students who indicate to be less risk-seeking. On the contrary, more innovative and/or proactive students take less risk than those who self-report to be less innovative and/or proactive. However, the result for innovativeness is not statistically significant in the Uncertainty sample.

In Model 3, we analyse whether heterogeneous sensitivity to possible gains, losses, or success chances explains the relationship between IEO and decision-making under risk and uncertainty in Model 1. The results show that the interaction effects of IEO with the gain amount and loss amount are statistically significant in both samples but that the interaction with the number of loss cards is insignificant. Hence, in both experimental settings, entrepreneurially oriented students are more sensitive to the gain and loss amount than students who are less entrepreneurially oriented.

Table 4.3

Results of the Regressions Explaining the Number of Cards Selected in a Trial (Risk Sample)

	Model 1		Model 2		Model 3	
IEO	-0.838***	(0.062)			-0.914***	(0.168)
IEO Risk-Taking			0.336***	(0.086)		
IEO Innovativeness			-0.588***	(0.074)		
IEO Proactiveness			-0.746***	(0.076)		
IEO × Gain amount (30)					0.615***	(0.136)
IEO × Loss amount (750)					-0.329**	(0.146)
IEO × #Loss cards (3)					0.194	(0.154)
Gain amount (30)	2.559***	(0.137)	2.311***	(0.138)	2.193***	(0.141)
Loss amount (750)	-3.230***	(0.146)	-3.113***	(0.143)	-3.049***	(0.146)
#Loss cards (3)	-4.447***	(0.159)	-4.700***	(0.156)	-4.821***	(0.159)
Task order (uncertainty first)	-1.143***	(0.131)	-1.336***	(0.141)	-1.201***	(0.140)
Sex (female)	1.157***	(0.127)	0.956***	(0.146)	0.632***	(0.142)
Age	-0.145**	(0.064)	-0.270***	(0.065)	-0.232***	(0.064)
Intercept	9.888***	(0.190)	10.887***	(0.193)	11.050***	(0.196)
Likelihood value	17819.34		17642.25		17668.15	
n	127		127		127	
n×*t*	6096		6096		6096	

Note. Coefficient with standard errors in parentheses; IEO = Individual Entrepreneurial Orientation; Reference categories are 10 for Gain amount, 250 for Loss amount, and 1 for #Loss cards; Continuous variables (IEO and its subscales and age) have been standardised to *Z*-scores; *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 4.4

Results of the Regressions Explaining the Number of Cards Selected in a Trial (Uncertainty Sample)

	Model 1		Model 2		Model 3	
IEO	-0.124**	(0.061)			-0.030	(0.139)
IEO Risk-Taking			0.196**	(0.078)		
IEO Innovativeness			-0.087	(0.061)		
IEO Proactiveness			-0.208***	(0.063)		
IEO × Gain amount (30)					0.543***	(0.128)
IEO × Loss amount (750)					-0.411***	(0.134)
IEO × #Loss cards (3)					0.126	(0.113)
Gain amount (30)	3.303***	(0.132)	3.160***	(0.129)	3.278***	(0.131)
Loss amount (750)	-4.177***	(0.137)	-3.958***	(0.134)	-4.153***	(0.137)
#Loss cards (3)	-0.094	(0.115)	-0.024	(0.114)	-0.087	(0.115)
Task order (uncertainty first)	1.992***	(0.125)	1.953***	(0.126)	1.976***	(0.124)
Sex (female)	1.168***	(0.122)	1.178***	(0.127)	1.192***	(0.122)
Age	-0.258***	(0.063)	-0.270***	(0.063)	-0.243***	(0.063)
Intercept	5.537***	(0.162)	5.709***	(0.161)	5.520***	(0.162)
Likelihood value	16625.83		16686.88		16608.48	
n	124		124		124	
n×*t*	5952		5952		5952	

Note. Coefficient with standard errors in parentheses; IEO = Individual Entrepreneurial Orientation; Reference categories are 10 for Gain amount, 250 for Loss amount, and 1 for #Loss cards; Continuous variables (IEO and its subscales and age) have been standardised to *Z*-scores; *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

4.4.2.2 Control Variables. The effects of the control variables are fairly consistent across the different experimental set-ups and models. First, females take more risk than males, and older students are more risk-averse compared to younger students. Second, the effects of the game parameters (gain amount, loss amount, and number of loss cards) are all in the expected direction, with the largest effect for probability, followed by the loss amount and then gain amount. In other words, participants are most sensitive to the number of loss cards present. However, this is only the case in the Risk sample (Table 4.3), and not in the Uncertainty sample (Table 4.4), where this information was withheld from participants and could thus not play a role in their decision-making. Indeed, in the Uncertainty sample, the number of loss cards present does not significantly impact the number of cards participants turn. Related to this, the effects of the loss and gain amount are stronger in the Uncertainty sample than in the Risk sample, possibly because when people have less information, they value the information that *is* available more. Finally, regarding task order, we find that irrespective of which task version is played first, participants take less risk (i.e., they select

fewer cards) in their second task. The absolute effect of task order is somewhat stronger in the Uncertainty sample: those participants for whom the Uncertainty game is their second task (due to counterbalancing) select ~2 cards fewer than participants for whom the Uncertainty game is their first task, while this effect is ~1 card in the Risk sample. Playing more and more trials in the CCT may make participants increasingly aware of the difficulty of decision-making under risk and uncertainty. As a result, on average, they tend to reduce the number of cards selected throughout the trials. This reduction seems to carry over to the second task the participants engage in.

4.4.3 Robustness Checks

4.4.3.1 Unconventional Behaviour and Fatigue. A few participants repeatedly selected 32 cards, a response that by definition leads to encountering a loss card. Although participants may select 32 cards once or twice to explore the structure of the CCT, they should soon realise that there are indeed loss cards present in the trials (and that they are not being deceived). In addition, some participants showed signs of fatigue, evidenced by them repeatedly selecting the "no card" option (thus selecting zero cards) later in the task. Therefore, we performed a robustness check in which we excluded participants who selected 32 cards more than twice across the full 48 trials and/or who selected zero cards five times or more in the last nine trials. The results of the regressions with these participants excluded are presented in Tables S4.2 and S4.3 in the Supplementary Materials. The main results did not change qualitatively, indicating that unconventional behaviour and fatigue did not drive the main results. A few changes were observed for the control variables. In the Risk sample, males now took more risk than females, and age was no longer significant. This difference seems to result from dropping relatively many male participants who frequently chose the "no card" option in the last few trials, and the smaller analysis sample. In the Uncertainty sample, age and the interaction terms between IEO and the game settings in Model 3 were no longer significant.

4.4.3.2 Overall Performance. While the number of cards selected is the main outcome in the CCT, performance in terms of the points earned is also relevant as the number of cards selected is the way through which participants try to maximise their compensation for taking part in the study. Tables S4.4 and S4.5 in the Supplementary Materials present the results of regressions explaining the average number of points earned in a trial. In the Risk sample, only sex is significant, with males earning more than females. In the Uncertainty sample, next to sex, also task order and age have a significant impact. In case a participant first played the Risk version, they perform better in the Uncertainty version. Furthermore, older

participants perform better. Interestingly, neither IEO (Model 1) nor its subscales (Model 2) significantly affect performance in either of the two versions of the CCT. The absence of a relationship between IEO and points earned in the CCT backs the customary interpretation of the number of cards selected (and not the points earned) as a proxy for risk preferences. After all, the number of cards a participant selects is a direct proxy of their risk preferences, whereas the eventual payoff is more distal and subject to other factors such as chance.

4.4.3.3 The Proactiveness Subscale. As the reliability of the Proactiveness subscale of the IEO measure was particularly low ($\alpha = .42$), we examined whether the results of Model 2 in Tables 4.3 and 4.4 were driven by the inclusion of somewhat unreliably measured subdimensions of the IEO construct. For this reason, we analysed model specifications in which the three Proactiveness items were added separately. The results are presented in Table S4.6 in the Supplementary Materials. In the Risk sample, we find that the coefficients for Risk-Taking and Innovativeness remain very similar in size and significance. The negative coefficient for Proactiveness in Model 2 seems mostly driven by Item 8 ("I usually act in anticipation of future problems, needs, or changes") and Item 10 ("I prefer to 'step-up' and get things going on projects rather than sit and wait for someone else to do it"). Item 9 ("I tend to plan ahead on projects") is, however, not significantly associated with the model outcome. In the Uncertainty sample, we observe essentially the same patterns although the coefficient for Risk-Taking (while still being positive) loses statistical significance in the robustness check. Still, taken together, the results in Table S4.6 suggest that the results of Model 2 in Table 4.3 and Table 4.4 are not driven by the low internal reliability of the Proactiveness subdimensions of IEO.

4.5 Discussion and Conclusion

The results of our experiments show that the IEO of the students in our sample is related to decision-making under risk in the warm CCT. However, contrary to our expectation and somewhat surprising considering the literature, this relationship is *negative*. The results for the subscales of IEO are less surprising: individuals who indicate to be more risk-seeking indeed take more risk in the behavioural task, whereas those who indicate to be innovative or proactive exhibit more risk-averse behaviour. The moderation analyses indicate that heterogeneous sensitivity to possible gains and losses rather than sensitivity to success chances explains the relationship between IEO and decision-making under risk and uncertainty. We also modified the warm CCT in such a way that the number of loss cards was not known to participants. Overall, we find similar results in the modified CCT, although

the larger uncertainty does translate into smaller effect sizes of IEO and its subscales on decision-making.

With risk aversion at the core of many theories about entrepreneurship (Van Praag, 1999), our main result regarding the negative relationship between IEO and decision-making under risk and uncertainty needs an explanation. This explanation may lie in the multidimensionality of the IEO construct because we do find that individuals self-reporting to be risk-taking indeed take more risks in the CCT (both versions). Thus, not surprisingly, individuals reporting a high propensity for taking (entrepreneurial) risks do take more financial risks in a laboratory experiment than those who report to be less risk-seeking. Our results, however, suggest that the joint effect of proactiveness and innovativeness (which are negatively related to risk-taking behaviour in the CCT) is larger than the effect of the risk-taking dimension of IEO. Being highly proactive may mean trying to reduce future uncertainties, resulting in a negative relationship between proactiveness and actual risky behaviour. Similarly, more innovative people may also prefer to mitigate or actively manage risks, to have better chances of bringing their new products on the market. As such, in line with other studies (e.g., Craig et al., 2014; Dai et al., 2014; Kreiser & Davis, 2010), our results underscore the importance for future studies to appreciate heterogeneity in the entrepreneurship construct as the subscales of IEO relate differently to decision-making under risk and uncertainty. The IEO subscales and other traits that have been shown to be relevant to entrepreneurship (such as autonomy and competitive aggressiveness; Lumpkin & Dess, 1996) together are what separates those who are entrepreneurially oriented and may become entrepreneurs from those who are not entrepreneurially oriented and will gravitate more towards other types of employment.

Because of their experimental set-up and identification strategy, it is most relevant to compare our findings to the results of the studies by Holm et al. (2013) and Koudstaal et al. (2016). Overall, these lab-in-the-field studies find that individuals indicating to be risk-averse do not exhibit particular risk-averse behaviour in lottery tasks. With the CCT, an experimental task known for its ecological validity, we do find a positive relationship between the self-reported risk-taking dimension of IEO and the number of selected cards in the CCT. Our results therefore suggest that there might be more value in self-reports about risk preferences than concluded by Holm et al. (2013) and Koudstaal et al. (2016) based on their lottery experiments. Moreover, the results suggest that the CCT provides a useful experimental set-up to study the relationship between decision-making under risk and uncertainty and EO and, by extension, other aspects of entrepreneurship.

The difference between our results and the results of Holm et al. (2013) and Koudstaal et al. (2016) may be explained not only by the different experimental paradigm used but also by the different entrepreneurship measures employed: students' IEO is not the same as actual entrepreneurial behaviour. When we on the one hand appreciate that students with high IEO exhibit intentions for entrepreneurship (e.g., Bolton & Lane, 2012) but that not all of them eventually become an entrepreneur and on the other hand take into account that some individuals without particularly strong entrepreneurial intentions during their academic studies may still engage in entrepreneurship later, it may be expected that the effect of IEO on decision-making under risk and uncertainty among students is larger than the relationship between actual entrepreneurial involvement and risk-taking behaviour. Thus, the sample composition could possibly explain why our statistical evidence is relatively strong. Obviously, it is important to verify this conjecture in future studies by replicating the present study design with actual entrepreneurs and preferably several control groups, such as managers, employees, and also students.

4.6 Supplementary Materials

Table S4.1

The 10 Individual Entrepreneurial Orientation (IEO) Scale Items (Bolton & Lane, 2012)

Subscale	Items	
Risk-Taking	1.	I like to take bold action by venturing into the unknown
	2.	I am willing to invest a lot of time and/or money on something that might yield a high return
	3.	I tend to act "boldly" in situations where risk is involved
Innovativeness	4.	I often like to try new and unusual activities that are not typical but not necessarily risky
	5.	In general, I prefer a strong emphasis in projects on unique, one-of-a-kind approaches rather than revisiting tried and true approaches used before
	6.	I prefer to try my own unique way when learning new things rather than doing it like everyone else does
	7.	I favor experimentation and original approaches to problem solving rather than using methods others generally use for solving their problems
Proactiveness	8.	I usually act in anticipation of future problems, needs or changes
	9.	I tend to plan ahead on projects
	10.	I prefer to "step-up" and get things going on projects rather than sit and wait for someone else to do it

Table S4.2

Results of the Regressions Explaining the Number of Cards Selected in a Trial, Excluding Individuals who Showed Unconventional Experimental Behaviour or Signs of Fatigue (Risk Sample)

	Model 1		Model 2		Model 3	
IEO	-0.731***	(0.070)			-0.853***	(0.161)
IEO Risk-Taking			0.386***	(0.083)		
IEO Innovativeness			-0.629***	(0.068)		
IEO Proactiveness			-0.570***	(0.067)		
IEO × Gain amount (30)					0.583***	(0.134)
IEO × Loss amount (750)					-0.529***	(0.136)
IEO × #Loss cards (3)					0.279*	(0.149)
Gain amount (30)	1.829***	(0.141)	2.040***	(0.131)	1.748***	(0.141)
Loss amount (750)	-2.476***	(0.143)	-2.614***	(0.134)	-2.409***	(0.143)
#Loss cards (3)	-4.543***	(0.155)	-4.178***	(0.148)	-4.590***	(0.155)
Task order (uncertainty first)	-1.210***	(0.142)	-1.287***	(0.134)	-1.233***	(0.141)
Sex (female)	-0.516***	(0.155)	-0.088	(0.145)	-0.473***	(0.152)
Age	0.029	(0.069)	0.052	(0.068)	0.034	(0.069)
Intercept	11.258***	(0.202)	10.746***	(0.190)	11.274***	(0.202)
Likelihood value	11953.02		12037.86		11935.84	
n	88		88		88	
n×*t*	4224		4224		4224	

Note. Coefficient with standard errors in parentheses; IEO = Individual Entrepreneurial Orientation; Reference categories are 10 for Gain amount, 250 for Loss amount, and 1 for #Loss cards; Continuous variables (IEO and its subscales and age) have been standardised to *Z*-scores; *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table S4.3

Results of the Regressions Explaining the Number of Cards Selected in a Trial, Excluding Individuals who Showed Unconventional Experimental Behaviour or Signs of Fatigue (Uncertainty Sample)

	Model 1		Model 2		Model 3	
IEO	-0.124**	(0.060)			-0.202	(0.131)
IEO Risk-Taking			0.184**	(0.074)		
IEO Innovativeness			-0.185***	(0.060)		
IEO Proactiveness			-0.113*	(0.061)		
IEO × Gain amount (30)					0.166	(0.121)
IEO × Loss amount (750)					-0.097	(0.127)
IEO × #Loss cards (3)					0.199*	(0.112)
Gain amount (30)	2.702***	(0.126)	2.702***	(0.126)	2.701***	(0.126)
Loss amount (750)	-3.614***	(0.131)	-3.591***	(0.131)	-3.608***	(0.131)
#Loss cards (3)	-0.111	(0.116)	-0.106	(0.116)	-0.115	(0.116)
Task order (uncertainty first)	1.842***	(0.121)	1.749***	(0.124)	1.842***	(0.121)
Sex (female)	0.732***	(0.124)	0.879***	(0.131)	0.733***	(0.124)
Age	-0.108	(0.066)	-0.128*	(0.067)	-0.106	(0.066)
Intercept	5.469***	(0.162)	5.414***	(0.163)	5.464***	(0.162)
Likelihood value	12547.41		12541.42		12544.39	
n	96		96		96	
n×*t*	4608		4608		4608	

Note. Coefficient with standard errors in parentheses; IEO = Individual Entrepreneurial Orientation; Reference categories are 10 for Gain amount, 250 for Loss amount, and 1 for #Loss cards; Continuous variables (IEO and its subscales and age) have been standardised to *Z*-scores; *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table S4.4

Results of the Regressions Explaining the Average Number of Points Earned Across 48 Trials (Risk Sample)

	Model 1		Model 2	
IEO	0.038	(0.089)		
IEO Risk-Taking			-0.035	(0.104)
IEO Innovativeness			0.023	(0.100)
IEO Proactiveness			0.069	(0.090)
Task order (uncertainty first)	0.241	(0.174)	0.246	(0.176)
Sex (female)	-0.529***	(0.178)	-0.563***	(0.186)
Age	0.121	(0.087)	0.126	(0.088)
Intercept	0.174	(0.149)	0.19	(0.152)
R-squared	0.09		0.10	
n	127		127	

Note. Coefficient with standard errors in parentheses; IEO = Individual Entrepreneurial Orientation; Continuous variables (IEO and its subscales and age) have been standardised to *Z*-scores; *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table S4.5

Results of the Regressions Explaining the Average Number of Points Earned Across 48 Trials (Uncertainty Sample)

	Model 1		Model 2	
IEO	0.132	(0.087)		
IEO Risk-Taking			-0.008	(0.101)
IEO Innovativeness			0.124	(0.096)
IEO Proactiveness			0.054	(0.087)
Task order (uncertainty first)	-0.479***	(0.168)	-0.463***	(0.171)
Sex (female)	-0.484***	(0.173)	-0.514***	(0.180)
Age	0.158*	(0.085)	0.158*	(0.086)
Intercept	0.501***	(0.145)	0.510***	(0.147)
R-squared	0.17		0.18	
n	124		124	

Note. Coefficient with standard errors in parentheses; IEO = Individual Entrepreneurial Orientation; Continuous variables (IEO and its subscales and age) have been standardised to *Z*-scores; *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table S4.6

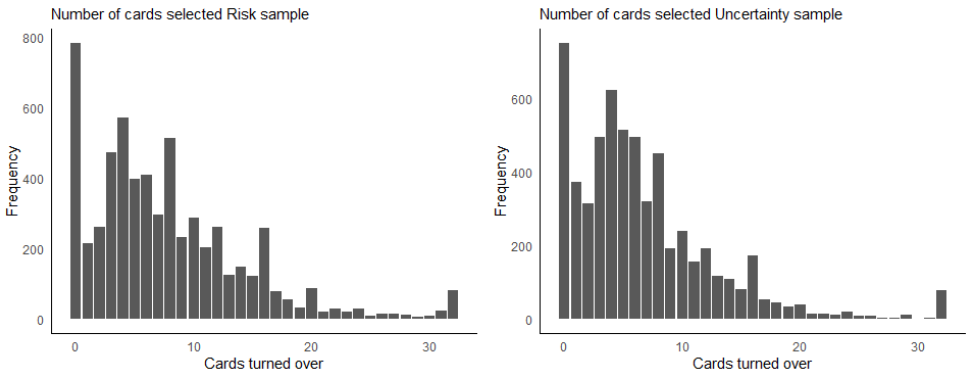
Results of the Regressions Explaining the Number of Cards Selected in a Trial, Excluding Individual Items From the Proactiveness Subscale

	Risk sample		Uncertainty sample	
IEO Risk-Taking	0.337***	(0.086)	0.088	(0.081)
IEO Innovativeness	-0.430***	(0.078)	-0.04	(0.062)
IEO Proactiveness – Item 8	-0.181**	(0.078)	-0.176**	(0.068)
IEO Proactiveness – Item 9	0.004	(0.084)	0.100	(0.065)
IEO Proactiveness – Item 10	-0.849***	(0.077)	-0.246***	(0.067)
Gain amount (30)	2.297***	(0.137)	3.288***	(0.132)
Loss amount (750)	-3.126***	(0.142)	-4.123***	(0.137)
#Loss cards (3)	-4.734***	(0.155)	-0.102	(0.115)
Task order (uncertainty first)	-1.519***	(0.147)	1.976***	(0.137)
Sex (female)	0.772***	(0.150)	1.164***	(0.134)
Age	-0.286***	(0.064)	-0.275***	(0.063)
Intercept	11.124***	(0.196)	5.532***	(0.164)
Likelihood value	17617.29		16615.35	
n	127		124	
n×*t*	6096		5952	

Note. Coefficient with standard errors in parentheses; IEO = Individual Entrepreneurial Orientation; Reference categories are 10 for Gain amount, 250 for Loss amount, and 1 for #Loss cards; Continuous variables (IEO and its subscales and age) have been standardised to *Z*-scores; *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Figure S4.1

Number of Cards Selected per Trial in the Risk Sample and Uncertainty Sample



PART II

Brain Activity

Chapter 5

Event-Related Potentials in Response to Feedback Following Risk-Taking in the Hot Version of the Columbia Card Task

Abstract

Given the importance of risk-taking in individuals' personal and professional life, several behavioural tasks for measuring the construct have been developed. Recently, a new task was introduced, the Columbia Card Task (CCT). This task measures participants' risk levels and establishes how sensitive participants are to gains, losses, and probabilities when taking risk. So far, the CCT has been examined in behavioural studies and in combination with several (neuro)biological techniques. However, no electroencephalography (EEG) research has been done on the task. The present study fills this gap and helps to validate this relatively new experimental task. To this end, n = 126 students were asked to complete self-reports (reward responsiveness, impulsiveness, and sensation-seeking) and to perform the CCT (and other risk tasks) in an EEG set-up. The results show that feedback appraisal after risky decision-making in the CCT was accompanied by a Feedback-Related Negativity (FRN) and a P300, which were stronger in response to negative than positive feedback. Correlations between the FRN and P300 difference wave on the one hand and risk-related self-reports and behaviour on the other were non-significant and small, but were mostly in the expected direction. This pattern did not change after excluding participants with psychiatric/neurological disorders and outliers. Excluding participants with reversed (positive > negative) difference waves strengthened FRN correlations. The impact such individuals can have on the data should be taken into account in future studies. Regarding the CCT in particular, future studies should also address its oddball structure and its masking of true values (censoring).

This chapter has been published as: De Groot, K., & Van Strien, J. W. (2019). Event-related potentials in response to feedback following risk-taking in the hot version of the Columbia Card Task. *Psychophysiology, 56*, Article e13390. https://doi.org/10.1111/psyp.13390

5.1 Introduction

Since deciding whether or not to take a risk can have large consequences for personal and professional ventures, risk-taking propensity is examined in multiple scientific fields, such as neuroscience, psychology, criminology, economics, and management. For this purpose, several (computerised) behavioural tasks for measuring the construct have been developed. Well-known tasks include the Iowa Gambling Task (IGT; Bechara et al., 1994), the Balloon Analogue Risk Task (BART; Lejuez et al., 2002), the Cambridge Gambling Task (CGT; Rogers et al., 1999), and the Game of Dice Task (GDT; Brand et al., 2005). Recently, Figner et al. (2009) introduced a new computer task to measure risk-taking propensity: the Columbia Card Task (CCT).

In the CCT, participants turn virtual cards from a 32-card array. Most of these cards are win cards, which earn participants points. However, a small number of cards are loss cards, which make them lose points. In every trial, participants are given three information parameters to help them decide how many cards to turn: (1) the number of points they gain when turning a win card, (2) the number of points they lose when encountering a loss card, and (3) the number of loss cards present in the trial. The way in which cards are turned differs across task versions. In the so-called "hot" CCT, participants turn the cards one by one (thereby accumulating points) until they voluntarily stop and cash the points or until they turn a loss card, at which point the specified loss amount is subtracted from the points earned (and the remainder is cashed). In the "cold" CCT, however, participants indicate at the start of every trial how many cards they want to turn, after which the computer determines the trial's outcome, unseen by the participant. The key difference between the versions is whether or not participants receive feedback following their choices. In the cold CCT, participants get to see only the final result of the game, which elicits deliberative decision-making. In the hot CCT, they receive feedback after every card turn, eliciting affective decision-making. Later work by Huang et al. (2013) introduced a third, "in-between" version: the "warm" CCT. Here, participants select the cards they would like to turn at the start of a trial and then press a button that prompts the selected cards to turn, thereby providing delayed feedback.

Regardless of which version, the CCT is characterised by two advantages. First, since the task gives participants information on the probability of losing (i.e., the number of loss cards present in a trial), it is an apparent risk task that leaves no room for conceptual ambiguity (De Groot & Thurik, 2018). This is different from, for example, the BART (where people are unaware of probabilities and therefore decide under uncertainty) and the IGT (where people learn the probabilities while progressing through the task and thus gradually

shift from making decisions under uncertainty to making decisions under risk). The second advantage of the CCT is its use of a so-called full factorial design, which independently varies the three information parameters given to participants (the number of points awarded for turning win cards, the number of points subtracted when turning a loss card, and the number of loss cards present) across trials so that all possible combinations are presented a given number of times. This design prevents risk and expected value from being confounded. In the IGT, BART, CGT, and GDT, (presumed) riskier options have a lower expected value than less risky options (Figner et al., 2009; Schonberg et al., 2011). This causes a decomposition problem, since it is unclear to what extent someone's observed level of risk-taking is driven by risk attitude (information on probabilities), sensitivity to reward (information on gains), or sensitivity to loss (information on losses). Its full factorial design enables the CCT to establish the extent to which these factors separately affect a participant's level of risk-taking.

The CCT's advantages have so far been used in several behavioural studies and in combination with various (neuro)biological techniques. For example, increased risk-taking in the cold CCT has been associated with higher impulsivity (Penolazzi et al., 2012) and more errors in an executive function task (Buelow, 2015). Increased risk-taking in the hot CCT has been related to high grandiosity (Brunell & Buelow, 2017) and reward responsiveness (Penolazzi et al., 2012). Using both the cold and hot CCT, several studies have found dissociations. For example, adolescents took more risk than adults in the hot but not in the cold CCT (Figner et al., 2009). A similar pattern was observed for patients with ventromedial prefrontal cortex (VMPFC) lesions compared to healthy controls (Spaniol et al., 2018). Biological dissociations have also been reported: electrodermal activity (EDA) only increased from baseline to decision phase in the hot and not the cold version of the task (Figner et al., 2009). Extending this finding, Holper and Murphy (2014) showed an opposite pattern for EDA and brain activity as measured with functional near-infrared spectroscopy (fNIRS): whereas skin conductance was larger in the hot than in the cold CCT, prefrontal total haemoglobin concentration (tHb) changes were larger in the cold version of the task. Another dissociation was reported in a study on hemispheric asymmetry using transcranial direct current stimulation (tDCS), showing that anodal left/cathodal right but not anodal right/cathodal left stimulation over the dorsolateral prefrontal cortex (DLPFC) decreased risk-taking in the cold CCT, which fits with the hypothesised involvement of the left DLPFC in deliberative information processing (Pripfl et al., 2013).

In addition to these findings on absolute risk levels, several studies have examined how individuals use the information (on gains, losses, and probabilities) that is provided to them in every trial. Distinctive patterns of information use have, for example, been observed in

adolescents (Figner et al., 2009) and older adults (Huang et al., 2015): both were shown to take less information into account than young and middle-aged adults when making decisions in the hot and warm CCT, respectively. Aberrant sensitivity to information has also been observed in several patient groups: compared to healthy controls, crack cocaine users (Kluwe-Schiavon et al., 2016), heroin-dependent persons (Saleme et al., 2018), and individuals with lesions in the VMPFC (Spaniol et al., 2018) paid less attention to probabilities. Decreased information sensitivity has also been reported in healthy individuals. For example, the positive association between the use of habitual cognitive reappraisal and risk-taking in the cold CCT as reported by Panno et al. (2013) was accompanied by reduced sensitivity to loss and probability information, suggesting that reappraisal operates via decreasing the attention to the negative aspects of a choice. Likewise, Penolazzi et al. (2012) showed that the association between reward responsiveness and risk-taking in the hot CCT interacted with information on gains and losses in such a way that individuals who scored high on reward responsiveness were sensitive to high gains while neglecting concomitant high loss. These studies, among others, clearly illustrate the benefits of the CCT's full factorial design by examining not only risk behaviour itself but also the motives that drive it.

Whereas many of the older decision tasks (such as the BART and the IGT) have been explored with electroencephalography (EEG), no EEG research has yet been done on the CCT. Previous EEG research on the BART (e.g., Kardos et al., 2016; L. Kessler et al., 2017; Takács et al., 2015) and the IGT (e.g., Mapelli et al., 2014; Oberg et al., 2014) primarily focused on the feedback phase of the tasks in which the rapid appraisal of the decision outcomes is usually captured by two Event-Related Potentials (ERPs): the Feedback-Related Negativity (FRN) and the feedback-related positivity 300 (P300). Since previous (neuro)biological research on the CCT employed measures with lower temporal resolution (EDA [Figner et al., 2009; Holper & Murphy, 2014], fMRI [Van Duijvenvoorde et al., 2015], and fNIRS [Holper & Murphy, 2014]) or focused on stimulating rather than recording brain activity (tDCS; Pripfl et al., 2013), examining the ERPs for the CCT could aid in validating this relatively new experimental task.

The first ERP of interest, the FRN, is a negative deflection peaking at frontocentral sites, and reaches its maximum 200-300 ms after feedback presentation (Holroyd & Coles, 2002; Miltner et al., 1997). Generation of the potential is closely linked to the mesolimbic dopaminergic system (Nieuwenhuis et al., 2004; Walsh & Anderson, 2012). When an outcome is worse than expected, mesencephalic dopaminergic firing decreases (Holroyd & Coles, 2002). These transient dopaminergic dips signal disinhibition of apical dendrites in the anterior cingulate cortex (ACC), which uses the signal to determine the most suitable

behaviour for the situation at hand. The FRN reflects an early and rapid bad versus good evaluation of feedback. Accordingly, it is influenced by only the valence and not by the magnitude of rewards, showing stronger amplitudes following negative than following positive feedback (Hajcak et al., 2006; Miltner et al., 1997; Yeung & Sanfey, 2004). With regard to risk-taking, stronger amplitudes have been related to increased risk aversion (Schuermann et al., 2012); blunted absolute and relative (difference) waves have been observed in individuals who typically take more risk, such as people dealing with borderline personality disorder (Endrass et al., 2016), family alcohol problems (Fein & Chang, 2008), or problematic internet use (Yau et al., 2015). These findings suggest a relationship between increased risk-taking and underdeveloped internal models and warning signals. In line with this, larger FRN difference waves have been associated with higher executive function (Kóbor et al., 2015) and a preference for low-risk decisions (Endrass et al., 2016).

The FRN is typically followed by the second ERP of interest: the P300, a positive, parietally distributed deflection that peaks approximately 300-500 ms after feedback presentation (Kopp & Wolff, 2000; Sutton et al., 1965). This potential is linked to the noradrenergic system and hence to locus coeruleus activity (Nieuwenhuis et al., 2005; Polich, 2007). Candidate regions for its neural basis are the cingulate cortex and adjacent areas involved in the circuit between frontal and parietal regions (Linden, 2005; Nieuwenhuis et al., 2005). Contrary to the FRN, the P300 reflects elaborate appraisal of feedback, varies with the motivational significance of this feedback (Kleih et al., 2010; Nieuwenhuis et al., 2005), and is sensitive to top-down attentional control (H. M. Gray et al., 2004; Nieuwenhuis et al., 2005; Polich, 2007). The literature is inconsistent as to whether the P300 is sensitive to valence. Some studies indicate no effect (Yeung & Sanfey, 2004), whereas others show a stronger P300 following positive feedback (Wu & Zhou, 2009; Z. Zhou, Yu, & Zhou, 2010) or negative feedback (Crowley et al., 2009; Endrass et al., 2016; Euser, Evans, et al., 2013; Euser, Greaves-Lord, et al., 2013; Fein & Chang, 2008; Kóbor et al., 2015; Schuermann et al., 2012). With regard to risk-taking, absolute amplitudes are larger for high-risk than for low-risk decisions (Endrass et al., 2016; Schuermann et al., 2012). These higher amplitudes, especially in response to negative feedback, are related to greater risk avoidance. Reduced (absolute and difference) waves are observed in risk-prone people, such as people who are alcohol-intoxicated (Euser et al., 2011) and individuals who have a parental history of substance abuse (Euser, Greaves-Lord, et al., 2013), show features of problematic internet use (Yau et al., 2015), or are diagnosed with borderline personality disorder (Endrass et al., 2016). Blunted absolute P300s may reflect a diminished ability to engage in feedback appraisal, outcome prediction, and cognitive control. Reduced difference scores can be the result of a heightened response to gains and/or a weaker response to losses.

The present study aids in validating the CCT by examining ERPs in response to feedback in the hot version of the task. The hot version is particularly suitable for this purpose, since the cold CCT does not provide feedback, and since the later-developed warm CCT (which does provide feedback) is still in a pioneering phase. First, we examine the FRN and P300, two ERPs that are commonly observed during feedback appraisal in other behavioural risk tasks, such as the BART and the IGT (L. Kessler et al., 2017; Oberg et al., 2011). Based on this literature, we expected that feedback appraisal in the hot CCT would also be accompanied by an FRN and a P300, and that both potentials would be more potent following negative than following positive feedback. Second, we examine the correlations between the observed CCT ERPs and risk-related self-reports and behaviour. Here, the absolute ERPs are transformed into difference waves by subtracting the positive feedback-locked waveform from the negative one (see, e.g., Fein & Chang, 2008; Kóbor et al., 2015). The advantage of doing so is twofold: it eliminates exogenous components – that is, elements that are elicited in response to all stimuli and hence across all conditions (Miltner et al., 1997) – and it corrects for individual differences in general wave amplitude, given that absolute waves may reflect a general tendency for small or large amplitudes, rather than the underlying construct (which is especially problematic for correlations). Correlations are calculated between the ERP difference waves and the following variables: behavioural measures derived from the hot CCT itself (average number of card turns; number of loss card encounters; sensitivity to gain, loss, and probability); risk-taking on the cold CCT and the BART (which has been shown to correlate with risk-taking on the hot CCT: Buelow & Blaine, 2015; Saleme et al., 2018); gender (which has been shown to correlate with several tasks and types of risk-taking and which has been related to the FRN and P300: Byrnes et al., 1999; Ding et al., 2017; Hirayasu et al., 2000); and three self-report constructs that have been related to either hot CCT behaviour itself or to behaviour and electrophysiology in other tasks probing affective decision-making: reward responsiveness (Penolazzi et al., 2012) and impulsivity and sensation-seeking (Euser et al., 2011; Lejuez et al., 2002). Based on this literature, we hypothesised that reduced FRN and P300 difference scores (i.e., smaller differences between responses to negative and positive feedback) are associated with the following: being male; higher risk-taking in the BART, the cold CCT, and the hot CCT; a higher number of loss cards encountered; higher reward responsiveness, impulsivity, and sensation-seeking; increased sensitivity to gains; and lower sensitivity to losses and probabilities.

5.2 Method

5.2.1 Participants

The sample consisted of n = 126 students (52.38% female) recruited from two universities, with a mean age of M = 21.01 (SD = 2.62), range 17-31 years. Most participants were studying social sciences (38.89%), economics (26.19%), or management (19.84%), although all main fields of study (including law, mathematics, and medicine) were represented. In exchange for participation, students received either course credit or a standard fee of €25. They were informed that they could earn extra money (up to €7.50) based on their task performance. All participants provided written informed consent. The study was approved by the institutional review board, and all procedures performed were in accordance with the 1964 Helsinki declaration and its later amendments.

5.2.2 Procedure

The measures were part of a larger study on decision-making under uncertain and risky conditions. Participants signed up online based on a brief description of the study design, after which they received an email with more elaborate information and the request to not drink alcohol, coffee, or energy drinks on the day of the appointment to prevent these substances from impacting the measurements. This email also contained a link to a web-based survey including the self-report measures, which participants were required to complete before their appointment at the laboratory. During the appointment itself, the procedure was explained to the participant, and they were asked to provide written informed consent. Then the participant was seated in a light- and sound-attenuated EEG room, was wired to the electrodes, and was presented with the BART and both CCTs. The order in which the tasks were presented was counterbalanced across participants. After finishing the tasks, the participant was debriefed. The full session lasted approximately 1.5 hours.

5.2.3 Self-Report Measures

5.2.3.1 Reward Responsiveness. Reward responsiveness (RR) was measured using the RR subscale of the Behavioural Approach System (BAS) questionnaire (Carver & White, 1994), which consists of five items (4, 7, 14, 18, and 23) and is answered on a 4-point scale with labels “completely disagree”, “disagree”, “agree”, and “completely agree”. The RR score ranges from 5 to 20, with higher scores indicating higher trait reward responsiveness. In the present study, scores ranged from 9 to 20. Cronbach’s alpha was $\alpha = 0.68$.

5.2.3.2 Impulsiveness. Impulsiveness was measured using the Barratt Impulsiveness Scale 11 (BIS-11; Patton et al., 1995), which consists of 30 items that are answered on a 4-point scale with labels “rarely/never”, “occasionally”, “often”, and “almost always/always”. The BIS-11 score ranges from 30 to 120, with higher scores being indicative of higher trait impulsiveness. Scores in the present study ranged from 43 to 87. Cronbach’s alpha was $\alpha = 0.76$.

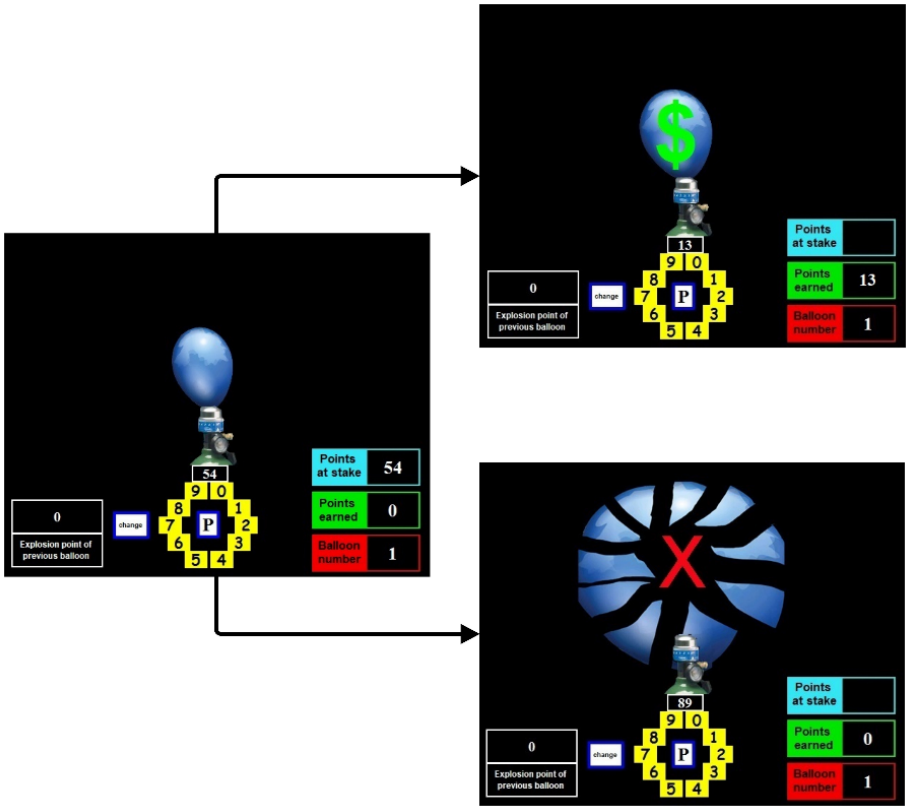
5.2.3.3 Sensation Seeking. Sensation seeking was measured using the Brief Sensation Seeking Scale (BSSS; Hoyle et al., 2002), which consists of eight items that are answered on a 5-point scale with labels “strongly disagree”, “disagree”, “neither disagree nor agree”, “agree”, and “strongly agree”. The BSSS scale ranges from 8 to 40, with higher scores indicating higher trait sensation-seeking. Scores in the present study ranged from 11 to 37. Cronbach’s alpha was $\alpha = 0.77$.

5.2.4 Behavioural Tasks

5.2.4.1 Automatic Balloon Analogue Risk Task (BART). In the automatic BART (Euser, Evans, et al., 2013; Euser, Greaves-Lord, et al., 2013; Pleskac et al., 2008; Yau et al., 2015), participants pump up a virtual balloon. More pumps equal more points but also increase the chance that the balloon pops, in which case all points accumulated in that trial are lost. The task set-up is presented in Figure 5.1. On the left side of the screen, the explosion point of the previous balloon is shown. On the right, three parameters are provided: how many points are at stake in the present trial (blue box), how many points have been accumulated so far (green box), and the trial number (red box). In the middle, a number dial is shown. Participants had to pump up 60 balloons. Via the number dial, participants first indicated how many times they wanted to pump the balloon, ranging from 1 to 128. Then, they pressed “P”, after which the balloon started inflating. After inflation, either positive or negative feedback followed. In case of positive feedback, the balloon remained intact and a green dollar sign appeared (Figure 5.1, upper right). In case of negative feedback, the balloon popped and a red cross appeared (Figure 5.1, lower right). The explosion likelihood distribution was equal for every possible pump, with an average of 64 pumps, and was the same for all participants. As in the original BART, participants were not informed about the explosion likelihood distribution. The variable of interest for the BART was the average number of pumps across all trials.

Figure 5.1

Task Set-Up of the Balloon Analogue Risk Task (BART)



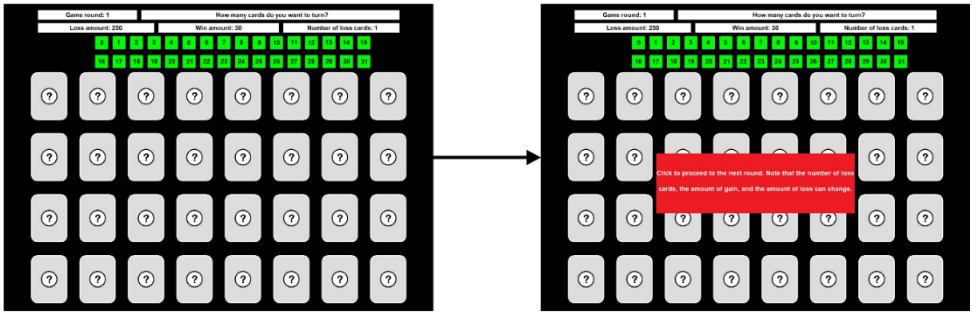
Note. The left screen shows the starting position. The upper right screen shows a situation in which the participant receives positive feedback. The lower right screen shows a situation where negative feedback is provided.

5.2.4.2 Cold and Hot Columbia Card Task (CCT). In the cold and hot CCT (Figner et al., 2009), participants are presented with a virtual array of 32 (4 × 8) cards. The majority of these cards earn the participant points (win cards), but in every array a small number of loss cards is hidden, for which points are subtracted. In the cold CCT, participants choose the total number of cards they would like to turn at the start of every round by clicking a number from 0 to 31. After selecting a number, a message appears, informing participants that they are continuing to the next round. This task set-up is presented in Figure 5.2. No feedback is provided during the cold CCT; participants are only informed about their final points after finishing the task. In the hot CCT, in contrast, participants turn over cards one by one in a self-paced manner and receive immediate feedback (i.e., without delay) in the

form of a happy face (win card) or a sad face (loss card). Participants can decide to stop turning cards at any point, terminating the round. However, if they encounter a loss card (which is shown for 2000 ms), the round terminates automatically, and the specified loss amount is subtracted from the points earned in that round. The task set-up of the hot CCT is presented in Figure 5.3. In both the cold and the hot CCT, participants are given three information parameters to help them decide how many cards to turn: (1) the number of points they gain when turning a win card, (2) the number of points they lose when turning a loss card, and (3) the number of loss cards hidden in a round. These parameters are presented at the top of the screen and are independently varied across trials by means of a full factorial design. The number of levels within a parameter differs across studies, with most studies using two levels per parameter: 250 or 750 loss, 10 or 30 win, and 1 or 3 chance (Brunell & Buelow, 2017; Buelow, 2015; Holper & Murphy, 2014; Huang et al., 2013; Panno et al., 2013; Penolazzi et al., 2012; Pripfl et al., 2013; Schumpe et al., 2017). In the present study, we ran this $2 \times 2 \times 2$ factorial six times, resulting in 48 trials ($[2 \times 2 \times 2] \times 6$). Given that the focus of the present study is on the hot CCT, the only variable of interest extracted from the cold CCT was the absolute risk level (i.e., the number of cards chosen). The variables of interest for the hot CCT were the average number of card turns; the number of loss card encounters; and sensitivity to gains, losses, and probabilities. Since data from the hot CCT are inherently censored (i.e., people's observed risk level in trials that forcedly end when turning a loss card does not necessarily reflect their true risk level), the variables of interest were in addition calculated using only data from trials in which participants voluntarily stopped turning cards.

Figure 5.2

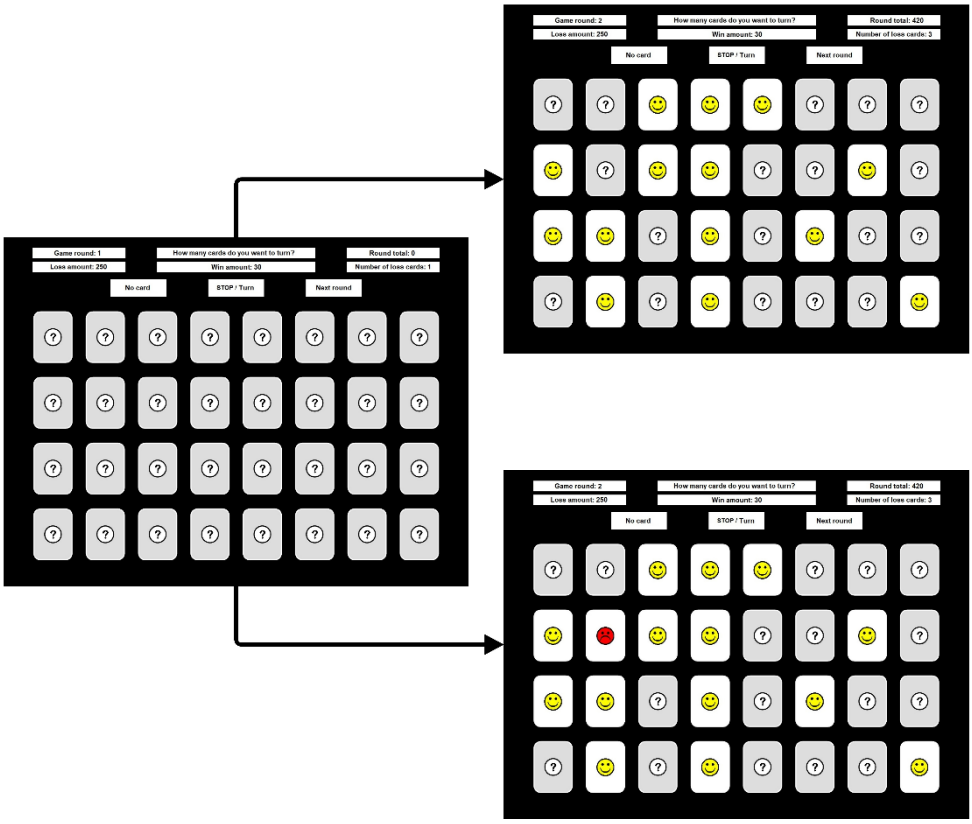
Task Set-Up of the Cold Columbia Card Task (CCT)



Note. The left screen represents the initial set-up in which the participant indicates how many cards he/she wants to turn by clicking a number from 0 to 31. The right screen shows the message participants see after selecting a number, which informs them that the next round is about to start and that the information parameters may change.

Figure 5.3

Task Set-Up of the Hot Columbia Card Task (CCT)



Note. The left screen represents the initial set-up the participant encounters. The upper right screen shows the set-up during turning the cards. The lower right screen shows the set-up when a loss card is encountered.

5.2.5 Electrophysiological Recordings and Signal Processing

EEG was recorded using a 32-channel amplifier and ActiveTwo data acquisition software. Ag/AgCl active electrodes were placed on the scalp by means of a head cap according to the 10-20 placing system. The electrooculogram (EOG) was recorded by placing flat electrodes above and below the left eye (vertical EOG) and at the outer canthi of both eyes (horizontal EOG). Two reference electrodes were placed on the mastoids. An active (common mode sense) and passive (driven right leg) electrode comprised a feedback loop for amplifier referencing. All signals were digitised with a sampling rate of 512 Hz.

The data were analysed offline using Brain Vision Analyzer. First, all EEG channels were referenced to the mathematically linked mastoid electrodes. Then we applied a high-pass filter of 0.10 Hz, a low-pass filter of 30.00 Hz, and a notch filter of 50.00 Hz (to filter out powerline artifacts). Data were segmented into epochs ranging from 100 ms before onset of the feedback presentation to 1000 ms after onset of the feedback presentation. Then ocular artifact correction (Gratton et al., 1983) and baseline correction (using the 100 ms pre-feedback presentation window) were applied. Finally, extreme amplitudes (below -75 μV or above 75 μV) were removed using automatic artifact rejection. The average number of segments used for calculating individuals' total ERP was M = 172.80 following positive feedback (a happy icon) and M = 8.52 following negative feedback (a sad icon). The time window used for analysing the FRN was 220-300 ms; the P300 was analysed across a time window of 300-450 ms. All epochs were averaged across midline (Fz, Cz, Pz, Oz) and adjacent (F3, F4, C3, C4, P3, P4, O1, O2) electrodes in order to minimise myogenic artifacts.

5.2.6 Analyses

The first aim of the study was examining whether feedback appraisal after risky decision-making in the CCT was accompanied by an FRN and P300, and whether these ERPs differed between positive and negative feedback. To this end, the averaged absolute and difference waves as recorded from 100 ms before to 1000 ms after feedback presentation were plotted. In addition, two repeated-measures analyses of variance (ANOVAs) (2 [valence: positive, negative] × 4 [cluster: frontal, central, parietal, occipital]) were performed to examine whether the ERP in response to positive feedback significantly differed from the ERP in response to negative feedback for the 220-300 FRN time period and for the 300-450 P300 time period. A Bonferroni-corrected 5% alpha level was used. Since repeated-measures ANOVAs are susceptible to violation of the sphericity assumption, this assumption was tested using Mauchly's test of sphericity. Pre-empting the findings, the observed violations were relatively severe ($\hat{\varepsilon} < 0.75$). Therefore, following the advice of Field (2013), these violations were corrected for by adjusting the degrees of freedom using the Greenhouse-Geisser estimate.

The second research aim was to examine the validity of the CCT by pairwise correlating the EEG difference waves with risk-related self-reports and behavioural constructs: gender; average number of hot CCT card turns; number of hot CCT loss card encounters; sensitivity to gain, loss, and probability information (calculated using per-trial correlations between gain/loss/probability values and the number of cards chosen); average number of cold CCT card turns; average number of BART balloon pumps; and self-reported

reward responsiveness, impulsiveness, and sensation-seeking. Given the relatively large number of correlations (22), we would expect one correlation to be wrongly marked as "significant". Because of the limitations associated with significance testing (J. Cohen, 1990), we also focused on the magnitude of the effects (the correlation coefficients).

Finally, three sets of robustness analyses were performed. The first robustness check examined whether the correlational findings lasted when the behavioural CCT data were solely based on trials in which participants voluntarily stopped turning cards (hence trials for which data was uncensored). A second check examined the robustness of the ERPs itself and their correlations with the self-reports and behavioural constructs when excluding (1) individuals who reported a current psychiatric or neurological disorder, (2) individuals whose FRN and/or P300 difference scores were "reversed" (i.e., the opposite of what was expected, namely a stronger response to positive than negative feedback), and (3) individuals with univariate and/or bivariate outlying values as detected via visual inspection of the histograms, boxplots, and scatterplots, and by checking for extreme (> |3.29|) standardised residuals. In a third check, we abandoned the difference wave approach and examined correlations between self-report/behavioural measures and absolute ERPs (i.e., the separate measures for FRN gain, FRN loss, P300 gain, and P300 loss). Given that a difference score is computed using absolute scores, its correlation may be conflated, and its interpretation may be unclear (Meyer et al., 2017). In particular, if the gain and loss ERPs correlate with each other but are correlated with risk-taking in opposite directions, individual correlations are suppressed when using difference scores. This last robustness check examined whether this was the case for the present data.

5.3 Results

5.3.1 Visual Representation and Interpretation of the ERPs

Figures 5.4 and 5.5 show the scalp distributions of respectively the FRN and the P300 for positive feedback, negative feedback, and the difference wave. FRN activity peaked at frontocentral sites, while its difference activity was located more parietally. P300 activity showed a central-parietal distribution. The grand averaged ERP waveforms are presented in Figure 5.6. The waveform appeared robust, with a clear FRN in the 220-300 ms window and a clear P300 in the 300-450 ms window, both of which were stronger following negative than following positive feedback. These observations were confirmed by the repeated-measures ANOVAs, which were (after discarding ERP segments in the preprocessing phase)

based on data from n = 121 individuals. For the FRN, the Greenhouse-Geisser correction was applied to the main effect of cluster ($\chi^2(5) = 269.65$, $p < 0.001$, $\hat{\varepsilon} = 0.50$) and to the interaction ($\chi^2(5) = 214.20$, $p < 0.001$, $\hat{\varepsilon} = 0.56$). The main effects showed that the potential was stronger in response to negative than to positive feedback (Δ2.70 μV, $F(1, 120) = 31.72$, $p < 0.001$, $\eta_p^2 = 0.21$), and that it differed across clusters ($F(1.50, 179.96) = 25.55$, $p < 0.001$, $\eta_p^2 = 0.18$). These factors interacted as well ($F(1.68, 201.97) = 28.24$, $p < 0.001$, $\eta_p^2 = 0.19$): the effect of valence was strongest at central (Δ4.53 μV) and parietal (Δ3.77 μV) electrodes, and weaker at frontal (Δ1.24 μV) and occipital (Δ1.25 μV) ones. For the P300, the Greenhouse-Geisser correction was again applied to the main effect of cluster ($\chi^2(5) = 215.28$, $p < 0.001$, $\hat{\varepsilon} = 0.62$) and to the interaction ($\chi^2(5) = 204.98$, $p < 0.001$, $\hat{\varepsilon} = .60$). The main effects confirmed that the potential was stronger in response to negative than to positive feedback (Δ10.57 μV, $F(1, 120) = 352.87$, $p < 0.001$, $\eta_p^2 = 0.75$), and that it differed across clusters ($F(1.87, 223.94) = 85.60$, $p < 0.001$, $\eta_p^2 = 0.42$). The interaction ($F(1.80, 215.69) = 126.95$, $p < 0.001$, $\eta_p^2 = 0.51$) showed that the difference in valence was weaker at occipital (Δ4.90 μV) than at central (Δ14.57 μV), frontal (Δ11.72 μV), and parietal (Δ11.09 μV) sites.

Figure 5.4

Topographical Distribution of FRN Activity Across the Scalp for Positive Feedback, Negative Feedback, and the Difference Between These Two (Negative Minus Positive)

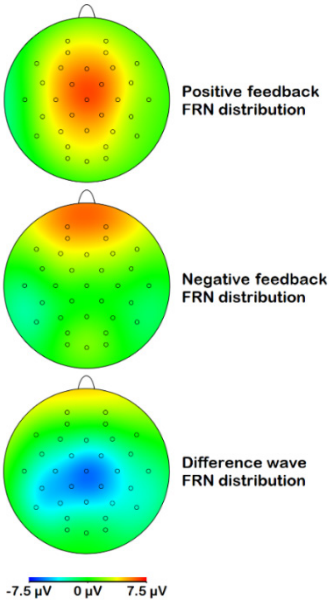


Figure 5.5

Topographical Distribution of P300 Activity Across the Scalp for Positive Feedback, Negative Feedback, and the Difference Between These Two (Negative Minus Positive)

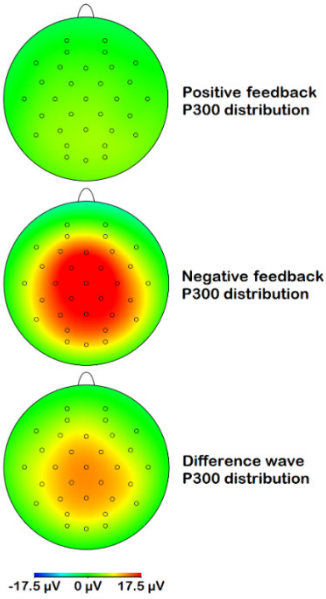
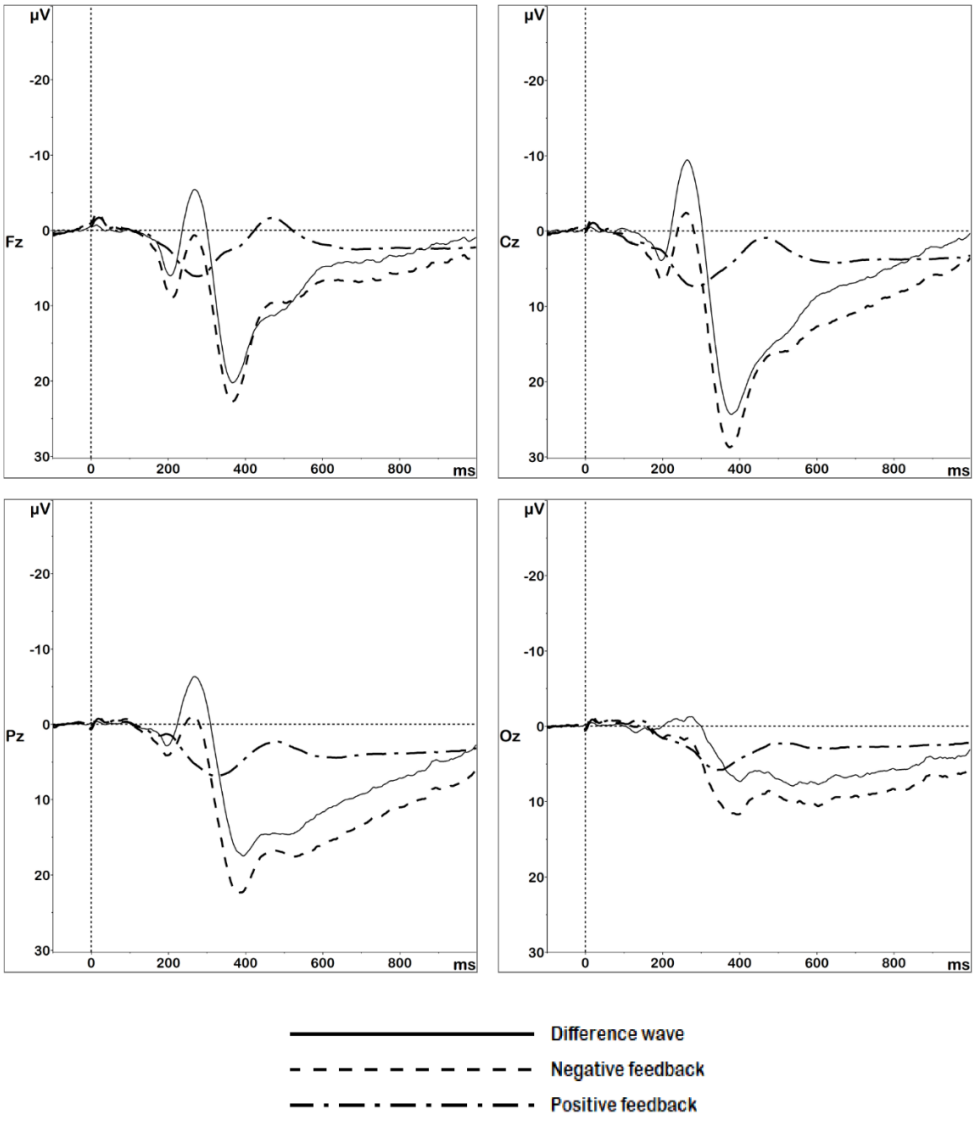


Figure 5.6

Grand Averaged ERPs for the Hot Columbia Card Task (CCT)



5.3.2 Correlational Analyses

Table 5.1 shows the correlations between the FRN and P300 difference waves on the one hand and the risk-related self-reports and behavioural constructs on the other. The correlations involving gender are point-biserial (r_{pb}); the remaining correlations are bivariate (r). Analyses were performed pairwise on an n between 119 and 121, with most missing values resulting from incomplete surveys or missing ERP data. None of the correlations was significant at a 5% level, which was fewer than the one significant correlation that was expected by chance. The majority of the effects were small but in the expected direction.

Since the FRN is a negative potential, its difference score is generally negative. Therefore, we expected a positive correlation with risk-taking (i.e., the more negative the FRN difference score, the smaller self-reported or behavioural risk-taking, and vice versa). In line with this, smaller FRN difference scores (reflecting reduced electrophysiological response to the feedback) were correlated with encountering more loss cards in the hot CCT (r = 0.05), choosing more cards in the cold (r = 0.10) and hot (r = 0.03) CCT, pressing more balloon pumps in the BART (r = 0.09), being less sensitive to information on losses (r = 0.14), and reporting higher impulsiveness (r = 0.13) and sensation-seeking (r = 0.09). Three correlations were in the opposite direction of what was expected, namely those for gender (r = 0.09), sensitivity to gains (r = -0.10), and reward responsiveness (r = -0.08).

Since the P300 is a positive potential, its difference scores are generally positive, and hence negative correlations were expected between this difference score and risk-taking (i.e., the more positive the P300 difference score, the smaller self-reported or behavioural risk-taking, and vice versa). In line with this, smaller P300 difference scores (reflecting reduced electrophysiological response to the feedback) were correlated with encountering more loss cards in the hot CCT (r = -0.06), being less sensitive to information on losses (r = -0.05) and probabilities (r = -0.05), and reporting higher reward responsiveness (r = -0.16), impulsiveness (r = -0.03), and sensation seeking (r = -0.04). Four correlations were in the opposite direction of what was expected, namely those for gender (r = -0.04), sensitivity to gains (r = 0.16), cold CCT cards chosen (r = 0.07), and BART balloon pumps (r = 0.13).

Table 5.1

Correlations Between the FRN and P300 Difference Waves and Risk-Related Self-Reports and Behavioural Constructs

					Correlations	
	M	*SD*	Min.	Max.	FRN CCT diff. wave	P300 CCT diff. wave
Gender (male = 0)	0.52	0.50	0.00	1.00	0.09[a]	-0.04[a]
Hot CCT average card turns	7.25	2.18	1.65	12.65	0.03	-0.01
Hot CCT loss card encounters	22.21	8.31	6.00	44.00	0.05	-0.06
Hot CCT gain sensitivity	0.16	0.18	-0.22	0.57	-0.10	0.16
Hot CCT loss sensitivity	-0.22	0.18	-0.64	0.22	0.14	-0.05
Hot CCT probability sensitivity	-0.51	0.12	-0.86	-0.11	0.01	-0.05
Cold CCT number of chosen cards	8.57	3.71	1.60	21.81	0.10	0.07
BART number of chosen pumps	58.49	11.72	13.42	85.25	0.09	0.13
Reward Responsiveness	16.63	2.15	9.00	20.00	-0.08	-0.16
Impulsiveness	65.46	8.32	43.00	87.00	0.13	-0.03
Sensation Seeking	25.48	5.62	11.00	37.00	0.09	-0.04

Note. [a] indicates point-biserial correlations (r_{pb}).

5.3.3 Robustness Checks

The first robustness analysis examined whether the correlational findings lasted when only using uncensored behavioural CCT data. To this end, eight correlations were rerun, namely those including the average number of card turns in the hot CCT, gain sensitivity, loss sensitivity, and probability sensitivity. Half of these correlations (most relatively large) remained similar in size and direction; three (all of them small) changed direction; and one changed substantially in size, namely the correlation between hot CCT card turns and the P300 difference wave, which changed from -0.01 to -0.10 and thereby became more in line with the hypothesis that a smaller P300 difference wave is associated with taking more risk (see Table S5.1).

A second set of robustness checks examined the effect of participant exclusion (Tables S5.2 to S5.4, Figures S5.1 to S5.3). The exclusion of individuals with psychiatric or neurological disorders (n = 5) did not impact the direction of the correlations, nor did it substantially change correlation size. The same was true for exclusion of individuals with outlying values (n = 5), except for the correlation of the P300 difference wave with impulsiveness (which changed direction: from r = -0.03 to r = 0.05) and with the BART (which turned significant: from r = 0.13 to r = 0.20). Neither exclusion impacted the grand averaged waveform or the topographical distribution of the ERPs. The largest change in findings was observed when excluding individuals with reversed difference scores (n = 48).

This caused a moderate increase in FRN amplitude and in most FRN difference score correlations, which became more in line with our hypotheses, such as the correlation with hot CCT card turns (from $r = 0.03$ to $r = 0.20$), loss card encounters (from $r = 0.05$ to $r = 0.20$), and loss sensitivity (from $r = 0.14$ to $r = 0.30$). The correlations for the P300 difference wave, however, changed in a less consistent manner and overall became less in line with our hypotheses. Since most exclusions resulted from reversed FRN scores, the impact on P300 correlations may indeed have been more equivocal (by also discarding "regular" scores).

The third and final set of robustness checks examined whether the use of difference scores suppressed the correlations between the absolute ERPs and the risk-related self-reports and behavioural constructs. The gain and loss FRN ERPs were correlated with each other ($r = 0.31$), as were the gain and loss P300 ERPs ($r = 0.37$). However, few of the gain and loss ERPs correlated with risk-related constructs in opposite direction (5 out of 11 for the FRN and 3 out of 11 for the P300), hence indicating no major risk of conflated difference scores. Examination of the individual correlations (Table S5.5) showed that the majority of FRN loss and P300 gain correlations were not in line with expectations, and hence did not outperform the correlations based on difference scores. However, most FRN gain and P300 loss correlations were in line with expectations. A stronger FRN in response to gains was associated with being male ($r = 0.19$), higher self-reported and behavioural risk-taking ($r = -0.02$ to $r = -0.23$), and lower focus on loss ($r = -0.19$) and probability ($r = -0.10$) information. A stronger P300 in response to losses was associated with being female ($r = 0.05$), lower self-reported and behavioural risk-taking ($r = -0.06$ to $r = -0.13$), and a stronger focus on information on loss ($r = -0.10$) and probability ($r = -0.07$). Thus, the FRN gain and P300 loss correlations provided information beyond that offered by the correlations for the difference scores.

5.4 Discussion

The present study examined ERPs in response to feedback in the hot version of the CCT (Figner et al., 2009). In line with research on feedback-related ERPs in for example the IGT and the BART, feedback appraisal in the CCT was accompanied by a clear FRN and P300, which were stronger in response to losses than to gains. This pattern did not change after excluding individuals with psychiatric or neurological disorders, individuals with outlying scores, or individuals with reversed (positive > negative) ERP difference scores. Hence, the ERPs appeared robust. Despite this, correlations between the ERP difference waves and risk-related self-reports and behavioural measures were non-significant and small. Most correlations did show an effect in the expected direction though: for example,

smaller FRN difference scores were associated with taking more risk in the cold CCT, decreased sensitivity to information on losses, and higher impulsiveness; smaller P300 difference scores were most strongly associated with higher reward responsiveness. When correlating absolute instead of difference scores, the FRN gain and P300 loss (but not the FRN loss and P300 gain) also showed effects in the expected direction. Excluding individuals with reversed ERP difference waves strengthened most FRN correlations, thereby bringing the findings more in line with the hypotheses. Several possible explanations exist why individuals show such a reversed pattern. First, they may respond very weakly to losses and/or very strongly to gains, resulting in a more potent ERP in response to gains than to losses. Alternatively, these reversed waves may result from individuals' expectations. Especially the FRN has been suggested to represent a reward prediction error, an indicator of the difference between expected and observed outcomes (Holroyd & Coles, 2002; Sambrook & Goslin, 2015). Hence, encountering losses may elicit only modest ERPs when they are expected, compared to cases in which they are not. Given the impact participants with reversed scores had on the correlations in the robustness analyses, future studies may want to address this phenomenon and examine its underlying causes.

This study's results combined with previous findings also illustrate some challenges that are more specific to the CCT and that would benefit from further investigation. First, for studies using the CCT in combination with EEG, it will be key to have a more elaborate understanding of which (task) characteristics influence the ERPs, and whether such influence is desirable. In addition to the large sample size, one factor possibly influencing the strength and robustness of the CCT's potentials concerns the reward structure. Risk tasks can offer participants different types of incentive, such as monetary rewards (Xu et al., 2016), social rewards (Op de Macks et al., 2017), or sexual rewards (Lawyer, 2013). Moreover, tasks differ in how participants are penalised when losing a trial: it may result in not receiving any reward, or in losing a reward that was acquired before. In the CCT, the points participants earn are truly at stake since turning a loss card means that the specified loss amount is subtracted from the points earned. Furthermore, participants in the present study were offered real (vs. hypothetical) money, which increased the ecological validity of the task and which has been shown to elicit a stronger FRN for negative feedback (Xu et al., 2016). This combination of losing points that represent real money may induce a larger prediction error and therefore stronger ERPs. This contribution to the presently observed robust ERPs can be deemed desirable as it reflects the constructs presumed to underlie these ERPs. A second factor that may have influenced the ERPs seems less desirable: the stimulus-sequence history. In the CCT, most loss-card encounters are preceded by a series of win card encounters, with losses being roughly 20 times less frequent than wins. Such oddball

structures have been shown to impact ERPs: FRN amplitudes tend to be larger when successive encounters with a stimulus are followed by feedback of opposite valence compared to feedback of the same valence (Holroyd & Coles, 2002), and P300 amplitudes are larger after presenting a deviant or salient stimulus, especially when this stimulus is preceded by a series of other stimuli (Nieuwenhuis et al., 2005; K. C. Squires et al., 1976). The oddball-like structure of the CCT can reasonably have contributed to the strength of the ERPs (especially the negative-feedback waveform), thereby adding unintended systematic variance that future studies may want to mitigate.

In addition to this EEG-related concern, the hot version of the CCT poses a more general challenge: censoring. As briefly discussed in the method section, data from the hot CCT are inherently censored as people's observed risk level in trials in which they turn a loss card does not necessarily reflect their true risk level, since they might have taken more risk (i.e., turned more cards) if they had had the chance to do so. In the present study, we accommodated for censoring by running two sets of analyses: the main analyses, using data from all trials; and robustness analyses, using only data from trials in which participants had voluntarily stopped turning cards. A similar approach was employed by Kluwe-Schiavon et al. (2015), who reported no major changes in their final results. In the present study, one of eight rerun correlations substantially changed in size, demonstrating the effect that censoring can have on a study's findings. An alternative solution to censoring is offered by Figner et al. (2009), who prevented censoring ex ante by rigging the task. In their task set-up, 54 experimental trials are supplemented by nine trick trials. In the experimental trials, the loss card is programmed to be the last possible card, so that participants never encounter it and so that all stopping points are voluntary. These uncensored data are used for analysis. Credibility is upheld by randomly interspersing the nine trick trials, which are programmed in such a way that participants quickly encounter a loss card. However, Figner et al.'s (2009) solution to censoring seems problematic, as the rigged percentage of trials in which participants encounter a loss card seems unrealistically low: $9 \div (54 + 9) \times 100 \approx 14.29$. In the present unrigged study, participants on average encountered a loss card in $22.21 \div 48 \times 100 \approx 46.27$ percent of trials, which was shown to be significantly higher using a one sample *t*-test: $t(125) = 20.73$, $p < 0.001$. Arguably, presenting people with (too) little negative feedback could cause them to take more risk. Tentative evidence for this conjecture is found in the large difference between risk levels found in studies using a rigged CCT (~23 cards [Figner et al., 2009], 27 cards [Markiewicz & Kubińska, 2015], and 21 cards [Penolazzi et al., 2012]) and studies in which the cards are truly shuffled (7.25 [8.48 uncensored] in the present study, and ~12 in Holper and Murphy, 2014).

A final, easier solution to censoring is omitting the hot CCT altogether and instead using the warm CCT, which also measures affective risk-taking but delays the feedback, so that participants first decide how many cards they want to turn and only then observe the (per-card) outcome of their decision. Although this does solve the issue of censoring, it offers no solution to the other challenge we discussed with regard to the CCT, that is, its oddball structure. Notably, the rigged design by Figner et al. (2009) does solve the oddball problem (at least within the trial). Whereas loss card encounters in an unrigged design are generally preceded by a series of win cards, loss card encounters in Figner et al.'s (2009) design are artificially positioned at the start of trials, thereby mitigating the oddball effect. This does not change the fact that win cards are more frequent than loss cards across the task but does impact the local probability. Therefore, if the percentage of rigged trials in which participants encounter a loss card in Figner et al.'s (2009) design were set to a more plausible, naturalistic value, both challenges observed in the present study would be resolved: the influence of the hot CCT's oddball structure on its ERPs and its masking of true values in trials that forcedly end (censoring). We recommend future studies to keep these two challenges and our proposed solutions to them in mind when further validating or using the CCT either in a behavioural study or in combination with a (neuro)biological measure such as EEG.

5.5 Supplementary Materials

Table S5.1

Correlations Using Uncensored Behavioural CCT Data

					Correlations	
	M	*SD*	Min.	Max.	FRN CCT diff. wave	P300 CCT diff. wave
Hot CCT average card turns	8.49	4.40	1.45	27.00	-0.03	-0.10
Hot CCT gain sensitivity	0.21	0.27	-0.61	0.85	-0.11	0.13
Hot CCT loss sensitivity	-0.31	0.24	-0.81	0.54	0.17	0.04
Hot CCT probability sensitivity	-0.58	0.18	-0.94	-0.07	-0.04	-0.06

Table S5.2

Correlations After Excluding n = 5 Participants with a Current Psychiatric or Neurological Disorder

					Correlations	
	M	*SD*	Min.	Max.	FRN CCT diff. wave	P300 CCT diff. wave
Gender (male = 0)	0.53	0.50	0.00	1.00	0.12[a]	-0.03[a]
Hot CCT average card turns	7.31	2.15	1.65	12.65	0.05	-0.02
Hot CCT loss card encounters	22.39	8.23	6.00	44.00	0.07	-0.06
Hot CCT gain sensitivity	0.16	0.17	-0.22	0.57	-0.12	0.18
Hot CCT loss sensitivity	-0.22	0.18	-0.64	0.22	0.14	-0.07
Hot CCT probability sensitivity	-0.52	0.13	-0.86	-0.11	<|0.01|	-0.05
Cold CCT number of chosen cards	8.56	3.66	1.60	21.81	0.08	0.07
BART number of chosen pumps	58.52	11.89	13.42	85.25	0.09	0.14
Reward Responsiveness	16.60	2.16	9.00	20.00	-0.10	-0.14
Impulsiveness	65.31	8.38	43.00	87.00	0.12	-0.04
Sensation Seeking	25.42	5.64	11.00	37.00	0.07	-0.06

Note. [a] indicates point-biserial correlations (r_{pb}).

Table S5.3

Correlations After Excluding n = 5 Participants with Outlying Scores

					Correlations	
	M	*SD*	Min.	Max.	FRN CCT diff. wave	P300 CCT diff. wave
Gender (male = 0)	0.52	0.50	0.00	1.00	0.09[a]	-0.04[a]
Hot CCT average card turns	7.31	2.13	2.54	12.65	0.03	-0.06
Hot CCT loss card encounters	22.15	7.90	6.00	44.00	0.06	-0.07
Hot CCT gain sensitivity	0.16	0.18	-0.22	0.57	-0.10	0.16
Hot CCT loss sensitivity	-0.21	0.18	-0.64	0.22	0.14	-0.08
Hot CCT probability sensitivity	-0.52	0.12	-0.86	-0.19	0.04	-0.04
Cold CCT number of chosen cards	8.48	3.57	1.60	16.94	0.10	0.11
BART number of chosen pumps	58.98	10.93	25.37	85.25	0.12	0.20*
Reward Responsiveness	16.76	1.94	10.00	20.00	-0.06	-0.15
Impulsiveness	65.46	8.13	46.00	87.00	0.16	0.05
Sensation Seeking	25.66	5.65	11.00	37.00	0.09	-0.06

Note. [a] indicates point-biserial correlations (r_{pb}); * is significant at a 5% level.

Table S5.4

Correlations After Excluding n = 48 Participants with Reversed FRN and/or P300 Difference Scores

					Correlations	
	M	*SD*	Min.	Max.	FRN CCT diff. wave	P300 CCT diff. wave
Gender (male = 0)	0.47	0.50	0.00	1.00	0.12[a]	0.01[a]
Hot CCT average card turns	7.33	1.98	2.54	11.63	0.20	0.06
Hot CCT loss card encounters	22.42	8.00	6.00	44.00	0.20	-0.07
Hot CCT gain sensitivity	0.18	0.16	-0.21	0.57	-0.14	0.17
Hot CCT loss sensitivity	-0.21	0.18	-0.61	0.14	0.30**	-0.13
Hot CCT probability sensitivity	-0.52	0.12	-0.86	-0.11	-0.02	-0.05
Cold CCT number of chosen cards	8.57	3.75	1.60	21.81	0.09	0.03
BART number of chosen pumps	57.63	11.24	13.42	81.95	<|0.01|	0.27*
Reward Responsiveness	16.71	2.13	9.00	20.00	0.04	0.12
Impulsiveness	64.83	8.06	46.00	83.00	0.16	0.13
Sensation Seeking	25.14	5.82	11.00	37.00	0.12	0.04

Note. a indicates point-biserial correlations (rpb); * is significant at a 5% level; ** is significant at a 1% level.

Table S5.5

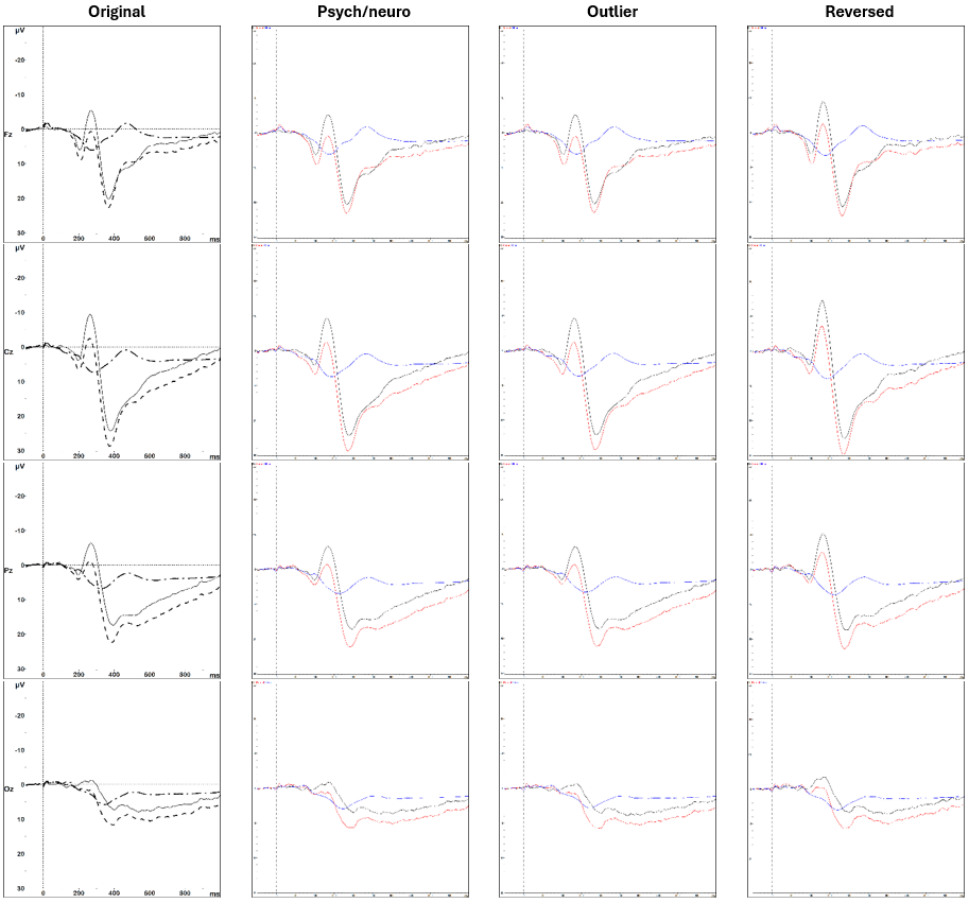
Correlations Using Absolute ERPs

					Correlations			
	M	*SD*	Min.	Max.	FRN gain	FRN loss	P300 gain	P300 loss
Gender (male = 0)	0.52	0.50	0.00	1.00	0.19[a]*	0.17[a]	0.25[a]**	0.05[a]
Hot CCT average card turns	7.25	2.18	1.65	12.65	-0.23**	-0.09	-0.18*	-0.09
Hot CCT loss card encounters	22.21	8.31	6.00	44.00	-0.22*	-0.05	-0.14	-0.11
Hot CCT gain sensitivity	0.16	0.18	-0.22	0.57	0.13	-0.01	0.10	0.21*
Hot CCT loss sensitivity	-0.22	0.18	-0.64	0.22	-0.19*	0.07	-0.14	-0.10
Hot CCT probability sensitivity	-0.51	0.12	-0.86	-0.11	-0.10	-0.04	-0.08	-0.07
Cold CCT number of chosen cards	8.57	3.71	1.60	21.81	-0.05	0.08	-0.05	0.04
BART number of chosen pumps	58.49	11.72	13.42	85.25	-0.02	0.09	-0.06	0.10
Reward Responsiveness	16.63	2.15	9.00	20.00	-0.06	-0.09	0.01	-0.13
Impulsiveness	65.46	8.32	43.00	87.00	-0.22*	0.03	-0.25**	-0.11
Sensation Seeking	25.48	5.62	11.00	37.00	-0.14	<|0.01|	-0.06	-0.06

Note. [a] indicates point-biserial correlations (r_{pb}); * is significant at a 5% level; ** is significant at a 1% level.

Figure S5.1

Effect of Exclusions on the ERPs



Note. The dash-dotted line (“original” panels) / blue line (other panels) represents the electrophysiological response to positive feedback; the dashed line (“original” panels) / red line (other panels) represents the electrophysiological response to negative feedback; and the solid line (“original” panels) / black line (other panels) is the calculated difference wave.

Figure S5.2

Effect of Exclusions on the Spatial Distribution of the FRN

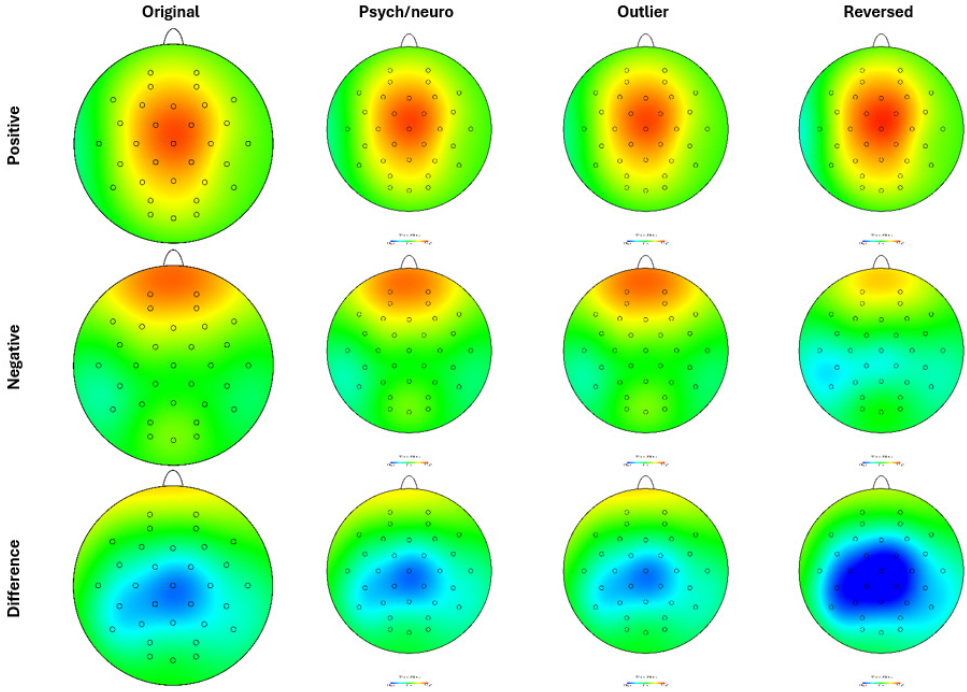
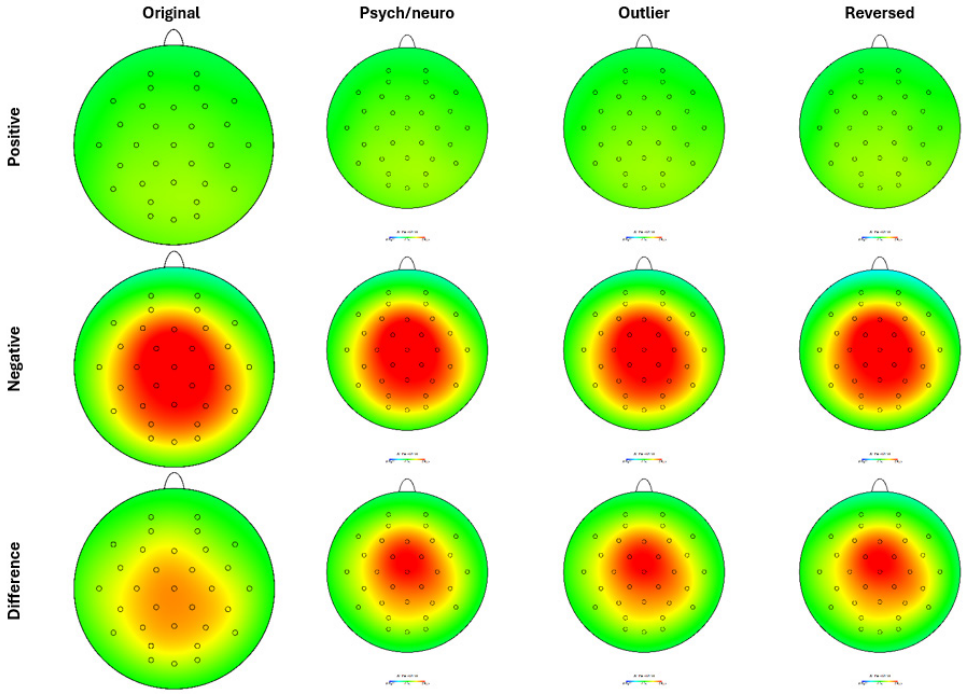


Figure S5.3

Effect of Exclusions on the Spatial Distribution of the P300



Chapter 6

Birds of a Feather Flock Together: Evidence of Prominent Correlations Within but not Between Self-Report, Behavioural, and Electrophysiological Measures of Impulsivity

Abstract

Despite many studies examining a combination of self-report, behavioural, 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, behavioural, 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 behavioural 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 behaviour and/or electrophysiology. Therefore, we suggest using sufficiently large samples when investigating associations between different levels of measurement.

This chapter has been published as: Bernoster, I., De Groot, K., Wieser, M. J., Thurik, R., & Franken, I. H. (2019). Birds of a feather flock together: Evidence of prominent correlations within but not between self-report, behavioural, and electrophysiological measures of impulsivity. *Biological Psychology, 145*, 112-123. https://doi.org/10.1016/j.biopsycho.2019.04.008

6.1 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 et al., 2001, p. 1784). It is a normal aspect of behaviour which is often functional, but can also be dysfunctional. Impulsivity is a multidimensional construct (Gerbing et al., 1987; Khadka et al., 2017; Meda et al., 2009), and is closely related to other constructs such as the Behavioural Activation System (BAS). BAS is in turn associated with reward responsiveness (Carver & White, 1994), which consists of reward sensitivity and rash impulsiveness (Dawe et al., 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 (H. A. 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 et al., 2018): various facets of impulsivity such as urgency and a lack of premeditation and perseverance characterise for example Attention Deficit/Hyperactivity Disorder (ADHD; Lopez et al., 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), behavioural measures (e.g., Sharma et al., 2014), and neurophysiological measures such as electroencephalography (EEG; e.g., Taylor et al., 2018). However, only few studies address whether these different levels of measurement (i.e., self-report, behaviour, 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 et al. (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 behavioural measures (acquired from the dot-probe task), nor between electrophysiology and self-reports (obtained from the Pain Catastrophising Scale, Pain Anxiety Symptoms Scale, and Pain Hypervigilance and Awareness Questionnaire). With respect to behaviour 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 et al., 2015). Self-report measures consisted of two parental reports: the Child Behaviour Questionnaire and Child Behaviour Checklist. Further, children performed 15 behavioural 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 behavioural 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 behavioural 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, behaviour, 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 behavioural outcomes, such as decreased behavioural inhibition in a Go/No-Go task (Littel, Van den Berg, et al., 2012), increased uncertain decision-making in the Balloon Analogue Risk Task (BART; Lauriola et al., 2014; Lejuez et al., 2002), and slower stopping reaction times in a stop-signal task (Logan et al., 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, Van den Berg, et al., 2012), and on punishment (Potts, George, et al., 2006; Potts, Martin, et al., 2006) or incorrect (Luijten et al., 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 et al., 2014), the continuous performance task (Kam et al., 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 et al., 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 (Z. H. Zhou, Yuan, et al., 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 et al., 2003), reduced ERN amplitudes in the Eriksen Flanker task (Zheng et al., 2014), and reduced P3 amplitudes in a passive oddball paradigm (W. Wang & Wang, 2001). Furthermore, self-reported reward responsiveness has been related to shorter reaction times on the Go/No-Go task (De Pascalis et al., 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 et al., 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 et al., 2013; Jonkman et al., 2007; Van Meel et al., 2007; Wiersema et al., 2005), and to attenuated P3 (Liotti et al., 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 et al., 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 behavioural 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, behaviour, and electrophysiology). Most studies investigate the association between a *single* self-reported construct and either behavioural 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 et al., 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 et al., 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 et al., 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, behavioural 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 et al., 2014), and Parkinson's disease (Chaudhuri et al., 2011), as well as in addiction and pathological gambling (Limbrick-Oldfield et al., 2013). Furthermore, impulsivity is a rather well-suited construct for examining the associations between self-reports, behaviour, 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 et al., 2002), the Reward Responsiveness (RR) scale (Van den Berg et al., 2010), and the ADHD Self-Report Scale (ASRS-6; R. C. Kessler et al., 2005). To obtain behavioural and electrophysiological measures, participants performed the Eriksen Flanker task (Eriksen & Eriksen, 1974; Marhe et al., 2013), the Go/No-Go task (Donders, 1869/1969; Littel, Van den Berg, 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 et al., 2011; Lejuez et al., 2002; Pleskac et al., 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 et al., 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, behavioural 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, behavioural measures, and electrophysiological measures to show only small (but significant) correlations.

6.2 Data and Method

The present section describes the two samples (Sample 1 and Sample 2) and the methods used to analyse 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.

6.2.1 Sample 1

6.2.1.1 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[4] resulting in a final sample of $n = 133$ (average age of 22.23 ($SD = 2.46$) and 39% women).

[4] 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).

6.2.1.2 Procedure. At least two days before the lab session, participants received an email asking them to not drink coffee or smoke cigarettes 90 minutes prior to the 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 tasks) participants would receive a 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 room. Participants were wired to the EEG and performed two tasks: a Go/No-Go (Donders, 1869/1969; Littel, Van den Berg, et al., 2012) and an Eriksen Flanker task (Eriksen & Eriksen, 1974; Marhe et al., 2013), during which EEG was recorded. The full session lasted two hours. All tasks were programmed in E-Prime software. Session design was approved by the institutional review board. Part of the data is reported in a previous study (Rietdijk et al., 2014) on the internal consistency of the EEG measures.

6.2.1.3 Self-Report Measures. The questionnaire included 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 et al., 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"). *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 (R. C. 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 organisation?", "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 included "never", "rarely", "sometimes", "often", and "very often". Cronbach's alpha equalled .52.

6.2.1.4 Behavioural Measures. Participants completed two behavioural tasks: the Go/No-Go task and the Eriksen Flanker task. The Go/No-Go task (Donders, 1869/1969; Littel, Van den Berg, 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 700 ms, and between consecutive vowels a fixation cue ("+") was shown for 300 ms. Both the vowels and fixation cues were presented in white on a black background. Four behavioural 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 (*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 predefined 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 250 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 500 ms break. Three behavioural 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*).

6.2.1.5 Electrophysiological Measures. EEG was recorded during the Go/No-Go task and Eriksen Flanker task using a Biosemi Active-Two amplifier system. A total 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 electro-oculogram), and the infraorbital and supraorbital region of the left eye (for recording a vertical electro-oculogram). Signals were digitalised 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. 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 et al. (1983) algorithm. 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). 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 response). The pre-stimulus 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.

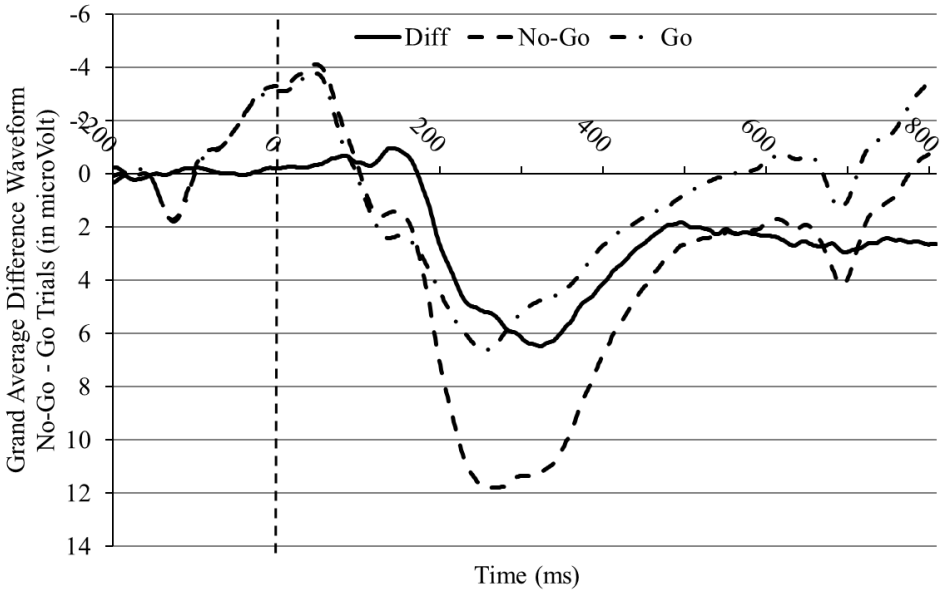
The electrophysiological measures of interest in the Go/No-Go task are the N2 (representing mismatch detection) and the P3 (representing more elaborate appraisal of the stimuli). We opted for analysing difference waves, which has the advantage of eradicating exogenous components, that is, elements that are elicited in response to all stimuli and hence across all conditions (Miltner et al., 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, Van den Berg, et al., 2012; Marhe et al., 2013; Rietdijk et al., 2014), and were compatible with visual inspection of the present grand averaged waveforms (see Figures 6.1 and 6.2).

Figure 6.1

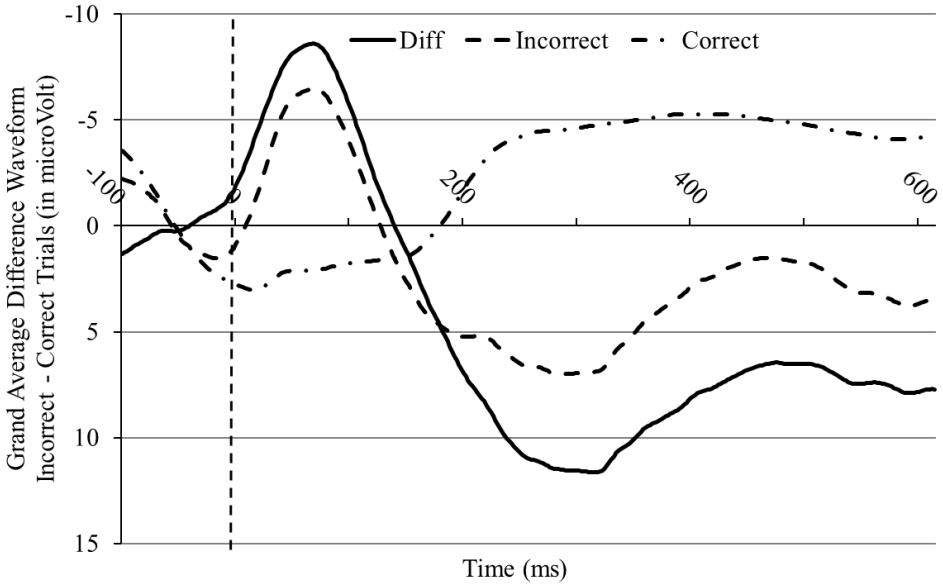
Grand Averaged Difference and Absolute Waveforms for the Go/No-Go task



Note. Average taken across midline electrodes. The difference waveform is similar to Rietdijk et al. (2014), where the data are used for a different purpose.

Figure 6.2

Grand Averaged Difference and Absolute Waveforms for the Eriksen Flanker task



Note. Average taken across midline electrodes. The difference waveform is similar to Rietdijk et al. (2014), where the data are used for a different purpose.

6.2.2 Sample 2

6.2.2.1 Participants. The second sample again consists of university students (n = 181) and was collected between May 2015 and April 2016. Incomplete observations were excluded[5] resulting in a final sample of n = 142 (average age of 20.63 (SD = 2.04) and 54% women).

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

[5] Incomplete observations included 16 no-shows for the lab session, six participants with incorrect electrophysiological measurements on only the BART, 10 participants with incorrect electrophysiological measurements on only the Reward task, and seven 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).

euros and could earn an additional 7.5 euros by performing well on the tasks.[6] 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 behavioural tasks, a Reward task (Franken, Van den Berg, & Van Strien, 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 software. Session design was approved by the local institutional review board.

6.2.2.3 Self-Report Measures. The questionnaire included 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 equalled .78.

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

[6] 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).

6.2.2.4 Behavioural Measures. Participants completed two behavioural tasks: the passive Reward task and the automatic BART. The Reward task (Franken, Van den Berg, & Van Strien, 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.[7] 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 behavioural 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 labelled "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 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 behavioural measures were obtained from the BART: (1) the average number of pumps (*BART Average Pumps*), indicating a riskier choice; and (2) the average response time (*BART Average Response Time*), that is, the time it took participants to choose a number between 1 and 128 and to press the "P".

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

6.2.2.5 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. 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.

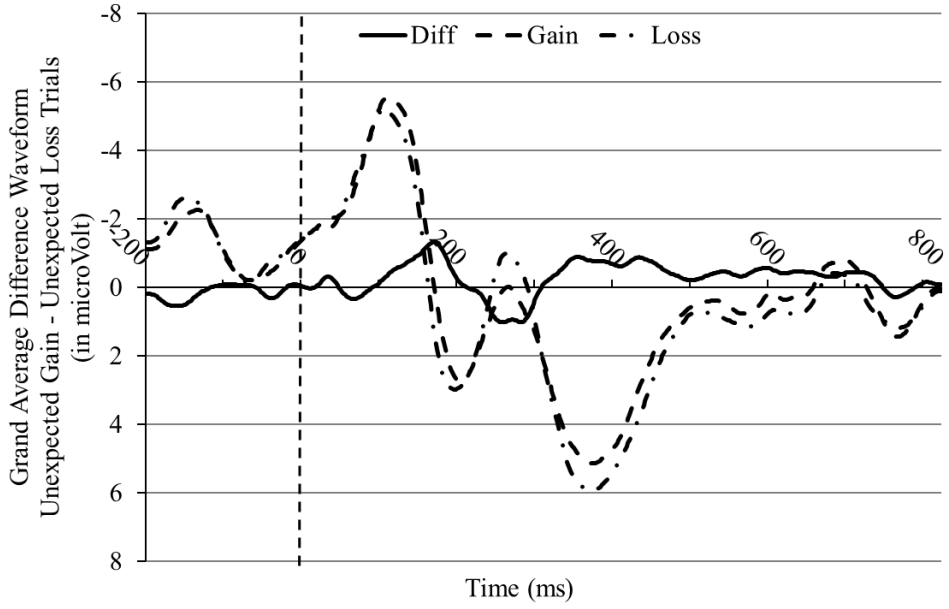
The electrophysiological measures of interest in the Reward task are the N2 (representing mismatch detection), the P2 (representing attention to (deviating) stimuli), and the P3 (representing elaborate stimulus appraisal). The analyses employed difference scores obtained from midline electrodes. Justifications 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 et al., 2015; Warren & Holroyd, 2012), and were compatible with visual inspection of the present grand averaged waveforms (see Figures 6.3 and 6.4).

Figure 6.3

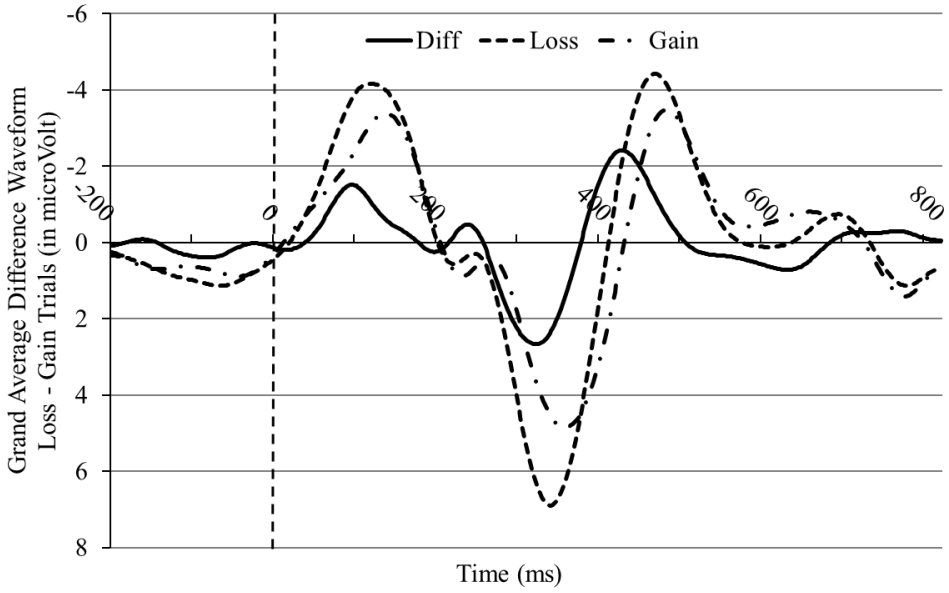
Grand Averaged Difference and Absolute Waveforms for the Reward Task



Note. Average taken across midline electrodes.

Figure 6.4

Grand Averaged Difference and Absolute Waveforms for the BART



Note. Average taken across midline electrodes.

6.2.3 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 inflation factors (VIFs), which indicate the level of multicollinearity[8], 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.

[8] 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 (Diamantopoulos et al., 2008; Hair et al., 2010), there is no indication of multicollinearity.

Third, we used linear regression models to further investigate whether behavioural 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 analysed three multiple regression models: the first model only included behavioural predictors, the second only included electrophysiological predictors, and the third included both behaviour and electrophysiology. The coefficients of the regression models were estimated using Ordinary Least Squares (OLS). To allow for comparison between the models, coefficients were standardised.

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 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.

6.3 Results

6.3.1 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 et al., 2003). Therefore, we examined the possible common method bias using Harman's single factor test. The first principal component explained 11.94% of the variance in Sample 1, and 14.76% in Sample 2. Since this is below the threshold of 50%, the risk of common method bias in our data is small. The VIFs are reported in Tables 6.1 and 6.2 for respectively Samples 1 and 2. The highest VIF in Table 6.1 is 3.34 (for *GNG Number Post-Incorrect Incorrect*), and that in Table 6.2 is 4.55 (for *REWARD N2*). Hence, there is no indication of multicollinearity.

6.3.2 Correlation Analyses

Tables 6.1 and 6.2 show the descriptive statistics for the variables in Sample 1 and Sample 2, respectively. For Sample 1, 100% of the correlations within the impulsivity(-related) self-report measures, 57.14% of the correlations within behavioural measures, and 50% of the correlations within the electrophysiological measures was significant. However,

only 19.05% of correlations between behavioural and self-reported measures, 8.33% of correlations between electrophysiological and self-reported measures, and 17.86% of correlations between behavioural and electrophysiological measures reached significance.

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

6.3.3 Regression Analyses

Tables 6.3 and 6.4 show the results of the OLS regressions investigating whether the joint behavioural measures, the joint electrophysiological measures, or all behavioural 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 behavioural and electrophysiological measures, that is, excluding those with age and gender. For Sample 1, the models including only behaviour (Models 1) and the models including only electrophysiology (Models 2) together have a total of 33 relevant associations. As we allow a 5% 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 behaviour 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, 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 behavioural and electrophysiological measures on the other.

Table 6.1

Descriptive Statistics (Mean, SD, Min, Max, VIF, Cronbach's alpha (diagonal), and Correlations) for Sample 1 (n = 133)

						Correlations and Cronbach's alpha														
	Mean	*SD*	Min.	Max.	VIF	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
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	-	0.04	-0.11	-0.19*	-0.11	-										
6. GNG Number Incorrect NoGo (behaviour)	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 (behaviour)	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 (behaviour)	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 RT (behaviour)	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 (behaviour)	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 RT Incongruent (behaviour)	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 RT Post-Incorrect - Post-Correct (behaviour)	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	.29***	0.20*

Note. ***: $p < .001$; **: $p < .01$; *: $p < .05$; GNG = Go/No-Go task; EF = Eriksen Flanker task; RT = response time in ms; [a]: difference score.

Table 6.2

Descriptive Statistics (Mean, SD, Min, Max, VIF, Cronbach's alpha (diagonal), and Correlations) for Sample 2 (n = 142)

						Correlations and Cronbach's alpha										
	Mean	*SD*	Min.	Max.	VIF	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 (behaviour)	61.86	10.09	24.87	90.83	1.18	-0.09	0.03	0.14	0.14	-0.22**	-					
7. BART Average RT (behaviour)	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. ***: $p < .001$; **: $p < .01$; *: $p < .05$; BART = Balloon Analogue Risk Task; REWARD = Reward task; RT = response time in ms; [a]: difference score.

Table 6.3

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 (behaviour)	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 (behaviour)	-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 (behaviour)	-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 RT (behaviour)	-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 (behaviour)	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 RT Incongruent (behaviour)	-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 RT Post-Incorrect - Post-Correct (behaviour)	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. *: $p < .05$; GNG = Go/No-Go task; EF = Eriksen Flanker task; RT = response time in ms; [a]: difference score.

Table 6.4

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 (behaviour)	-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 RT (behaviour)	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. **: $p < .01$; *: $p < .05$; BART = Balloon Analogue Risk Task; REWARD = Reward task; RT = response time in ms; [a]: difference score.

Table 6.5

Bootstrapped Mean % of Significant Correlations/Associations (Based on 1000 Iterations)

		Sample 1			Sample 2		
	Subsample Size	20	30	40	20	30	40
Correlations	Behaviour 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
	Behaviour vs. Electrophysiology	8.42	9.98	11.87	5.06	4.68	4.69
Associations	Behaviour/Electrophysiology vs. Self-Report (Models 1 and 2)	5.48	5.86	6.13	4.88	4.42	4.11

6.3.4 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 5% significance level) we would have found if we had used such small samples. The results of this bootstrapping analysis are summarised in Table 6.5. 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 analysing the full sample, analysing 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 5% significance level).

6.4 Discussion

The present paper examined the association between self-report measures, behavioural 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, behaviour, 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 behaviour 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 behaviour/electrophysiology. The discrepancy between our current null-findings and previous research possibly results from the limitations that characterise 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 (J. A. Gray, 1987) consists of two separate dimensions of which especially one (rash impulsiveness) is related to impulsivity. Therefore, future studies examining impulsivity could benefit from using well-defined models to operationalise 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 premeditation. Finally, we analysed 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 et al., 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 behavioural 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 behavioural/electrophysiological measures on the one hand and self-report measures on the other hand differ. Typically, behavioural 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, behavioural 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 (behaviour 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. Behaviour and electrophysiology are implicit measures because they largely operate outside of one's awareness, whereas self-reports represent the more conscious processes and are therefore explicit measures (Dittmar et al., 2011; M. W. 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 behaviour 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 et al. (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 et al., 2014; Reutter et al., 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 favourable, ranging from moderate to excellent. Furthermore, the issue raised by Hedge et al. (2017) is limited to explaining the lack of correlations/associations between behaviour 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 ($1 - \beta$), 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 et al., 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 et al. (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.

Chapter 7

Quantifying Evidential Value for Event-Related Potentials as Biomarkers of Substance Use Disorders: A Literature Review and *P*-Curve Analysis

Abstract

Empirical research has put forward multiple Event-Related Potentials (ERPs) as possible biomarkers of Substance Use Disorders (SUDs). However, it is unclear whether the reported relations between these ERPs and SUDs reflect genuine effects or can be explained by selective reporting of statistically significant findings. To determine which ERPs have potential as biomarkers of SUDs, the present study reviewed the existing literature for 14 ERPs, and used *p*-curve analysis to investigate whether these literatures had "evidential value" (i.e., cannot solely be accounted for by selective reporting). In addition, the selective-reporting-corrected statistical power of the literatures was estimated. Following PRISMA guidelines, we identified 65 studies that examined one or more of the reviewed ERPs. *P*-curve analysis was done on literatures for which five or more studies were available: the No-Go N2b and P3, the oddball P3b, the Error-Related Negativity (ERN), the P3 and LPP in response to drug-related stimuli, and the Feedback-Related Negativity (FRN) / Reward Positivity (RewP). All literatures demonstrated evidential value, meaning that the statistically significant findings that were reported could not be explained by selective reporting. However, the majority of the literatures was underpowered, with especially the FRN / RewP (23% estimated power) and No-Go P3 (39%) literatures performing poorly. These ERPs also scored poorly on a sensitivity analysis, showing that evidential value disappeared when the lowest included *p* value was removed from their analyses. In conclusion, strongest support for true neurobiological differences between persons with and without addiction was found for the No-Go N2b, the oddball P3b, the ERN, and the drug-P3 and drug-LPP.

Joint work with Dr. Eric Slob, Dr. Oliver Lindemann, and Prof.dr. Jan van Strien (manuscript in preparation).

7.1 Introduction

Over the past few decades, interest in the neurobiological correlates of addiction has risen, in part brought on by major advancements in genetics research and neuroscience (Volkow & Boyle, 2018). While diagnosis of substance use disorders (SUDs) is still based on reported or observed behaviour, like the continued use of a substance despite adverse consequences, research is increasingly trying to identify the *neurobiological processes* that underlie these behavioural manifestations. In fact, various research frameworks have advocated focusing *solely* on such processes, arguing that since diagnostic categories have failed to align with findings emerging from neuroscience and genetics, and have failed to address the large within-diagnosis clinical heterogeneity that exists, research should move towards a dimensional approach in which mental illnesses are not studied via symptoms but via the constructs that underlie them (Insel et al., 2010).

For addiction research, such constructs have been proposed by the Addiction Neuroclinical Assessment (ANA) framework, which states that substance use disorders comprise disruptions in three domains: *Executive Function, Incentive Salience,* and *Negative Emotionality* (Kwako et al., 2016). Importantly, by examining these underlying disruptions, the framework allows and accommodates the search for biomarkers: quantifiable biological indicators of healthy or pathological processes. One way to examine biomarkers is through electroencephalography (EEG) Event-Related Potentials (ERPs), as their high temporal resolution is well suited for measuring the rapid fluctuations in neural activity that characterise the proposed dysfunctional processes (Houston & Schlienz, 2018). Indeed, research has shown that various ERPs related to the processes implicated in the ANA dimensions present differently in persons with SUDs than in those without.

In the present study, we provide an integrated, updated overview of this research, structured along the ANA dimensions, and systematically examine, using *p*-curve analysis, whether the research contains evidential value and sufficient statistical power. This way, we attempt to determine which ERPs have potential as biomarkers for SUDs and which do not at this moment.

7.1.1 Executive Function

The largest body of ERP addiction studies is on the proposed impairments in Executive Function: the cognitive processes responsible for the execution of goal-directed and self-regulatory behaviour, which are associated with hypoactivity of lateral prefrontal areas, including the dorso- and ventrolateral prefrontal cortex, and the dorsal anterior cingulate

cortex (Goldstein & Volkow, 2011), and which are crucial in countering the pulling force of addictive agents (i.e., their incentive salience).

7.1.1.1 Inhibitory Control. One component of Executive Function that is thought to be affected in individuals with substance use disorders is inhibitory control: the ability to suppress prepotent but unwarranted responses. This ability and its electrophysiological correlates are assessed using cognitive paradigms in which one class of stimuli requires a behavioural response, whereas another requires one to refrain from responding. Of such tasks, the most commonly used is the Go/No-Go task (Donders, 1869/1969), which instructs participants to respond as quickly as possible to frequently occurring "Go" stimuli (generally by pressing a button), but to not respond to the (often less frequently occurring) "No-Go" signals. The degree to which a person succeeds at the latter, represents their ability to suppress what becomes the prepotent motor response (i.e., pressing the button).

The main electrophysiological markers observed in the Go/No-Go task and related tasks are the N2b and P3 components. The negative-going N2b appears approximately 200 ms after presentation of a No-Go stimulus, at fronto-central sites, and while earlier believed to reflect response inhibition (Pfefferbaum et al., 1985; Van Boxtel et al., 2001), is now thought to represent an earlier cognitive process necessary to control the preparation of incorrect responses: conflict monitoring (Donkers & Van Boxtel, 2004; Kopp et al., 1996; Nieuwenhuis et al., 2003). The actual inhibition of the motor response is thought to be reflected by the positive-going, more centrally located P3 wave that follows the N2b 300-500 ms after the No-Go stimulus is presented (Randall & Smith, 2011). This view is supported by research on their neural origins: while the N2b is shown to originate in the anterior cingulate cortex and inferior prefrontal areas (Bekker et al., 2005; Nieuwenhuis et al., 2003), the No-Go P3 is in addition generated in areas associated with (pre-)motor function (Huster et al., 2010; Kiefer et al., 1998).

With regard to substance use disorders, various studies have found attenuated No-Go N2b and P3 amplitudes in those with SUDs compared to controls (e.g., Morie et al., 2014; Sokhadze et al., 2008), which attests to these ERPs' possible role as a biomarker of inhibitory control dysfunction. However, others have found increased No-Go P3's in SUDs (e.g., Y. Chen et al., 2022; Fathi et al., 2022), and a systematic review (Luijten et al., 2014) and recent meta-analysis (Y. Zhang et al., 2021) only reported support for reduced N2b amplitudes, while finding inconsistent effects for the P3, suggesting that conflict detection but not necessarily later inhibition may be impaired in SUDs.

7.1.1.2 Attention Allocation. A second component of Executive Function that has been studied in relation to substance abuse is attentional resource allocation, or attentional control, for which electrophysiological correlates have mostly been examined using an oddball paradigm (N. K. Squires et al., 1975). In a typical oddball task, participants are presented with a rapid sequence of repetitive audio or visual stimuli in which rare target stimuli ("oddballs") are interspersed. Participants must respond to these oddballs, generally by pressing a button or by mentally counting them, which requires continuous attention to be paid to the stimuli.

In these tasks, circa 300 ms after the target (the oddball) is presented, a P3 with a centro-parietal distribution is elicited, activity that is believed to reflect the top-down deployment of attentional resources to the stimulus and the subsequent updating of the mental model of the stimulus environment (Donchin, 1981; but see Verleger, 2020). This component is called the P3b. A second P3 that can be elicited within an oddball paradigm is the P3a (or novelty P3), which presents slightly earlier, has a fronto-central distribution, and habituates rapidly (Courchesne et al., 1975; N. K. Squires et al., 1975). This component can develop when a subject is not actively paying attention to the target stimuli, or in response to distracter stimuli that may be included in the task and that participants are instructed not to attend to (R. T. Knight, 1984; Simons et al., 2001; Snyder & Hillyard, 1976). The P3a therefore reflects stimulus-driven, automatic attention allocation, related to the orienting response. In line with this, the generators of the P3a have been localised primarily within the frontal cortex and the anterior cingulate cortex, while P3b sources additionally include temporoparietal regions that play a role in memory operations (Volpe et al., 2007; Wronka et al., 2012).

Regarding addiction, several studies have reported reduced P3a amplitudes in individuals with substance use disorders compared to healthy control subjects (e.g., Hada et al., 2000; X. Liu et al., 2020), although the majority of SUDs research is concerned with the P3b. Here, reduced amplitudes in response to target stimuli have been widely observed for various dependencies, including that to alcohol (e.g., H. L. Cohen et al., 2002), tobacco (e.g., Anokhin et al., 2000), and cocaine and opioids (e.g., Bauer, 2001), as well as in persons with behavioural addictions like that to gaming (e.g., Park et al., 2017). These individual findings have been further supported by a large meta-analysis, suggesting that an attenuated oddball-P3b may be a promising biomarker for SUDs (Euser et al., 2012).

7.1.1.3 Error Processing. A third component of Executive Function that has received considerable attention in addiction research concerns performance monitoring, or more specifically, error processing: the ability to adequately detect and process the mistakes that

one makes in order to adjust subsequent behaviour. This can be examined using any task in which motor response errors occur frequently (like the Go/No-Go, Stroop, or Simon task), but is mostly studied using the Eriksen Flanker task (Eriksen & Eriksen, 1974). Here, participants are shown a centrally positioned stimulus (often a letter) flanked by either congruent or incongruent stimuli (e.g., "HHHHH" or "HHSHH"), and are asked to press a button with one hand if the target is an "H" and with the other hand if it is an "S". When target and flankers are congruent, participants generally respond swiftly and accurately, but when they are incongruent, performance is slower and error-prone.

Whenever participants err in these tasks, two ERPs can be observed: the Error-Related Negativity (ERN, Gehring et al., 1993; or Ne, Falkenstein et al., 1991), which peaks between 25 and 100 ms after the commission of an error at fronto-central sites; and the subsequent error positivity (Pe; Falkenstein et al., 1991), which has a centro-parietal distribution and develops between 200 and 400 ms after an incorrect response. Although their precise functional significance is still debated (Olvet & Hajcak, 2008; Overbeek et al., 2005), the ERN is thought to reflect early error processing activity of the caudal ACC (O'Connell et al., 2007; Van Boxtel et al., 2005), and develops regardless of whether subjects are consciously aware of committing the error (Nieuwenhuis et al., 2001). Like the No-Go N2, which has a similar scalp topography and neural origin (Van Veen & Carter, 2002), it has been proposed to reflect conflict monitoring (C. S. Carter et al., 1998; Yeung et al., 2004) – with the N2 appearing prior to high-conflict correct responses, and the ERN following errors, when conflict arises between the erroneous and the correct response. The Pe is thought to reflect later, more conscious processing of the error (Nieuwenhuis et al., 2001; Overbeek et al., 2005), and originates from more rostral ACC sites (O'Connell et al., 2007; Van Boxtel et al., 2005).

ERP studies on error processing in addiction have reported both reduced ERN (e.g., Sokhadze et al., 2008) and Pe (e.g., Franken et al., 2007) amplitudes on error trials in those with SUDs compared to control subjects. These findings were corroborated by two meta-analyses on ERN and Pe amplitudes in persons with externalising disorders, including addiction (Lutz et al., 2021; Pasion & Barbosa, 2019), and by a systematic review (Luijten et al., 2014) and meta-regression (Y. Liu et al., 2023) specifically on SUDs, indicating that decreased sensitivity to erring as measured by these ERPs may constitute a biomarker. However, two further meta-analyses on SUDs reported non-significant effects for the ERN, while not examining the Pe (Webber et al., 2024; Y. Zhang et al., 2021).

7.1.2 Incentive Salience

Next to research on executive dysfunction, ERP addiction studies have extensively examined another dimension proposed by the ANA to be disrupted in individuals with substance use disorders: Incentive Salience, the attribution of motivational value to stimuli, mediated by mesocorticolimbic dopamine structures among which the (ventral) striatum and the orbitofrontal and ventromedial prefrontal cortices (Goldstein & Volkow, 2011). Although originally believed to encode the hedonic value of rewarding stimuli, this neuronal circuit has in fact been shown to encode how much something is physiologically "wanted" and not how much it is "liked" (Berridge & Robinson, 2016).

7.1.2.1 Drug Cue Salience. One aspect of Incentive Salience that is studied within the context of substance use disorders is the salience attributed to substance-related stimuli. All major substances of abuse, irrespective of their precise pharmacological properties, are acutely rewarding through triggering a release of dopamine. These dopamine bursts that follow drug-taking also strengthen conditioned associations, such that, with repeated use, not only the drug itself but also cues related to drug delivery (like drug paraphernalia) acquire incentive value (T. E. Robinson & Berridge, 1993). EEG research on this increased valuation of drug cues is often done within cue-reactivity paradigms, which involve passive viewing of neutral and SUDs-related stimuli. In addition, to demonstrate that observed effects arise from involuntary bias toward substance-related stimuli, rather than from intentional viewing strategies (e.g., consciously focussing on drug cues), paradigms in which attention is manipulated have also been applied. These include oddball and Go/No-Go tasks that use substance-related stimuli.

In response to these substance-related stimuli, two ERPs are observed, either when the stimuli are explicitly or implicitly attended to: a P3 that is maximal at parietal sites (Keil et al., 2002); and a later, likewise positive component referred to as the Late Positive Potential (LPP) or Slow Potential (SP). This component is often defined as starting between 600 and 800 ms, can last for several seconds, and while starting off strongest at posterior sites (Cuthbert et al., 2000; Keil et al., 2002), is equally strong at anterior sites after about one second (Hajcak et al., 2007). Both the P3 and LPP are enhanced in response to affectively arousing (pleasant and unpleasant) stimuli relative to neutral stimuli, and are thought to represent increased attention to and processing of stimuli that are motivationally relevant (Hajcak et al., 2010). In addition, given its sustained duration and diffuse topography (Sabatinelli et al., 2007), the LPP is thought to reflect memory encoding and storage processes, rather than merely reflecting a continuation of the P3 (Foti et al., 2009; Koenig & Mecklinger, 2008).

With regard to substance abuse, numerous studies have shown that in individuals with SUDs, more than in healthy controls, substance-related stimuli produce higher P3 and LPP amplitudes than neutral stimuli. This has been demonstrated in both cue-reactivity (e.g., Cheng et al., 2016) and active paradigms (e.g., Lubman et al., 2007), for a large variety of addictions including behavioural ones, and has been corroborated by various meta-analyses (Littel, Euser, et al., 2012; Norberg et al., 2016; Webber et al., 2022; Y. Zhang et al., 2021), all showing medium to large effects. Taken together, this indicates strong potential for these two ERPs as biomarkers of increased motivated attention to drug cues.

7.1.2.2 Reward Responsiveness. Next to sensitivity to substance-related stimuli, a second component of Incentive Salience that has been studied in relation to substance use is sensitivity to non-drug rewards, mostly monetary ones. One way through which this is studied in EEG research are two-choice gambling tasks, like the Doors Task (Dunning & Hajcak, 2007). Here, two options are provided from which participants must choose, after which they are presented with the rewards or losses associated with that choice. In addition, it is examined with more complex tasks in which outcomes depend on participants' performance, like in the Balloon Analogue Risk Task (BART; Lejuez et al., 2002).

When the outcomes of their decisions are presented to participants, two ERPs are observed. First, the Feedback-Related Negativity (FRN), which occurs at fronto-central sites 200-300 ms after outcomes are presented and was originally thought to elicit stronger negativity after losses than gains (Gehring & Willoughby, 2002; Miltner et al., 1997). However, recent findings suggest that it is not driven by loss-related negativity but rather by reward-related positivity that is suppressed following non-reward (Holroyd et al., 2008). Therefore, the component is also referred to as the Reward Positivity (RewP; Hajcak Proudfit, 2015). It originates from changes in midbrain dopamine impacting the ACC, reflecting prediction error signals (i.e., differences between received and predicted outcomes), which index reward learning (Holroyd & Coles, 2002) and reward responsiveness (Bress & Hajcak, 2013). Indeed, the RewP is related to activity in reward-related areas like the ventral striatum (Becker et al., 2014). A second ERP implicated in reward processing is the Feedback-P3: a centro-parietal positivity peaking 300-600 ms after outcome presentation (Sutton et al., 1965). Contrary to the RewP, the Feedback-P3 is modulated not by valence but by the value of an outcome, being stronger for larger outcomes (Yeung & Sanfey, 2004; but see San Martín, 2012). It is also larger when outcomes are unexpected (Pfabigan et al., 2011), which is consistent with the Feedback-P3 reflecting top-down coding of motivationally significant information, much like what the P3b observed in oddball tasks represents (Nieuwenhuis et al., 2005; Yeung & Sanfey, 2004).

As for substance use disorders, neural hyposensitivity to feedback represented by diminished RewP and Feedback-P3 potentials has been reported by various studies (e.g., Kamarajan et al., 2010; Morie et al., 2016; Yau et al., 2015), and is in line with theories stating that persons with SUDs have reduced reward responsiveness as a result of genetic vulnerability causing deficient dopamine function (Blum et al., 1996); and/or because of increased reward thresholds caused by long-term abuse (allostatic adaptation; Koob & Le Moal, 2001). Nonetheless, neural hypersensitivity to rewards has also been reported (e.g., J. Li & Yao, 2023; Wei et al., 2018), and in a recent meta-analysis, the RewP did not significantly differ between those with SUDs and healthy controls (Webber et al., 2024).

7.1.3 Negative Emotionality

The third and final dimension the ANA proposes to be disrupted in substance use disorders is Negative Emotionality: the propensity to experience negative feelings, and reduced ability to experience positive ones. This has been put forward as a possible driver for the use of addictive agents (Kwako et al., 2016), with addiction itself further progressing from positive to negative reinforcement: taking substances not to induce a pleasant sensation, but to alleviate the negative emotions inherent to withdrawal, which are brought on by activation of the brain's sensitised stress and anti-reward systems, including corticotropin-releasing factor in the extended amygdala, and dynorphin in the ventral striatum; and downregulation of the brain's anti-stress systems (Koob et al., 2014).

7.1.3.1 Affective Image Processing. One component of Negative Emotionality that has been studied with the use of EEG is the processing of affective content, generally examined within passive viewing or active response (like the Stroop task) paradigms in which participants are presented with neutral, pleasant, and unpleasant stimuli – sometimes in addition to being presented with drug cues. Pleasant images often come in the form of stimuli that are considered intrinsically appeasing, like erotic images (Cassidy et al., 2014; Lubman et al., 2008) or images of food (Lubman et al., 2009; Minnix et al., 2013). Unpleasant images include scenes that are naturally appalling, like scenes of violence, accidents, and mutilation (Minnix et al., 2013; L. Yang et al., 2015) or dangerous animals (Cassidy et al., 2014; L. Yang et al., 2015).

The electrophysiological responses to these pleasant and unpleasant stimuli are similar to those discussed in relation to the incentive salience of drug cues: a parietal P3 emerging about 300 ms after stimulus presentation, representing increased attention to and facilitated processing of the stimulus, and a subsequent LPP that in addition reflects memory formation

processes, both of which are enhanced for affectively arousing (thus pleasant and unpleasant) stimuli relative to neutral ones (Hajcak et al., 2010).

Mixed findings have been reported as to whether these electrophysiological responses to affective images differ between persons with and without SUDs. Some studies have found the P3 and LPP to emotion-laden images to be reduced in individuals with SUDs compared to healthy controls (e.g., Dunning et al., 2011; Lubman et al., 2008), though this effect is mainly found for pleasant stimuli (e.g., Parvaz et al., 2017; J. Wang & Li, 2023), suggesting that diminished P3 and LPP amplitudes may have potential as biomarkers of hyposensitivity to positive affective content. However, some studies have reported opposite effects, for example showing a larger LPP to pleasant stimuli in smokers than non-smokers (Minnix et al., 2013; J. D. Robinson et al., 2015), and various studies have found no significant differences for either pleasant or unpleasant stimuli (e.g., Asmaro et al., 2014; Cassidy et al., 2014; Deweese et al., 2018; L. Yang et al., 2015). In line with this, a recent meta-analysis found no significant differences between persons with SUDs and healthy controls in their electrophysiological responses to pleasant and unpleasant images, granted that for pleasant images the P3 and LPP were smaller in substance users, but that this effect did not reach significance (Webber et al., 2022).

7.1.3.2 Emotional Facial Expression Processing. A second body of EEG literature that relates to Negative Emotionality concerns individuals' response to one specific type of affective stimuli: faces. While some studies on affective processing do incorporate facial stimuli (e.g., Dunning et al., 2011), a separate set of studies focusses exclusively on so-called Emotional Facial Expression (EFE) processing, as faces are highly salient and recurrent carriers of affective information. The electrophysiological correlates of EFE are often examined with tasks that require participants to discriminate which of two emotions a facial stimulus is depicting (e.g., angry vs. sad) by as quickly and as accurately as possible pressing the button corresponding to that emotion. Often, emotions are presented at different intensities, accomplished by morphing the stimuli (Hoffman et al., 2019). In addition, tasks in which attention is manipulated are used, like oddball (Maurage et al., 2007) and Stroop (Opitz et al., 2024) tasks that involve faces.

The main ERPs observed in response to these stimuli are the N170 and P3. The former is a negative wave peaking 170 ms after stimulus onset, showing a larger amplitude for faces and face-related information than for other objects (Bentin et al., 1996). In addition, it is sensitive to expression, with particularly angry, fearful, and to a lesser extent happy faces eliciting large N170s (Hinojosa et al., 2015; Schindler & Bublatzky, 2020). The component is strongest at (right) temporo-occipital sites, which is consistent with its neural source being

located at the superior temporal sulcus and fusiform gyrus (Rossion et al., 2003; Sadeh et al., 2010), and is associated with the structural encoding – not identification – of faces (Bentin et al., 1996) and other objects of visual expertise (Rossion et al., 2002). At fronto-central sites a positive counterpart called the Vertex Positive Potential (VPP; Bötzel & Grüsser, 1989; Jeffreys, 1989) is observed, which is thought to represent the same neural processes (Joyce & Rossion, 2005). The subsequent parietal P3 observed in response to faces reflects their higher-order processing and affective labelling (Schupp et al., 2006), and although not face-specific, is enhanced for affectively arousing stimuli and therefore for EFEs (Schindler & Bublatzky, 2020).

Regarding substance abuse, studies have reported smaller N170s (e.g., Criado & Ehlers, 2007; Maurage et al., 2007; Maurage, Philippot, et al., 2008) and P3s (e.g., Bohan et al., 2020; Maurage et al., 2007) to EFE stimuli in persons with SUDs compared to healthy control subjects, although for the P3 similar effects were found when people were tasked to discriminate the sex of (neutral) faces, suggesting that the impairment may not be specific to emotion (Hoffman et al., 2019; Maurage, Campanella, et al., 2008). Furthermore, especially for the N170, several studies have reported null results (e.g., Bohan et al., 2020; Hoffman et al., 2019). A systematic review on the N170 in psychiatric disorders concluded that findings regarding alcohol use disorder and various other disorders were inconsistent, though reported group differences generally showed smaller amplitudes to EFE in diagnosed versus healthy people, offering some support for the N170 being a biomarker of deficient facial affect processing (Feuerriegel et al., 2015).

7.1.4 Strength of the Evidence and Challenges Within the Field

The above review of existing literature shows that for each dimension proposed by the ANA to be involved in the development and perpetuation of substance use disorders – Executive Function, Incentive Salience, and Negative Emotionality – differences in electrophysiological signals have been observed between individuals with and without SUDs. However, the level of support for these observed differences, and thus the potential these ERPs have as biomarkers of substance use disorders, varies across the domains and individual signals. For example, while research on the N170 in response to emotional faces is somewhat limited and has produced mixed findings, the cue-P3 indexing motivated attention to drug cues has been studied extensively, and has consistently – also in meta-analyses – shown to differ between persons with and persons without SUDs.

Yet, even when robustly supported by large numbers of papers, determining the potential that ERPs have as biomarkers of SUDs is not straightforward, as what is reported in the literature is not necessarily representative of reality. Due to researchers and journals having a preference for statistical significance, employing an arbitrary *p* value of .05 to decide whether something is interesting enough to be shared, published literature is riddled with studies reporting significant findings, while studies reporting non-significant ones are much less likely to get published – a phenomenon reported as early as 1959 by Sterling and later coined the file-drawer problem by Rosenthal (1979). A consequence of this bias against non-significant findings is that it can impact researchers' behaviour (Simonsohn et al., 2014a). Throughout the process of designing a study and collecting, analysing, and reporting the data, researchers are faced with many decisions that have no objective "right" answer. For instance, they need to determine if their *n* is sufficient or if more data needs to be collected; what values of each measure constitute an outlier and whether these need to be removed; and which control variables or covariates need to be considered (Simmons et al., 2011). Decision points like these are called Researcher Degrees of Freedom (RDF), referring to the flexibility and ambiguity inherent to the scientific process. This flexibility and ambiguity, however, combined with a desire for statistical significance, can lead to the selective reporting of analyses that "worked" – not disclosing the alternative analysis routes one explored that did not render significant outcomes. As an example, one may fail to disclose that effects were only significant after including more participants, or after excluding specific outliers. These practices, called *p*-hacking[9] (Simonsohn et al., 2014a), are not necessarily deliberate or come from ill intent. As Simmons et al. (2011) point out, in contexts where choices are arbitrary – and thus all options equally reasonable – it is easy for researchers to come up with plausible (to themselves and others) post-hoc justifications for why specific choices were "right". The problem that arises from these behaviours, however, is that the chance of a type I error – a false-positive outcome – increases. This can even occur when researchers follow only one analysis path: whenever choices are not prespecified, each one (like removing certain outliers) is contingent on the data. Because the data determine the decision tree, the probability of finding a bogus effect is larger than the assumed 5% (Gelman & Loken, 2013).

In the field of EEG and other neurophysiological research, this risk of researcher flexibility causing an increased probability of false-positive findings is amplified by the need for extensive pre-processing of the data. Even simple EEG studies generate enormous

[9] Other terms have also been employed, but *p*-hacking is arguably the most familiar one. We refer to Gelman and Loken (2013) for an overview of terms and an argument as to why other terms may be preferred.

amounts of raw data, requiring researchers to make a large number of decisions for which various reasonable choices exist. For example, an ERN observed in a Flanker task could be measured as the average amplitude recorded at electrode FCz in the 25-100 ms time window, after removing signals larger than 100 μV. However, taking the ERN as a peak measure from electrode Cz would be equally reasonable, as would be shortening the time window to 50-100 ms, and removing signals larger than 70 μV. Likewise, either an average or mastoid reference could be used, a baseline correction of 50 ms or 100 ms or anything in between, and frequencies above 30 Hz but also those above 40 or 50 Hz may be filtered out. These combined choices amount to thousands – if not more – ways to pre-process the same data, with different pathways leading to different results (Clayson et al., 2021; Nikolin et al., 2022; Šoškić et al., 2024), thus creating many opportunities to obtain statistically significant findings that reflect random variation rather than true effects (Luck & Gaspelin, 2017). The odds of this happening are further increased by two analytical practices common in EEG research: selecting time windows and electrode sites not *a priori* but based on visual inspection of the grand-averaged waveforms, essentially making multiple implicit comparisons before selecting the "definitive" statistical comparison; and conducting ANOVAs that in addition to group or condition include factors for which no hypothesis is presented (like electrode site), without correcting the familywise error rate, thereby wrongfully treating all main and interaction effects that arise as planned comparisons. Indeed, these practices lead to higher Type I error rates, with simulations showing that the probability of obtaining a false-positive outcome can rise to over 50% (Kilner, 2013; Luck & Gaspelin, 2017). Taken together, it is possible that ERP effects, like the deviations reported in persons with addiction, are not genuine but due to the selective reporting of significant results. If this were the case, advocating these ERPs as biomarkers would be misguided, or at least premature. As such, it is critical to evaluate whether findings can be accounted for by selective reporting, before any claims as to the suitability of the ERPs as biomarkers of SUDs are made.

7.1.5 Testing for Selective Reporting

Various methods exist to evaluate the bias that may be present in the published literature.[10] The most popular ones are funnel-plot-based, most prominently Egger's test of the asymmetry of the funnel (Egger et al., 1997), which examines whether the effect sizes

[10] We refer to Marks-Anglin and Chen (2020) for an extensive overview of bias evaluation methods. Here we summarise the methods commonly used in this subfield, and thus those used by the meta-analyses referenced in our literature review.

of studies are associated with their precision; and the trim-and-fill method, which removes (trims) real studies and adds (fills) fictitious ones to force the funnel plot to symmetry, using the new funnel to estimate a bias-corrected effect size (Duval and Tweedie, 2000). A problem with these methods, however, is that funnel plots can be asymmetric for a variety of reasons aside from bias (Egger et al., 1997; Sterne et al., 2005). Moreover, these methods rest on the unlikely assumption that selective reporting is driven by effect size rather than statistical significance (Simonsohn et al., 2014b). A second popular method is the fail-safe *N*: the number of unpublished or new studies averaging null results that would be required to make the overall effect statistically non-significant, as indicated by a *p* value above .05 (Rosenthal, 1979), or practically unimportant, as denoted by a low effect size (Orwin, 1983). The higher this number, the more confident one can be that the effect truly exists. However, this logic is invalidated by *p*-hacking, as by doing so large sets of false-positives can be attained without file-drawers full of non-significant studies (Simonsohn et al., 2014a). Furthermore, since the fail-safe method does not employ a statistical model, little can be learnt from it regarding the size or impact of the bias (Marks-Anglin & Chen, 2020). Finally, a more general problem that the fail-safe method and Egger's test share is that the question they ask ("Are there null-findings we are not observing in this literature?") is not actually the question we are interested in, nor is it a meaningful one. Since we know that most published studies report significant findings (Scheel et al., 2021; Sterling, 1959) and that the majority of research is underpowered (Button et al., 2013), it *must* be so that a substantial number of unpublished null results exist (Simonsohn, 2012). Hence, it makes little sense to test for this, as the answer should always be "Yes" – unless a field consists solely of registered reports, in which the case the question is still meaningless, but the answer will always be "No". What we want to test for, is whether the effects that we study are true. In other words: after taking into account selective reporting, is there still evidence of a genuine effect?

One method that answers this question is *p*-curve (Simonsohn et al., 2014a), a *p* value based selection model (Hedges, 1984) that makes use of predictable trends in the distribution of statistically significant *p* values. *P* values indicate how likely one is to observe an effect equal to or more extreme than the one observed if the null hypothesis were true. By definition, then, when there is no true effect, a set of independent *p* values should be distributed uniformly: each *p* value is equally likely to occur. Instead, when an effect does exist, the distribution should be right-skewed: we would expect to find more small (e.g., .01) than large (e.g., .04) significant *p* values, indicating that, overall, there is more strong than weak support for the hypothesis. In fact, for a true effect with sufficient statistical power, the majority of significant *p* values will be below .01, and only a few will be between .04 and .05 (Lakens & Evers, 2014; Sellke et al., 2001). The stronger the effect and the larger the

sample, the more right-skewed the distribution will be, though studies with lower power investigating a true effect should still demonstrate a right skew. The skew is undermined by selective reporting: when selectively reporting findings that "worked", or when analysis routes depend on the data, the resulting significant *p* values are likely due to chance, and are more often close to .05 than 0, reducing the right skew of the curve and potentially forming a left skew (Simonsohn et al., 2014a; Simmons & Simonsohn, 2017). The *p*-curve method *tests for the presence of a right skew in the observed distribution of statically significant p values for a set of studies examining a specific effect*. While it is not unlikely for an individual high-powered study of a true effect to yield a *p* value between .01 and .05, or for an individual study that employed selective reporting when the underlying effect was not true to yield one below .01, it is highly implausible these things occur across a *series* of studies. Therefore, the *p*-curve method asserts that when the observed *p*-curve for a *set* of studies is right-skewed, the set contains "evidential value", meaning that we can rule out selective reporting as the *sole* explanation for this set of findings being statistically significant, and that there is thus a genuine effect (Simonsohn et al., 2014a). Within the context of the present research, it would mean that taken as a whole, the statistically significant findings reported by existing studies reflect a true neurobiological difference between individuals with and those without addiction.

7.1.6 Previous P-Curves and the Present Study

Since its development, *p*-curve has been applied in various fields of research: mostly psychology and other social sciences, but also in neuroscience and neuropsychology (e.g., Burns et al., 2019; Hosseini-Kamkar et al., 2021). Within the field of psychophysiology specifically, Carbine et al. (2019) applied *p*-curve to studies published recently versus a decade prior, finding evidential value for both points in time. However, acknowledging the high RDF in psychophysiological research, they recommended using *p*-curve within specific sub-disciplines, stating that "hot topics" – which includes the search for biomarkers – may be more susceptible to selective reporting than the set of diverse studies they examined. Indeed, findings from two studies applying *p*-curve to narrower and more "trending" fields of research have been less compelling. In a study examining ERN and Pe amplitudes in relation to externalising problems, Lutz et al. (2021) reported a significant right skew for the ERN, demonstrating evidential value, but found an inconclusive result for the Pe, indicating a need for more data. In another study, examining the ERN and RewP in relation to depression, evidential value was found for both ERPs, but with the caveat that the effects depended on extreme findings reported by a small subset of studies (Clayson et al., 2020).

Specifically, evidential value for the ERN disappeared after dropping 1 of the 20 included *p* values, and that for the RewP after dropping 1 of 27. Hence, the evidence for these ERPs' evidential value was not robust. In addition, both Lutz et al. and Clayson et al. employed *p*-curve to estimate the (selective-reporting-corrected) average power present in the included studies, finding percentages far below what is considered adequate. Whilst this is not an unexpected finding (Clayson et al., 2019), it is cause for concern. When power is low, the probability of a false-negative finding is larger (Button et al., 2013). Since researchers will obtain fewer initially significant outcomes, this may trigger opportunistic RDF use (Stanley et al., 2018), which in turn can increase the chances of false-positive findings as well.

In line with Carbine et al.'s call for applying *p*-curve in specific areas of psychophysiology, and following the two studies that have already done so, the present meta-analysis examines, for each of the ERPs reviewed in relation to the ANA framework, whether evidential value can be found as to their relationship with substance use disorders. In other words, we examine whether selective reporting can be ruled out as sole explanation for the statistically significant ERP differences reported between individuals with and without SUDs. In addition, given Lutz et al. (2021) and Clayson et al.'s (2020) findings regarding low power, we use *p*-curve to estimate the selective-reporting-corrected average statistical power of each of the literatures. An overview of the ERPs included in the analyses, along with the ANA dimension they relate to and the paradigms in which they are studied, is presented in Table 7.1. Taken together, the current meta-analysis has two important contributions: (1) it is the first to synthesise empirical findings in the field of addiction research with the use of *p*-curve, and (2) offers the most extensive overview of ERP-addiction research to date, comprising 6 ERPs reflecting Executive Function, 4 reflecting Incentive Salience, and another 4 reflecting Negative Emotionality. With this, we provide a strong foundation for further research into electrophysiological biomarkers of addiction. Specifically, for literatures that show evidential value and high power, one can be confident that existing research provides genuine support for what is being studied: here, the relationship between an ERP and SUDs. Hence, one may use that ERP to study subgroups or moderators of substance use disorders, and even go as far as to explore the ERP's use in treatment settings (Houston & Schlienz, 2018). Instead, for literatures that do not show evidential value, are inconclusive, or are underpowered, advocating or using the ERP as biomarker is ill-advised. For these ERPs, more research is needed, which the outcomes from the *p*-curve analysis can help decide on.

7.2 Method

7.2.1 Literature Search and Manuscript Selection Criteria

We conducted a literature search in three databases: Scopus, PubMed, and PsycInfo. Search terms relating to SUDs (Drug dependence, Substance abuse, Gambling, …) were crossed with search terms relating to electroencephalography (EEG, Event-related potential, P3b, …), and with search terms relating to the ANA (sub)dimensions and research paradigms (Inhibitory control, Cue reactivity, Stroop task, …). In addition, the searches were limited to human subjects and to articles written in English. The full search queries for each database are found in Table S7.1 in the Supplementary Materials. The searches were performed September 5, 2024. Identified records were exported to RefWorks. Duplicates were located with RefWorks' *Find Duplicates* function (set to similarity, comparing entries' titles and authors) and were inspected manually in order to correct any false-positives before removal. Article selection took place according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Page et al., 2021) guidelines. The PRISMA flow diagram of the article selection process is shown in Figure 7.1. Studies were included in the *p*-curve analysis if they satisfied the following criteria:

1. Empirical, single papers (i.e., no reviews, no dissertations). Dissertations were not included given that their individual chapters were commonly – after revisions – published in journals as well.
2. Participants met DSM or ICD diagnostic criteria for addiction and/or were recruited from treatment facilities. Two exceptions were specified. First, for addictions not or only recently included in the DSM and ICD (in particular gambling disorder and internet (gaming) disorder), a score indicative of addiction as measured with a validated instrument also sufficed. This same exception was set for tobacco use disorder, which, despite long-term DSM and ICD inclusion, is rarely formally diagnosed. Furthermore, the use frequency of tobacco was considered, with studies also being included when the SUDs sample was reported to, on average, consume 5 cigarettes a day or more (which corresponds to a 75% chance of meeting DSM criteria, Oliver & Foulds, 2021). Studies were not excluded for polydrug use or for (part of) the sample using opioid substitution substances (like methadone) provided in treatment. Furthermore, studies were not excluded based on participants having comorbid disorders, except when samples solely consisted of persons with a specific, non-SUDs comorbidity (such as research into alcoholism in persons with PTSD). No distinction was made between current and lifetime diagnoses.

3. Reported effects in persons with SUDs were relative to those in healthy controls (HCs). Individuals were considered healthy when no psychiatric or neurological diagnosis was reported, and/or when the validated instrument used for defining the SUDs group was not indicative of addiction for HCs. Due to the frequent occurrence of tobacco use among both individuals with SUDs and controls, an exception was made for smoking: healthy controls in studies not focussing on tobacco addiction could smoke as long as their use did not exceed that of the SUDs group. For studies on addiction to tobacco, HCs had to be non-smokers.
4. Both the SUDs and HC sample were adults. Studies were not included when the average age of the sample(s) was lower than 18 years, and/or when participants below the age of 18 were reported to be present, as many ERPs change with maturation (e.g., Davies et al., 2004; Wetzel et al., 2006) and as diagnosis of SUDs is often given in adulthood.
5. The study examined one or more ERPs of interest, measured with a typical ERP amplitude method (such as mean or peak amplitude), in a relevant location, at a relevant time, and within a relevant paradigm (as specified in columns 4, 5, and 6 of Table 7.1). As ERPs are defined by their scalp location, timing, and by the specific event that gave rise to the neural activity, these are not necessarily interchangeable and were therefore bound by requirements.
6. ERPs were recorded under "neutral" conditions: participants could not be under the acute influence of substances (i.e., acute administration studies were not included), receive neuromodulation (such as transcranial direct current stimulation), be instructed to adapt their behaviour or feelings (for example by practising mindfulness during stimulus presentation), or be deliberately put into a certain mindset (for example by having stress induced directly before measuring the EEG). Data from pre-intervention tests could be included.
7. The finding of interest (see section 7.2.2) was statistically analysed. If our finding of interest was not analysed by the original author(s), the study was not included.
8. The data necessary for computing an exact *p* value (the *df* and test statistic) for the finding of interest were available from the published report or were provided by the authors upon request.
9. The exact *p* value of the finding of interest, as recomputed using the relevant *df* and test statistic, was $\leq .05$.
10. The sample was independent from other samples. If the same sample was used in more than one study that met the inclusion criteria, the report that was first published was included.

Table 7.1

ERPs Included in the P-Curve Analyses, Organised by ANA Dimension

Dimension	Subdimension	ERP Name	ERP Location	ERP Latency	Paradigm
Executive Function	Inhibitory control	N2b	F, C	~100-300 ms	Inhibition tasks with neutral (i.e., non-SUDs-related) stimuli that require participants to withhold their actions, like the Go/No-Go task, the Stop-Signal task, and the cued CPT
		NoGo-P3	F, C, P	~300-500 ms	
	Attention allocation	P3a	F, C	~300-500 ms	Visual or auditory oddball tasks with neutral (i.e., non-SUDs-related) stimuli, and related attention tasks like the CPT and RHT
		P3b	C, P	~300-500 ms	
	Error processing	ERN	F, C	~0-100 ms	Performance-monitoring tasks with neutral (i.e., non-SUDs-related) stimuli, like the Flanker task, Go/No-Go task, Stroop task, and Simon task
		Pe	C, P	~150-400 ms	
Incentive Salience	Drug cue salience	Cue-P3	C, P	~300-500 ms	Passive viewing or listening tasks and active response tasks (e.g., Flanker, Go/No-Go, Stroop, oddball) with SUDs-related and neutral stimuli
		LPP	C, P	~500-3000 ms	
	Reward responsiveness	FRN / RewP	F, C	~200-300 ms	Reward tasks like simple gambling tasks (e.g., Doors task) and performance-contingent tasks (e.g., BART, MIDT, games like Blackjack) with non-SUDs-related stimuli
		Feedback-P3	F, C, P	~300-500 ms	
Negative Emotionality	Affective image processing	Cue-P3	C, P	~300-500 ms	Passive viewing tasks and active response tasks (e.g., Flanker, Go/No-Go, Stroop, oddball) with non-SUDs-related affective stimuli
		LPP	C, T, P	~500-3000 ms	
	EFE processing	N170 / VPP	T, O / F, C	~120-200 ms	Emotion judgement/discrimination tasks and other active response tasks (e.g., Flanker, Go/No-Go, Stroop, oddball) with affective facial stimuli
		Cue-P3	C, P	~300-500 ms	

Note. F = frontal; C = central; T = temporal; P = parietal; O = occipital; CPT = Continuous Performance Task; RHT = Rotated-Heads Task; BART = Balloon Analogue Risk Task; IGT = Iowa Gambling Task; MIDT = Monetary Incentive Delay Task; EFE = Emotional Facial Expression.

7.2.2 Data Extraction

The *p* values to be included in the analyses were extracted according to the guidelines presented in Simonsohn et al. (2014a) to ensure that they a) tested the finding of interest, b) were uniformly distributed under the null, and c) were statistically independent from each other. To help satisfy these requirements and to offer transparency on the *p* value selection process, *p-curve disclosure tables* were composed, which can be found in Tables S7.2 to S7.13 of the Supplementary Materials.

For each study, first the hypothesis and thus the finding of interest was identified. We focussed on studies examining the difference in ERP-amplitude between persons with SUDs en healthy controls on trials relevant to the processes reflected by that ERP (and, if applicable, relative to neutral trials). Specifically, for the N2 and P3 reflecting inhibitory control, this was the difference on trials requiring inhibition. For the P3s indexing attention allocation, it was the difference observed on non-target trials for the P3a (automatic attention) and on target trials for the P3b (deliberate attention). For the ERN and Pe, it was the difference on trials where participants made errors; for the cue-P3 and LPP, the difference on trials presenting substance-related or affective stimuli; for the FRN / RewP and feedback-P3, the difference on trials conveying (non-)rewarding feedback; and finally, for the N170 and P3 reflecting emotional face processing, it concerned the difference observed on trials presenting emotional faces.

Second, the study design was identified and used to determine the statistical result that should be included in the *p*-curve. For example, for a 2×2 interaction testing a reversed effect, the *p*-curve had to include the results of the simple effects, while for a 2×2 interaction testing an attenuated effect, the interaction result had to be included, as in that case the *p* values for simple effects are not uniformly distributed under the null. When multiple *p* values were reported for a finding of interest (for instance tests at different electrode sites), all *p* values were included in the disclosure table. Then, to satisfy *p*-curve's independence criterion, only one was selected for inclusion in the *p*-curve. The others were considered in the robustness analyses (see section 7.2.3.4). As a rule, unless the study's hypothesis was specific to a topographic region or timeframe, we selected the first significant *p* value that tested the finding of interest and that did not involve location or timing, or else the first significant *p* value testing the finding of interest at a relevant electrode site and time (see columns 4 and 5 in Table 7.1). When multiple hypotheses were tested, only the first was considered.

Next to *p* values, we extracted the experimental characteristics of the included studies. Specifically, we recorded included studies' sample size, mean participant age, percentage of

women, type of addiction, medical and psychiatric comorbidity, participants' use status, and the research paradigm used (see Tables S7.14 to S7.25 of the Supplementary Materials). Although this information is not used in the present meta-analysis, it may be of value to future studies wanting to compute narrower *p*-curves, for example only including studies on a specific SUDs or using a specific paradigm.

7.2.3 Data Analysis

7.2.3.1 Primary *P*-Curve Analysis. The *p*-curve analyses were performed using version 4.10 of the online app (Simonsohn, 2024). For each included result, the test statistic and degrees of freedom were entered (e.g., $F(1, 47) = 5.35$). Based on this information, the exact *p* value was computed. This was required as *p* values are often reported as "smaller than" (e.g., < .05) and are not always reported correctly (Bakker & Wicherts, 2011). Separate *p*-curves were performed for each ERP included in Table 7.1. To allow for robust conclusions about the observed effects, the analyses were conducted on *p* values of outcomes of the same direction, for example, all showing *lower* ERN amplitudes in individuals with SUDs compared to healthy controls, and none showing *higher* amplitudes. This way, we could conclude that evidential value existed for *poorer* error processing in SUDs, and not merely for a difference between persons with SUDs and healthy controls. When not all significant *p* values extracted for a particular *p*-curve comprised an effect of the same signage, two separate *p*-curves were conducted, one for each directionality.

To determine whether an ERP literature contained evidential value, the extracted *p* values were treated as test statistics themselves. For each significant *p* value, the probability of observing a *p* value at least as extreme if the null hypothesis were true, was computed. This is the *p* value of the *p* value: the *pp* value. These *pp* values were aggregated using Stouffer's Method (Simonsohn et al., 2015), yielding an overall χ^2 test for skew, which determines if the right skew of the *p*-curve is significant (vs. the H0 of a uniform distribution). In addition, to address "ambitious" *p*-hacking where researchers may have aimed at reporting the lowest *p* value as opposed to any *p* value below .05, a second test of the right skew was performed using only *p* values < .025 (i.e., the "half" *p*-curve). Because the half *p*-curve does not include barely significant findings, it is less likely to mistake ambitious *p*-hacking for evidential value. As proposed by Simonsohn et al. (2015), we concluded that a set of studies contained evidential value for the effect of interest either if the half *p*-curve was significantly right-skewed at the 5% level, or if both the full and half *p*-curves were significantly right-skewed at the 10% level.

When the tests for right skew were not significant, we further examined whether this null result was due to the set of studies *lacking evidential value*. To this end, it was examined whether the *p*-curve was flat (i.e., not right-skewed). As statistical inference cannot establish that a distribution is exactly flat, *p*-curve instead tests whether the observed distribution is significantly flatter (i.e., less right-skewed) than would be expected if the studies were powered at only 33%. In other words, it tests the null of a small effect (33% power) against the alternative of an even smaller effect (Simonsohn et al., 2014a). As proposed by Simonsohn et al. (2015), we concluded that a set of studies lacked evidential value either if the full *p*-curve was significantly flatter than that expected with 33% power, at the 5% level; or if both the half *p*-curve and a binomial test (high ($p > .025$) vs. low ($p < .025$) *p* values) of the flatness were significant at the 10% level. In these cases, either the effect of interest does not exist, or is too small to be reliably observed from the included studies. Finally, when a p-curve was neither significantly right-skewed nor significantly flat, the analysis was deemed *inconclusive*, that is, too noisy and/or underpowered to allow for inferences.

7.2.3.2 Sensitivity Analysis. To assess the extent to which the *p*-curves results' relied on a small number of included studies, we examined the impact of excluding the lowest *p* values from the original *p*-curve analyses. Specifically, for each analysis that showed evidential value, we first excluded the lowest included *p* value, then the second lowest, and so on, until we had excluded half of the originally included *p* values. Note that instead of simply rerunning the analyses with a smaller subset of *p* values, we followed Simonsohn et al. (2015) and tested whether the *pp* values were uniform in the range of possible values that *remained* (since excluding values evidently changes the distribution). The number of *p* values that could be dropped without the analysis' conclusion changing, was reported for each analysis.

7.2.3.3 Power Estimation. For each set of studies, the average, underlying selective-reporting-corrected statistical power was estimated. This was achieved by comparing the actual, observed *p*-curve to modelled *p*-curves for each possible value of statistical power between 5% and 99%, and identifying the closest resemblance. The modelled *p*-curves were created by computing *pp* values for the null that all included studies are powered at a given level (5%, 6%, … up to 99%), and aggregating those *pp* values using Stouffer's method. The best-fitting level of power was that which led to an overall Stouffer $Z = 0$, $p = .5$ (Simonsohn et al., 2014b; Simonsohn et al., 2015). This was considered the average statistical power of the set of studies. In addition, a 90% confidence interval (CI) for the power estimate was computed. Here, instead of determining which level of power led to a Stouffer test outcome of $p = .5$, it was examined which level of power gave $p = .05$ and $p = .95$, forming respectively the lower and upper bound of the CI.

7.2.3.4 Robustness Analysis. To examine whether the findings of the main analyses depended on the *p* value selection rules we set (see section 7.2.2), robustness *p*-curve analyses were computed, in which the *p* values that were originally selected were replaced – if applicable – by the *last* significant *p* value that was reported for the finding of interest. For example, if a significant difference between persons with SUDs and HCs on the cue-P3 in response to substance-related stimuli was reported for electrodes Cz, CPz, and Pz, all outcomes were noted down in the *p*-curve disclosure table (Tables S7.2 to S7.13), but only the first (as per our selection rules) was used in the main *p*-curve analysis. In the robustness analysis, we then replaced the result for Cz with the one for Pz. This served as a test of the robustness of the results to the selection rules: if the data supporting an ERP-amplitude difference between persons with SUDs and healthy controls were robust, then changing the selected *p* value in this manner should have limited impact on the overall result of the *p*-curve analysis.

7.3 Results

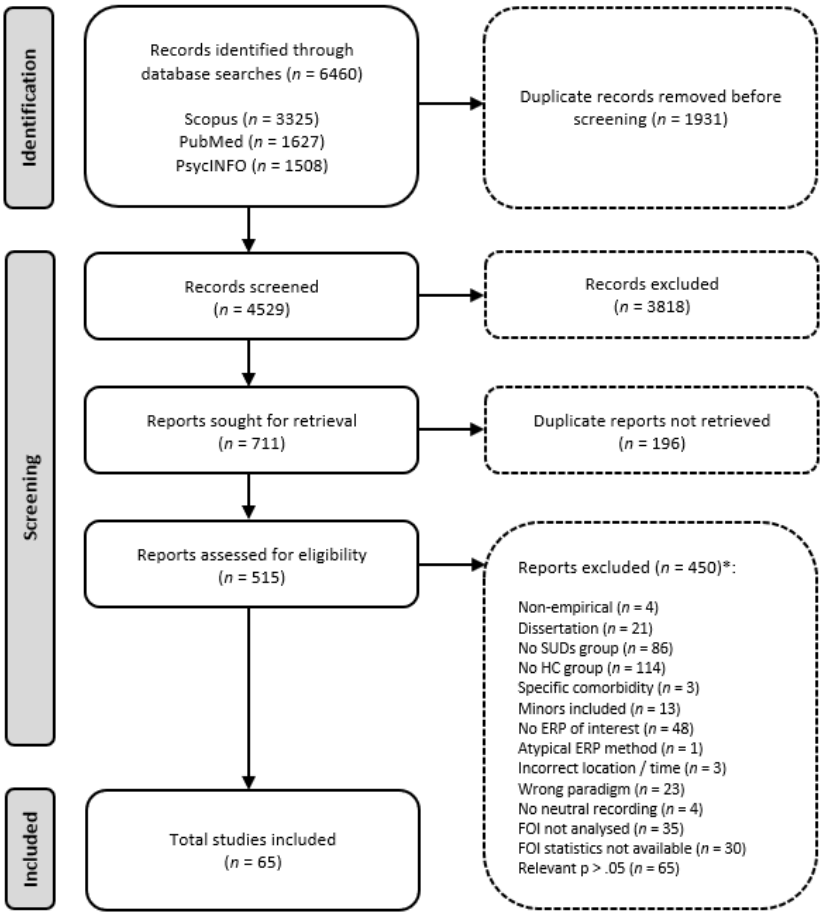
7.3.1 Search Results

As illustrated in Figure 7.1, the search of the three databases yielded a total of 6460 records, of which 4529 were retained after de-duplication using RefWorks' *Find Duplicates* function. The titles and abstracts of these records were screened, using the criteria set out in Section 7.2.1. When there was doubt regarding the inclusion of a record, for example because it was unclear from the title and abstract whether a control group was involved, the record was retained. From the original 4529 records, 711 remained, for which full texts were sought. These were retrieved for 515 records, with the remaining 196 records being duplicates which had not been detected by RefWorks' *Find Duplicates* function (among others due to incorrect author order, or unrecognised Greek letters). The 515 full texts were read and subsequently classified as "included" or "excluded", with the reason for exclusion being documented as well. Reports could frequently be excluded on the basis of multiple criteria, however, only the first criterium they failed to meet was recorded in order to prevent double counts. The largest number of exclusions was due to lack of a healthy control group (n = 114), either because one was not part of the study or because psychiatric or neurological disorders were reported to be present among members of the control group. After screening, n = 65 studies remained, published between 1990 and 2024. The most commonly studied addictions in this final dataset were alcohol use disorder (n = 21), internet gaming disorder (n = 12), and opioid use disorder (n = 9). About one-fourth of the included studies contained

findings on more than one ERP. The ERP most reported on was the P3b (n = 18), followed by the N2b and cue-P3 to drug cues (both n = 9). No studies on the affective image processing cue-P3 and LPP could be included, and for several other subdimensions (see Table 7.1) only few studies were available after exclusion. As a rule, we only computed *p*-curves for literatures (i.e., subdimensions) consisting of five or more studies to ensure a sufficient number of data points for aggregation and for plotting the curve. Still, details of all studies are provided in the *p*-curve disclosure tables (S7.2 to S7.13) and in the Experimental Characteristics tables (S7.14 to S7.25) in the Supplementary Materials.

Figure 7.1

PRISMA Flow Diagram of the Identification of Studies via Databases



Note. * Reports were excluded based on the first criterium they did not meet.

7.3.2 P-Curves

Table 7.2 provides a summary of the *p*-curve analyses outcomes, with Figure 7.2 providing a visual overview of all computed p-curves. Out of the 14 subdimensions reviewed in the Introduction Section, *p*-curves were computed for seven (i.e., seven subdimensions had five or more studies available): the N2b and No-Go P3 indicating Inhibitory Control; the P3b indicating Attention Allocation; the ERN indicating Error Processing; the Cue-P3 and LPP in response to drug-related stimuli indicating Drug Cue Salience; and the FRN / RewP indicating Reward Responsiveness. An overview of the input for all analyses can be found in Table S7.26 in the Supplementary Materials.

7.3.2.1 Inhibitory Control: N2b and No-Go P3. For the N2b, nine studies were available, of which eight reported attenuated and one reported stronger N2b amplitudes in individuals with SUDs compared to healthy controls. Therefore, a *p*-curve analysis was computed using the eight values indicating attenuated N2b's in SUDs. As detailed in Section 7.2.3.1, we concluded that a set of studies contained evidential value either if the half *p*-curve was significantly right-skewed at the 5% level, or if both the full and half *p*-curves were significantly right-skewed at the 10% level. Here, the half *p*-curve (based on six *p* values) was significantly right-skewed at 5% ($p < .001$), demonstrating evidential value for this literature. This conclusion was dependent on two very small *p* values, with the sensitivity analysis showing that only one *p* value could be removed before the set of studies did not show evidential value anymore. The underlying selective-reporting-corrected statistical power of the set of studies was good: 88%, 90% CI [63%, 97%]. Robustness analyses were not performed as no secondary values were extracted for the N2b analyses.

For the No-Go P3, eight studies were found that fulfilled the selection criteria, of which three reported stronger No-Go P3's and five reported attenuated No-Go P3's in persons with SUDs vs. healthy controls. Therefore, a *p*-curve analysis was computed using the five values showing an attenuated effect. Observed values were very spread out, resulting in a flat curve. Still, evidential value was present, as the full *p*-curve and the half *p*-curve (based on three *p* values) were both significantly right-skewed at the 10% level ($p = .072$ and $p = .058$). As was expected based on the observed distribution of *p* values, removal of the smallest value resulted in loss of evidential value. In other words, the significant right-skew was fully dependent on the inclusion of one study. Furthermore, the corrected average statistical power of the studies was low, 39%, 90% CI [5%, 86%], indicating a low probability to detect true effects. For the robustness analysis, one value was replaced, which did not change the conclusions.

7.3.2.2 Attention Allocation: P3a and P3b. For the P3a, only two studies were found that fulfilled the selection criteria, both reporting an attenuated P3a in persons with SUDs compared to controls. Since we only analysed literatures consisting of at least five studies, no *p*-curve analysis was done.

For the P3b, on the other hand, 18 studies were available, again all reporting attenuated amplitudes in SUDs compared to control groups. More small (e.g., .01) than large (e.g., .04) significant *p* values were observed, and the half *p*-curve (based on 14 *p* values) was significantly right-skewed at the 5% level ($p < .001$), demonstrating evidential value. The sensitivity analysis showed this conclusion to be robust against the exclusion of extremely low *p* values, as six values could be removed before the curve was not significantly right-skewed anymore. However, the corrected average statistical power of the studies was low: 53%, 90% CI [26%, 76%]. For the robustness analysis, five *p* values were replaced and one *p* value was excluded. This had no impact on the conclusions of either the main or sensitivity analysis, although power improved slightly to 60%.

7.3.2.3 Error Processing: ERN and Pe. For the ERN, eight studies were available, of which one reported stronger amplitudes and seven reported attenuated amplitudes in individuals with SUDs compared to healthy controls. Therefore, a *p*-curve analysis was performed using the seven available values that indicated an attenuated effect. The half *p*-curve (based on six *p* values) was significantly right-skewed at the 5% level ($p = .006$), thus demonstrating evidential value. Sensitivity analysis showed that one study could be dropped without loss of this evidential value. The underlying corrected statistical power of the set of studies was reasonable: 71%, 90% CI [30%, 92%]. For the robustness analysis, one value was replaced, which did not change any of the conclusions.

For the Pe, three studies were available, all reporting attenuated Pe amplitudes in persons with SUDs compared with healthy controls. Since we only analysed literatures consisting of at least five studies, no *p*-curve analysis was done.

7.3.2.4 Drug Cue Salience: Cue-P3 and LPP. For the P3 in response to SUDs-related stimuli, nine studies were found that fulfilled the selection criteria, all reporting stronger amplitudes in persons with SUDs compared to healthy controls. More small (e.g., .01) than large (e.g., .04) significant *p* values were observed, and the half *p*-curve (based on eight *p* values) was significantly right-skewed at the 5% level ($p = .006$), demonstrating evidential value. However, sensitivity analysis showed that this conclusion was fully dependent on the smallest *p* value that was included. Corrected average power of the studies was reasonable: 69%, 90% CI [30%, 91%]. For the robustness analysis, two *p* values were replaced, which did not change the main conclusions but did result in a considerable drop in power (to 56%).

For the LPP in response to SUDs-related stimuli, eight studies were available that all reported stronger amplitudes in individuals with SUDs compared to healthy controls. All included p values were < .025, and the half p-curve, which was thus based on all eight studies, was significantly right-skewed at the 5% level ($p < .001$), demonstrating evidential value. Sensitivity analysis showed this conclusion to be very robust against extreme values, as half of the included studies could be dropped without the literature losing evidential value (i.e., the p-curve was still significantly right-skewed after excluding the four lowest p values). Power was also very good: 92%, 90% CI [73%, 98%]. For the robustness analysis, six of the original eight values were replaced. This did not impact the evidential value, but did have a notable effect on the results of the sensitivity analysis and on the corrected average power: instead of four studies, now only one could be dropped without the literature losing evidential value; and instead of 92%, power was now estimated at 82%.

7.3.2.5 Reward Responsiveness: FRN and Feedback-P3. For the FRN/RewP, eight studies were found that fulfilled the selection criteria, of which two reported stronger FRN amplitudes in individuals with SUDs vs. health control subjects, and six reported attenuated amplitudes. Therefore, a p-curve analysis was done using the six values indicating an attenuated effect. Observed values formed two peaks, one clustering around .01, and one around .04. Evidential value was present based on the half p-curve (based on three p values) being significantly right-skewed at the 5% level ($p = .036$). However, sensitivity analysis showed that no studies could be dropped without resulting in loss of this evidential value, and the power estimate was extremely low: 23%, 90% CI [5%, 72%]. For the robustness analysis, one value was replaced, which did not change any of these conclusions.

For the Feedback-P3, four studies were found that fulfilled the selection criteria, of which one reported stronger Feedback-P3 amplitudes in individuals with SUDs vs. healthy controls, and the other three reported attenuated amplitudes. Since we only analysed literatures consisting of at least five studies, no p-curve analysis was done.

7.3.2.6 EFE Processing: N170 and EFE-P3. For the N170 and P3 in response to emotional facial expressions, respectively one and three studies fulfilled the inclusion criteria, all reporting attenuated amplitudes for persons with SUDs compared to healthy controls. Again, as we only analysed literatures consisting of at least five studies, no p-curve analyses were performed on these ERPs.

Table 7.2

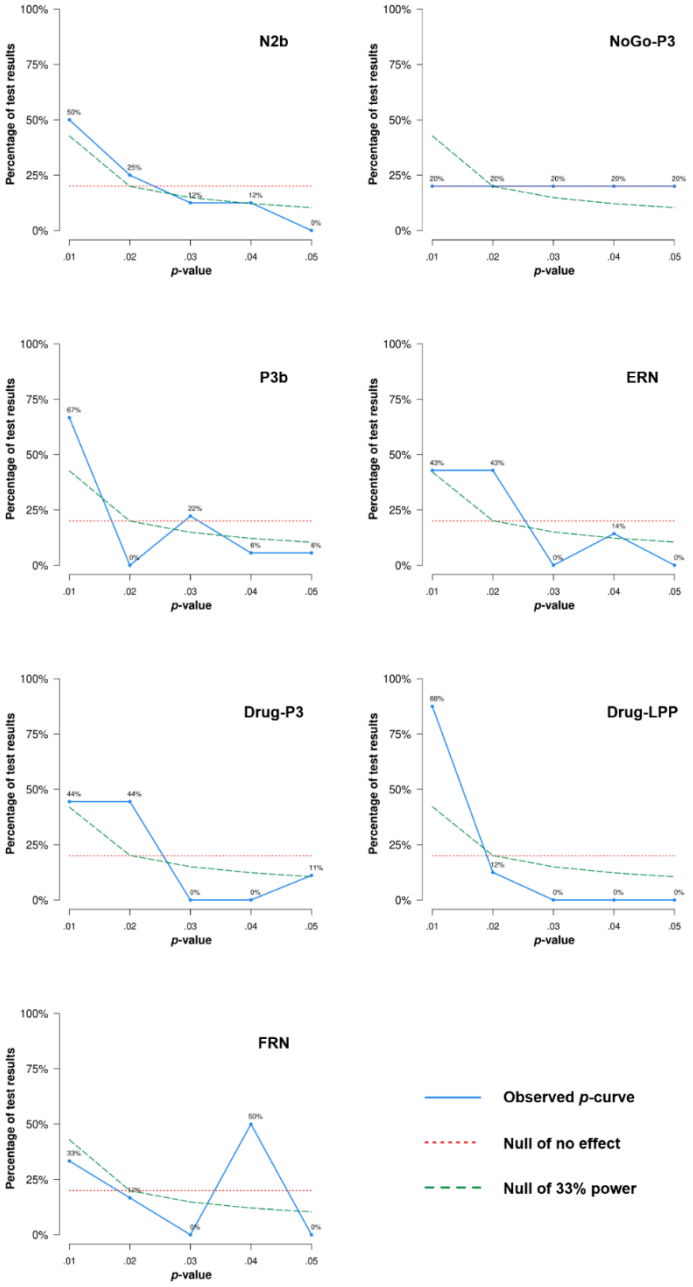
Summary of P-Curve Analyses Statistics

ERP	Direction	Main Analyses							Robustness analyses							
		# values	Evidential value		Lack of evidential value			Sensitivity	Power	# values	Evidential value		Lack of evidential value			Sensitivity	Power
			Full curve	Half curve	Full curve	Half curve	Binomial				Full curve	Half curve	Full curve	Half curve	Binomial		
N2b	Attenuated	8	Z = -4.90 p < .001	Z = -4.74 p < .001	Z = 2.90 p = .998	Z = 3.11 p = .999	p > .999	1	88% [63%, 97%]	-	-	-	-	-	-	-	-
No-Go P3	Attenuated	5	Z = -1.46 p = .072	Z = -1.57 p = .058	Z = 0.16 p = .564	Z = 0.63 p = .735	p = .988	0	39% [5%, 86%]	5	Z = -1.52 p = .065	Z = -1.70 p = .044	Z = 0.21 p = .584	Z = 0.74 p = .770	p = .988	0	41% [5%, 86%]
P3b	Attenuated	18	Z = -3.97 p < .001	Z = -3.43 p < .001	Z = 1.17 p = .880	Z = 1.19 p = .883	p > .999	6	53% [26%, 76%]	17	Z = -4.42 p < .001	Z = -3.37 p < .001	Z = 1.62 p = .947	Z = 1.06 p = .856	p > .999	6	60% [33%, 81%]
ERN	Attenuated	7	Z = -3.33 p < .001	Z = -2.54 p = .006	Z = 1.53 p = .937	Z = 1.08 p = .860	p > .999	1	71% [30%, 92%]	7	Z = -3.25 p < .001	Z = -2.41 p = .008	Z = 1.47 p = .929	Z = 0.97 p = .833	p > .999	1	70% [29%, 92%]
Drug-P3	Stronger	9	Z = -3.37 p < .001	Z = -2.53 p = .006	Z = 1.51 p = .935	Z = 1.03 p = .848	p > .999	0	69% [30%, 91%]	9	Z = -2.83 p = .002	Z = -2.01 p = .022	Z = 0.94 p = .826	Z = 0.47 p = .682	p > .999	0	56% [18%, 85%]
Drug-LPP	Stronger	8	Z = -5.60 p < .001	Z = -4.55 p < .001	Z = 3.51 p > .999	Z = 2.75 p = .997	p > .999	4 (50%)	92% [73%, 98%]	8	Z = -4.17 p < .001	Z = -2.95 p = .002	Z = 2.28 p = .989	Z = 1.45 p = .927	p > .999	1	82% [49%, 95%]
FRN	Attenuated	6	Z = -1.13 p = .129	Z = -1.80 p = .036	Z = -0.33 p = .369	Z = 0.72 p = .763	p = .970	0	23% [5%, 72%]	6	Z = -1.06 p = .145	Z = -1.68 p = .046	Z = -0.40 p = .344	Z = 0.61 p = .730	p = .970	0	21% [5%, 71%]

Note. Direction indicates how samples with SUDs scored in comparison with healthy controls. The sensitivity analysis reports the number of lowest p values that could be dropped without the analysis' conclusion regarding evidential value changing. The corrected power CI is 90%.

Figure 7.2

Overview of the Computed P-Curves



7.4 Discussion

The present study reviewed the literature on a large number of ERPs that have been proposed as biomarkers of addiction and performed *p*-curve analyses to examine whether these literatures have *evidential value* (i.e., represent genuine effects vs. selective reporting) and to estimate their selective-reporting-corrected statistical power. Both the review and *p*-curve analysis were structured along the dimensions that the Addiction Neuroclinical Assessment (ANA) framework proposes as central to the development and perpetuation of addiction: Executive Function, Incentive Salience, and Negative Emotionality.

The first ERPs we examined were those reflecting Executive Functions, beginning with the N2b and No-Go P3 reflecting *Inhibitory Control*. Here, existing literature showed lower N2b amplitudes in persons with addiction than in control subjects, but reported inconsistent findings for the No-Go P3 (Luijten et al., 2014; Y. Zhang et al., 2021). The *p*-curve analysis confirmed this: while eight out of the nine N2b studies identified for analysis showed attenuated amplitudes in persons with SUDs, this was only five out of eight for the No-Go P3, with the remaining three showing increased amplitudes in individuals with addiction. Furthermore, while we found evidential value for both ERPs, this was more robust for the N2b. Likewise, estimated power was higher for the N2b than for the No-Go P3. Regarding the ERPs reflecting *Attention Allocation*, we identified only a few studies for the P3a but a large number for the P3b. The P3a studies showed attenuated amplitudes in persons with SUDs compared to healthy controls, but no *p*-curve analysis was done due to the low number of studies. For the P3b, the review showed that studies consistently reported attenuated amplitudes in SUDs, which was supported by a meta-analysis (Euser et al., 2012). A *p*-curve analysis demonstrated that these findings could not be attributed to selective reporting, which was a robust finding. Surprisingly, however, corrected power for the P3b literature was low at around 50-60%. Finally, the ERN and Pe were considered, reflecting *Error Processing*. The review showed that previous studies reported attenuated amplitudes for both these ERPs in persons with addiction compared to control subjects, although meta-analyses differed in their conclusions (Y. Liu et al., 2023; Webber et al., 2024; Y. Zhang et al., 2021). The Pe was examined less than the ERN, which led to only the ERN being *p*-curved. This analysis demonstrated robust evidential value and reasonable power – thus indicating the existence of a genuine effect. Overall, across all six Executive Function ERPs we examined, the strongest evidence that the statistically significant findings reported by existing research reflect true neurobiological differences between persons with and persons without addiction was found for the N2b, the P3b, and the ERN.

The second series of ERPs we reviewed and applied *p*-curve analysis to were those concerned with Incentive Salience. First, we examined the *motivational value of stimuli associated with addiction*, like beer glasses or poker chips, as measured with the P3 and LPP. Existing literature showed that studies reported these to be higher in persons with addiction compared to healthy controls, which was corroborated by meta-analyses (Littel, Euser, et al., 2012; Norberg et al., 2016; Webber et al., 2022; Y. Zhang et al., 2021). The *p*-curve analysis demonstrated evidential value for both the P3 and LPP literature, although the robustness of this conclusion and the statistical power were more favourable for the LPP than for the P3. As a second aspect of Incentive Salience, we examined the *motivational value of non-drug rewards* as reflected by the FRN / RewP and the Feedback-P3. The majority of existing research showed these ERPs to be reduced in persons with SUDs compared to healthy control subjects, although the review also identified studies reporting the opposite effect, as well as a recent meta-analysis that found no effect for the FRN (Webber et al., 2024). For the Feedback-P3 too few studies were available for a *p*-curve analysis, and thus only the FRN literature was *p*-curved. While this analysis did find evidential value, this finding was not robust, and the average statistical power of this set of studies was poor. Overall, therefore, with regard to Incentive Salience, the strongest evidence of true neurobiological differences between persons with versus without addiction was found for the substance-related LPP followed by the substance-related P3.

Finally, we examined ERPs related to the third dimension of the ANA: Negative Emotionality. Here, we looked at the P3 and LPP when people processed *affective images*, and at the N170 and P3 when they processed *affective faces*. What stood out from the literature review on these ERPs was the inconsistency of the reported findings, plus the low number of studies investigating them in general (Feuerriegel et al., 2015; Webber et al., 2022). Consequently, no *p*-curve analysis was done for any of these ERPs, and contrary to the ERPs related to the other ANA dimensions, we cannot conclude which ones are and which ones are not likely to reflect genuine differences between persons with and without addiction. Rather, based on the available research, it seems most fitting to take an agnostic stance – not rejecting the notion that these ERPs differ based on addiction status, but not advocating this notion either. Moreover, we join the plea by Hoffman et al. (2019), who argue that impairments in social-emotional processing in people with SUDs have been underappreciated and under-researched, and recommend a push towards more empirical research on these constructs. By providing an overview of ERPs relevant to social-emotional processing and the paradigms in which they are studied, the present study offers a starting point for such an increase in research efforts.

The findings from the present study provide guidance for other future research directions too. Foremost, for ERPs which literatures show robust evidential value and high statistical power – the No-Go N2b and the drug-LPP – research should explore how they can be used in clinical practice. One of the key premises of the ANA framework and related approaches like the Research Domain Criteria (RDoC) project is that patients diagnosed with the same disorder, like SUDs, are highly heterogeneous, and that failure to take into account this heterogeneity contributes significantly to the low success rate of treatment (Insel et al., 2010; Kwako et al., 2016). For that reason, these frameworks argue that SUDs and other mental illnesses should be studied and treated based on the exact (neurobiological) processes that are disrupted in a person. As an example, two persons diagnosed with alcohol use disorder may have distinct profiles of dysfunction, regardless of having received the same DSM or ICD diagnostic label. The first person ("A") may have a strong salience for drug-related cues, but have only slightly compromised executive function and no negative affect bias; the second ("B") may likewise have a strong salience for alcohol cues but in addition have severe impairments in executive function (Kwako et al., 2018). When treating these two individuals based on their overall diagnosis, treatment will be generic as no distinction can be made. Instead, when looking at the neurobiological processes that are disrupted in each person, treatment can be *targeted*, for example by focussing on improving inhibitory control for person "B". ERPs are one way through which these disruptions can be identified (and possibly monitored) during treatment, provided that the research that informs the validity of an ERP for measuring these constructs reflects genuine effects.

Relatedly, in order to know which ERPs can be used to safely measure these constructs, the findings from the present study demonstrate that continued efforts are required to improve the statistical power of studies. For many of the literatures included in the present study, corrected power was insufficient: only for the N2b and drug-LPP did power surpass 80%. Across all reviewed literatures, samples on average consisted of 30 participants per group, which is only adequate for detecting large effects (Clayson et al., 2019; Kissel & Friedman, 2023) and which limits the ability of research to identify replicable biomarkers, as low power decreases the informational value of both significant and non-significant findings (Lakens & Evers, 2014). Thus, if any biomarkers of SUDs, other mental illnesses, or the processes that underlie them are to be identified, higher powered studies are essential. Inherently linked to this is the current lack of (a priori) power analyses in the EEG literature, which bars readers from judging the replicability of studies themselves (Larson & Carbine, 2017).

The present research has shortcomings of its own too. Primarily, despite attempting to run *p*-curve analyses on all reviewed ERPs, thus assessing evidential value and power for ERPs across all ANA dimensions, we only succeeded in doing so for seven: half of the initially intended effects. This can be attributed to various factors. First, we applied relatively stringent inclusion criteria. For example, we required the control group to be free of any psychiatric or neurological disorders, meaning that studies were excluded as soon as one member of the control group was diagnosed with such a disorder. Likewise, we required those with addiction to be formally diagnosed or identified through validated measures, among others resulting in the exclusion of several studies focussing on heavy cannabis use (e.g., Henry et al., 2014). Notably, these stringent criteria served a critical purpose: keeping heterogeneity low, as *p*-curve, like other meta-analytic approaches, decreases in performance when effect sizes are heterogeneous (Brunner & Schimmack, 2020; E. C. Carter et al., 2019; McShane et al., 2016), leading to the general recommendation to "select homogeneous subsets of studies on the basis of methodological or substantive characteristics" and meta-analyse studies only when they "are reasonably expected to have one underlying population effect" (Van Assen et al., 2015, p. 298 and p. 306). This also makes sense from a theoretical point of view: when the studies included in a meta-analysis differ widely in populations, measures, dependent variables or even research questions, one can wonder what is actually being estimated and hence what the outcome of the meta-analysis actually represents (Simonsohn et al., 2022). Thus, our more selective analyses, on specific ERPs elicited in specific tasks and in specific populations, are more straightforward to interpret and more informative compared to if we had employed a more lenient inclusion.

A second reason why many potential studies were excluded is related to limitations in reporting. In non-*p* value based meta-analyses, the key statistics (effect sizes) can be computed in various ways and hence few studies are excluded due to the required statistics being unavailable. On the other hand, for *p*-curve analysis and associated methods, the options are much more limited: one needs to recompute the *p* value, which can only be achieved by using the *t* or *F* statistic and the *degrees of freedom*. Unfortunately, these were not always available from the published report, and – as we experienced when emailing authors – are not always stored carefully either, resulting in a relatively large number of exclusions based on the relevant statistics not being available. We see this not so much as a limitation of the *p*-curve methodology, but more so as a limitation of our field. In many cases, full statistics were only provided for the omnibus test, regardless of whether that tested the hypothesis. As an example, studies on the ERN and Pe virtually always hypothesised a group difference on error trials. However, when researchers then performed a 2 × 2 ANOVA with group (addiction, healthy) and trial (error, correct) as factors, they often only provided

the *F* and *df* for the main effects and interactions. For the follow-up test to see if groups differed on the error trials only – the actual hypothesis – many times only the *p* value, or an approximation of this (e.g., < .05), was reported, which was insufficient for inclusion in the *p*-curve analysis.

Finally, not all intended *p*-curve analyses were performed due to choices made to protect against low precision of our analyses. Specifically, we decided against running *p*-curve analyses on the P3a, Pe, Feedback-P3, N170, and EFE-P3, as we believed these analyses would not result in informative outcomes due to the low number of studies available (one to three). How often *p*-curve is wrong depends on the statistical power of the included studies, on the degree of selective reporting in the set, and on the number of studies being *p*-curved (Simonsohn et al., 2014a). Simonsohn et al. (2014a) present simulations where, when studies are powered well (80%), *p*-curve almost never falsely concludes that the set lacks evidential value, even when as few as ten studies are included. However, as evidenced, EEG studies are frequently not well-powered, and therefore also the No-Go P3 *p*-curve (based on five studies) and the FRN / RewP *p*-curve (based on six) should be interpreted with caution. This is particularly true for the former, which (just about) showed evidential value despite an almost flat curve. Here it is important to update the analysis when new findings are published to see if the evidential value remains. More generally speaking, it is important to recognise that meta-analyses do not provide definitive conclusions (E. C. Carter et al., 2019): the present *p*-curve analyses should be treated as a beginning rather than an end of the examination into the evidential value of SUDs-related ERPs, with newer data and different methodologies incrementally contributing to answering this question.

7.5 Supplementary Materials

Table S7.1

Search Queries for Scopus, PubMed, and PsycInfo

Database	Full Query
Scopus	((TITLE-ABS-KEY ("Substance abuse") OR TITLE-ABS-KEY ("Substance use") OR TITLE-ABS-KEY ("Substance dependence") OR TITLE-ABS-KEY ("Drug abuse") OR TITLE-ABS-KEY ("Drug use") OR TITLE-ABS-KEY ("Drug dependence") OR TITLE-ABS-KEY (addict*) OR TITLE-ABS-KEY (alcohol*) OR TITLE-ABS-KEY (cocaine) OR TITLE-ABS-KEY (crack) OR TITLE-ABS-KEY (ghb) OR TITLE-ABS-KEY (heroin) OR TITLE-ABS-KEY (lsd) OR TITLE-ABS-KEY (methamphetamine) OR TITLE-ABS-KEY (opiate*) OR TITLE-ABS-KEY (opioid*) OR TITLE-ABS-KEY (mdma) OR TITLE-ABS-KEY (ecstasy) OR TITLE-ABS-KEY (cannabis) OR TITLE-ABS-KEY (marijuana) OR TITLE-ABS-KEY (nicotine) OR TITLE-ABS-KEY (cigarette) OR TITLE-ABS-KEY (tobacco) OR TITLE-ABS-KEY (smoking) OR TITLE-ABS-KEY (smoker*) OR TITLE-ABS-KEY (gambling) OR TITLE-ABS-KEY (gambler*) OR TITLE-ABS-KEY ("Gaming disorder") OR TITLE-ABS-KEY ("Computer gam*") OR TITLE-ABS-KEY ("Problematic internet") OR TITLE-ABS-KEY ("Problematic smartphone") OR TITLE-ABS-KEY ("Problematic porn*") OR TITLE-ABS-KEY ("Problematic social"))) AND ((TITLE-ABS-KEY (electroencephalogra*) OR TITLE-ABS-KEY (eeg) OR TITLE-ABS-KEY ("Event-related potential*") OR TITLE-ABS-KEY (erp) OR TITLE-ABS-KEY (n2) OR TITLE-ABS-KEY (n200) OR TITLE-ABS-KEY (n2b) OR TITLE-ABS-KEY (p3) OR TITLE-ABS-KEY (p300) OR TITLE-ABS-KEY (p3a) OR TITLE-ABS-KEY (p3b) OR TITLE-ABS-KEY ("Novelty P3") OR TITLE-ABS-KEY ("Error-related negativity") OR TITLE-ABS-KEY (ern) OR TITLE-ABS-KEY ("Error negativity") OR TITLE-ABS-KEY ("Error positivity") OR TITLE-ABS-KEY ("Late positive potential") OR TITLE-ABS-KEY (lpp) OR TITLE-ABS-KEY ("Late positive comp*") OR TITLE-ABS-KEY (lpc) OR TITLE-ABS-KEY ("Slow wave") OR TITLE-ABS-KEY ("Slow potential") OR TITLE-ABS-KEY ("Slow positive wave*") OR TITLE-ABS-KEY ("Feedback-related negativity") OR TITLE-ABS-KEY (frn) OR TITLE-ABS-KEY ("Feedback negativity") OR TITLE-ABS-KEY ("Reward positivity") OR TITLE-ABS-KEY (rewp) OR TITLE-ABS-KEY (n170) OR TITLE-ABS-KEY (p150) OR TITLE-ABS-KEY (p160) OR TITLE-ABS-KEY ("Vertex Positive Potential") OR TITLE-ABS-KEY (vpp))) AND ((TITLE-ABS-KEY ("Executive function") OR TITLE-ABS-KEY ("Executive control") OR TITLE-ABS-KEY ("Cognitive control") OR TITLE-ABS-KEY ("Information processing") OR TITLE-ABS-KEY ("Inhibitory control") OR TITLE-ABS-KEY (inhibit*) OR TITLE-ABS-KEY (no-go) OR TITLE-ABS-KEY (stop-signal) OR TITLE-ABS-KEY ("Attention allocation") OR TITLE-ABS-KEY ("Resource allocation") OR TITLE-ABS-KEY ("Attentional control") OR TITLE-ABS-KEY ("Selective attention") OR TITLE-ABS-KEY (oddball) OR TITLE-ABS-KEY (non-target) OR TITLE-ABS-KEY ("Continuous performance t*") OR TITLE-ABS-KEY (cpt) OR TITLE-ABS-KEY ("Rotated-heads") OR TITLE-ABS-KEY (rht) OR TITLE-ABS-KEY ("discrimination task") OR TITLE-ABS-KEY ("Error processing") OR TITLE-ABS-KEY ("Error monitoring") OR TITLE-ABS-KEY ("Performance monitoring") OR TITLE-ABS-KEY ("Conflict monitoring") OR TITLE-ABS-KEY (flanker) OR TITLE-ABS-KEY ("Stroop task") OR TITLE-ABS-KEY ("Simon task") OR TITLE-ABS-KEY ("Incentive salience") OR TITLE-ABS-KEY (cue-reactivity) OR TITLE-ABS-KEY ("Attention* bias*") OR TITLE-ABS-KEY ("Processing bias*") OR TITLE-ABS-KEY ("Motivational bias*") OR TITLE-ABS-KEY ("Motivated attention") OR TITLE-ABS-KEY ("Motivational signific*") OR TITLE-ABS-KEY (craving) OR TITLE-ABS-KEY (dot-probe) OR TITLE-ABS-KEY ("Reward respons*") OR TITLE-ABS-KEY ("Reward process*") OR TITLE-ABS-KEY ("Reward sensit*") OR TITLE-ABS-KEY (reward-related) OR TITLE-ABS-KEY ("Feedback processing") OR TITLE-ABS-KEY ("Gambling task") OR TITLE-ABS-KEY ("Doors task") OR TITLE-ABS-KEY ("Balloon Analogue Risk Task") OR TITLE-ABS-KEY (bart) OR TITLE-ABS-KEY ("Iowa Gambling Task") OR TITLE-ABS-KEY (igt) OR TITLE-ABS-KEY ("Monetary Incentive Delay Task") OR TITLE-ABS-KEY (midt) OR TITLE-ABS-KEY ("Negative emotionality") OR TITLE-ABS-KEY ("Emotional salience") OR TITLE-ABS-KEY ("Pleasant image*") OR TITLE-ABS-KEY ("Unpleasant image*") OR TITLE-ABS-KEY ("Neutral image*") OR TITLE-ABS-KEY ("Affective stimul*") OR TITLE-ABS-KEY ("Emotion* stimul*") OR TITLE-ABS-KEY ("Emotion* process*") OR TITLE-ABS-KEY ("Emotional Facial Expression*") OR TITLE-ABS-KEY ("Facial expression*") OR TITLE-ABS-KEY ("Facial process*") OR TITLE-ABS-KEY ("Face-process*"))) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (EXACTKEYWORD , "Human") OR LIMIT-TO (EXACTKEYWORD , "Humans"))

Table S7.1 (continued)

Search Queries for Scopus, PubMed, and PsycInfo

Database	Full Query
PubMed	("Substance abuse"[tw] OR "Substance use"[tw] OR "Substance dependence"[tw] OR "Drug abuse"[tw] OR "Drug use"[tw] OR "Drug dependence"[tw] OR addict*[tw] OR alcohol*[tw] OR cocaine[tw] OR crack[tw] OR ghb[tw] OR heroin[tw] OR lsd[tw] OR methamphetamine[tw] OR opiate*[tw] OR opioid*[tw] OR mdma[tw] OR ecstasy[tw] OR cannabis[tw] OR marijuana[tw] OR nicotine[tw] OR cigarette[tw] OR tobacco[tw] OR smoking[tw] OR smoker*[tw] OR gambling[tw] OR gambler*[tw] OR "Gaming disorder"[tw] OR "Computer gam*"[tw] OR "Problematic internet"[tw] OR "Problematic smartphone"[tw] OR "Problematic porn*"[tw] OR "Problematic social"[tw]) AND (electroencephalogra*[tw] OR eeg[tw] OR "Event-related potential*"[tw] OR erp[tw] OR n2[tw] OR n200[tw] OR n2b[tw] OR p3[tw] OR p300[tw] OR p3a[tw] OR p3b[tw] OR "Novelty P3"[tw] OR "Error-related negativity"[tw] OR ern[tw] OR "Error negativity"[tw] OR "Error positivity"[tw] OR "Late positive potential"[tw] OR lpp[tw] OR "Late positive comp*"[tw] OR lpc[tw] OR "Slow wave"[tw] OR "Slow potential"[tw] OR "Slow positive wave*"[tw] OR "Feedback-related negativity"[tw] OR frn[tw] OR "Feedback negativity"[tw] OR "Reward positivity"[tw] OR rewp[tw] OR n170[tw] OR p150[tw] OR p160[tw] OR "Vertex Positive Potential"[tw] OR vpp[tw]) AND ("Executive function"[tw] OR "Executive control"[tw] OR "Cognitive control"[tw] OR "Information processing"[tw] OR "Inhibitory control"[tw] OR inhibit*[tw] OR no-go[tw] OR stop-signal[tw] OR "Attention allocation"[tw] OR "Resource allocation"[tw] OR "Attentional control"[tw] OR "Selective attention"[tw] OR oddball[tw] OR non-target[tw] OR "Continuous performance t*"[tw] OR cpt[tw] OR "Rotated-heads"[tw] OR rht[tw] OR "discrimination task"[tw] OR "Error processing"[tw] OR "Error monitoring"[tw] OR "Performance monitoring"[tw] OR "Conflict monitoring"[tw] OR flanker[tw] OR "Stroop task"[tw] OR "Simon task"[tw] OR "Incentive salience"[tw] OR cue-reactivity[tw] OR "Attention* bias*"[tw] OR "Processing bias*"[tw] OR "Motivational bias*"[tw] OR "Motivated attention"[tw] OR "Motivational signific*"[tw] OR craving[tw] OR dot-probe[tw] OR "Reward respons*"[tw] OR "Reward process*"[tw] OR "Reward sensit*"[tw] OR reward-related[tw] OR "Feedback processing"[tw] OR "Gambling task"[tw] OR "Doors task"[tw] OR "Balloon Analogue Risk Task"[tw] OR bart[tw] OR "Iowa Gambling Task"[tw] OR igt[tw] OR "Monetary Incentive Delay Task"[tw] OR midt[tw] OR "Negative emotionality"[tw] OR "Emotional salience"[tw] OR "Pleasant image*"[tw] OR "Unpleasant image*"[tw] OR "Neutral image*"[tw] OR "Affective stimul*"[tw] OR "Emotion* stimul*"[tw] OR "Emotion* process*"[tw] OR "Emotional Facial Expression*"[tw] OR "Facial expression*"[tw] OR "Facial process*"[tw] OR "Face-process*"[tw]) AND "english"[la] AND humans[mh]
PsychINFO	(("Substance abuse" or "Substance use" or "Substance dependence" or "Drug abuse" or "Drug use" or "Drug dependence" or addict* or alcohol* or cocaine or crack or ghb or heroin or lsd or methamphetamine or opiate* or opioid* or mdma or ecstasy or cannabis or marijuana or nicotine or cigarette or tobacco or smoking or smoker* or gambling or gambler* or "Gaming disorder" or "Computer gam*" or "Problematic internet" or "Problematic smartphone" or "Problematic porn*" or "Problematic social") and (electroencephalogra* or eeg or "Event-related potential*" or erp or n2 or n200 or n2b or p3 or p300 or p3a or p3b or "Novelty P3" or "Error-related negativity" or ern or "Error negativity" or "Error positivity" or "Late positive potential" or lpp or "Late positive comp*" or lpc or "Slow wave" or "Slow potential" or "Slow positive wave*" or "Feedback-related negativity" or frn or "Feedback negativity" or "Reward positivity" or rewp or n170 or p150 or p160 or "Vertex Positive Potential" or vpp) and ("Executive function" or "Executive control" or "Cognitive control" or "Information processing" or "Inhibitory control" or inhibit* or no-go or stop-signal or "Attention allocation" or "Resource allocation" or "Attentional control" or "Selective attention" or oddball or non-target or "Continuous performance t*" or cpt or "Rotated-heads" or rht or "discrimination task" or "Error processing" or "Error monitoring" or "Performance monitoring" or "Conflict monitoring" or flanker or "Stroop task" or "Simon task" or "Incentive salience" or cue-reactivity or "Attention* bias*" or "Processing bias*" or "Motivational bias*" or "Motivated attention" or "Motivational signific*" or craving or dot-probe or "Reward respons*" or "Reward process*" or "Reward sensit*" or reward-related or "Feedback processing" or "Gambling task" or "Doors task" or "Balloon Analogue Risk Task" or bart or "Iowa Gambling Task" or igt or "Monetary Incentive Delay Task" or midt or "Negative emotionality" or "Emotional salience" or "Pleasant image*" or "Unpleasant image*" or "Neutral image*" or "Affective stimul*" or "Emotion* stimul*" or "Emotion* process*" or "Emotional Facial Expression*" or "Facial expression*" or "Facial process*" or "Face-process*")).mp. and english.lg. and human.po.

Table S7.2

P-curve Disclosure Table N2b

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
J. Chen et al. (2016)	We hypothesised that **relative to controls, excessive smartphone users would have a deficit in inhibitory control**, both **in blank context** and in the smartphone- related context (SC).	Repeated Measures Analyses of Variance (RM-ANOVA; with Greenhouse-Geisser adjusted *p*-values) were applied to analyse behavioural outcomes of performance on the Go/NoGo task, as well as ERP as the index of response inhibition. Simple effects were explored and interaction sources were systematically examined.	Difference of means (SUDs vs. HC)	Further analysis indicated a significant difference between the smartphone overuse and normal user groups in the NoGo condition; the excessive smartphone use group elicited significantly larger N2 mean amplitude than the normal users group in blank [$F(1, 30) = 13.57$, $p = 0.001$, $\eta_p^2 = 0.31$], neutral [$F(1, 30) = 8.67$, $p = 0.006$, $\eta_p^2 = 0.22$], and smartphone-related [$F(1, 30) = 12.40$, $p = 0.001$, $\eta_p^2 = 0.30$] contexts.	$F(1, 30) = 13.57$ $p = 0.00090$	N/A
Y. Chen et al. (2022)	Hypothesis 3. **Compared to the control group, the GD group shows differences in nogo-N2** and nogo-P3 **amplitudes during response inhibition.** This effect might be **stronger when processing task-unrelated emotional information rather than task-related information**.	A four-way repeated-measures analysis of variance (ANOVA) was conducted with respect to group (GD group and control group), task (unrelated and related), emotional valence (negative and positive), and electrode point (14 sites).	2-way interaction (attenuated)	Based on hypothesis 3, the task x group interaction was significant ($F(1, 47) = 5.35$, $p = 0.025$, $\eta_p^2 = 0.10$). The nogo-N2 amplitudes were significantly lower in the GD group than in the control group in the task-unrelated condition but not significantly different in the task-related condition.	$F(1, 47) = 5.35$ $p = 0.02515$	N/A
Dong et al. (2010)	**Internet-addicted participants were expected to show some difference in N2** and P3 **compared with their normal peers.**	ERP amplitudes and latencies were analysed using repeated measures ANOVA with electrode sites (frontal / central / parietal) × stimulus (Go / NoGo) as within-subject factors and group (Internet addicted / normal) as a between-subjects factor.	Difference of means (SUDs vs. HC)	Significant difference was found between IAD and normal groups in NoGo condition, the IAD group elicited significant lower N2 mean amplitude than normal group [$F(1,22) = 6.92$, $p < 0.05$].	$F(1, 22) = 6.92$ $p = 0.01527$	N/A
Fathi et al. (2022)	This study aimed to **investigate** both **proactive** and reactive **inhibitions by selective stop-signal task in individuals with VGA** who play action video games.	A mixed effects repeated measures analysis of variance (ANOVA) was conducted for each ERP with two within-subject factors: (i) trial type (Go, NoGo, and irrelevant trials) and, (ii) electrode locations (Fz, F3, F4, Cz, C3, and C4).	Difference of means (SUDs vs. HC)	In proactive stop trials, results showed that there were significant main effects of group [$F(1, 60) = 11.324$, $p = .001$, $\eta^2 = 0.156$], stop direction [$F(1, 60) = 19.18$, $p < .000$, $\eta^2 = 0.242$], and interaction of group with stop direction [$F(1, 60) = 45.932$, $p < .000$, $\eta^2 = 0.434$] on N2 amplitude. Follow-up analyses indicated that the VGA group had significantly smaller N2 amplitude than the control group in stop right trials ($p < .000$), but there was no significant difference in stop left trials.	$F(1, 60) = 11.324$ $p = 0.00134$	n.a.
Fathi et al. (2024)	We hypothesised that **individuals with VGA will exhibit impaired inhibitory control processes**, specifically in the preparatory process and response inhibition (…). This impairment may be **reflected by a reduction in N2 amplitude**.	A mixed effects two-way repeated measures analysis of variance was conducted for each ERP with two within-subject factors: (i) trial type (Go, NoGo, and irrelevant trials) and (ii) electrode locations (Fz, F3, F4, Cz, C3, and C4).	Difference of means (SUDs vs. HC)	Group had a significant main effect on N2 amplitude in the NoGo trials ($F(1, 46) = 6.133$, $p = 0.017$, $\eta^2 = 0.118$), in that the N2 amplitude of the NoGo trial was smaller in the VGA group compared with the control group, with the largest difference being in F4 electrode.	$F(1, 46) = 6.133$ $p = 0.01701$	N/A
Morie et al. (2014)	We hypothesised that **current cocaine abusers would demonstrate** reduction in task accuracy, reduced post error slowing, and **attenuation of ERP components related to inhibition**, performance monitoring and error awareness.	For the stimulus-locked N2 and P3, a 2 × 2 ANOVA was run for each with factors of Group (cocaine user versus controls) and Response Type (correct Go-response versus correct No-Go response).	Difference of means (SUDs vs. HC)	The ANOVA for the N2 revealed a main effect of response type ($F49 = 8.1$, $p < .01$) and a main effect of group ($F49 = 7.1$, $p < .02$), with an interaction of response type and group ($F49 = 7.0$, $p < .02$). Pairwise comparisons revealed the N2 amplitude to be smaller in the addicts than in the controls ($t49 = 5.8$, $p < .01$).	$t(49) = 5.8$ $p = 4.75098e-07$	N/A

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.2 (continued)

P-curve Disclosure Table N2b

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Pandey et al. (2012)	The aim of the present study was to evaluate **the N2 differences between alcoholic and normal control subjects** as well as between task conditions using an equal probability Go/NoGo task.	The mixed-effects model included group (controls, alcoholics), condition (Go/NoGo), region (frontal, central, parietal, occipital, left temporal, and right temporal), electrode (6 representative electrode sites nested within each region) age, age^2 and their interactions as fixed effects, where condition (Go/NoGo) and electrode coordinates (x, y, z) were treated as repeated measures.	Difference of means (SUDs vs. HC)	Post hoc analyses of the significant interaction effects revealed significant pair wise differences. For the Group x Condition interaction effects, alcoholics showed reduced N2 amplitudes compared to controls for both Go ($t(134) = -5.22$, $p < 0.0001$) and NoGo (**$t(134) = -7.21$, $p < 0.0001$**) conditions.	$t(134) = -7.21$ $p = 3.69524e\text{-}11$	N/A
Sokhadze et al. (2008)	**Patient group differences in inhibition efficiency were predicted to be reflected in a lower amplitude of the fronto-central ERP indices of the inhibition** (so-called anterior **NoGo-N2** and NoGo-P3).	Data for each EEG-based dependent variable are analysed using repeated measures ANOVA (SPSS 14.0). Within-subject factors are stimulus (Go-target, NoGo), congruence (congruent, incongruent), and error (commission, omission) type for the ERN. The between-subject factors is group (SUD, controls).	Difference of means (SUDs vs. HC)	A one-way ANOVA showed significant differences between the SUD and CNT groups only in congruent NoGo-N2 waves. Amplitude of the difference wave was lower in the SUD group compared to controls (0.32 ± 0.85 vs. –0.65 ± 1.56 μV, **$F = 4.90$, $p = .035$**). *REQUESTED: df belonging to reported F.*	$F(1, 32) = 4.90$ $p = 0.03411$	N/A
B. Yang et al. (2009)	We hypothesized that **heroin addicts would exhibit attenuated Nogo ERP effects in the fronto-central areas compared with healthy controls**.	A Site (5 midline sites: Fz, Fcz, Cz, Cpz, Pz) × Group (addicts, controls) repeated measures ANOVA was performed for peak N2d and P3d amplitudes.	Difference of means (SUDs vs. HC)	The repeated-measures ANOVA on N2 amplitudes of the Nogo-Go difference wave revealed a significant main effect of the Group (N2d controls = (−4.98 ± 0.57) μV, N2d addicts = (−2.72 ± 0.57) μV, **$F(1, 126) = 7.81$, $P < 0.01$,** $\eta^2 = 0.23$).	$F(1, 126) = 7.81$ $p = 0.00601$	N/A

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.3

P-curve Disclosure Table No-Go P3

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Y. Chen et al. (2022)	Hypothesis 3. **Compared to the control group, the GD group shows differences in** nogo-N2 and **nogo-P3 amplitudes during response inhibition**. This effect might be **stronger when processing task-unrelated emotional information rather than task-related information**.	A four-way repeated-measures analysis of variance (ANOVA) was conducted with respect to group (GD group and control group), task (unrelated and related), emotional valence (negative and positive), and electrode point (14 sites).	2-way interaction (attenuated)	Based on hypothesis 3, the task x group interaction was significant (**$F(1, 47) = 4.11$, $p = 0.048$**, $\eta_p^2 = 0.08$). Nogo-P3 amplitudes were significantly higher in the GD group than in the control group in the task-unrelated condition but not significantly different in the task-related condition.	$F(1, 47) = 4.11$ $p = 0.04832$	N/A
Colrain et al. (2011)	Consistent with other measures of inhibitory control (...), **NOGO P3 responses have been shown to be diminished in alcoholics**.	When an overall MANCOVA model showed significant effect of diagnosis, the relevant univariate ANCOVA values were then evaluated for each dependent variable.	Difference of means (SUDs vs. HC)	The univariate models showed alcoholics to have smaller P3 amplitude at Cz than controls (**$F(1, 30) = 4.73$, $p=0.038$**; Fig. 1).	$F(1, 30) = 4.73$ $p = 0.03766$	N/A
Dong et al. (2010)	**Internet-addicted participants were expected to show some difference in** N2 and **P3 compared with their normal peers.**	ERP amplitudes and latencies were analysed using repeated measures ANOVA with electrode sites (frontal / central / parietal) × stimulus (Go / NoGo) as within-subject factors and group (Internet addicted / normal) as a between-subjects factor.	Difference of means (SUDs vs. HC)	IAD group showed significant higher P3 amplitude than normal group in NoGo items [**$F(1, 22) = 6.43$, $p < 0.05$**].	$F(1, 22) = 6.43$ $p = 0.01883$	N/A
Fathi et al. (2022)	This study aimed to **investigate** both **proactive** and reactive **inhibitions by selective stop-signal task in individuals with VGA** who play action video games.	A mixed effects repeated measures analysis of variance (ANOVA) was conducted for each ERP with two within-subject factors: (i) trial type (Go, NoGo, and irrelevant trials) and, (ii) electrode locations (Fz, F3, F4, Cz, C3, and C4).	Difference of means (SUDs vs. HC)	The analysis of proactive stop trials showed a significant main effect of group on P3 amplitude [**$F(1, 133) = 4.66$, $p = .033$**, $\eta^2 = 0.034$].	$F(1, 133) = 4.66$ $p = 0.03267$	N/A
Ghin et al. (2022)	We examined (…) ERP components, which have repeatedly been shown to reflect distinct subprocesses involved in **inhibitory control**. Because the N2 and **P3** reflect distinguishable response inhibition subprocesses, we focused on the N2-P3 ERP complex in the current study.	We used mixed-effects ANOVAs with condition (Go vs. Nogo) and congruency (congruent vs. incongruent) as within-subject factors and group (AUD vs. control) as between-subject factors.	Difference of means (SUDs vs. HC)	The control group showed larger P3 amplitudes (5.33 $\mu V/m^2 \pm 0.30$) than the AUD group (4.35 $\mu V/m^2 \pm 0.30$) in incongruent Nogo trials (**$t = 2.298$, $p = 0.046$**). *REQUESTED: df belonging to reported t.*	$t(121) = 2.298$ $p = 0.02328$ Note: *p* value in paper is Bonferroni-corrected (*2)	N/A
Maij et al. (2017)	We expected to find **reduced response inhibition in CUD patients compared with controls as reflected by reduced NoGo**-N2 and **-P3 components**.	To analyse the ERPs of the Go/NoGo Task, a Group × Inhibition × Electrode RM-ANOVA was conducted.	Difference of means (SUDs vs. HC)	Post-hoc tests revealed that this was due to the CUD patients having lower NoGo P3 amplitudes than non-smokers: mean difference = 2.44 (*SE* = 0.98), **$p = .044$**. CUD patients who had lower P3 amplitudes on the Cz electrode than non-smokers in the NoGo condition (*M* difference = 3.04, *SE* = 1.08, $p = .018$). (…) The three-way interaction was caused by CUD patients who had lower P3 amplitudes on the FCz electrode than non-smokers in the NoGo condition (*M* difference = 2.75, *SE* = 1.08, **$p = .037$**). *REQUESTED: df and t belonging to reported p.*	$t(106) = 2.48$ $p = 0.01471$ Note: *p* value in paper is Bonferroni-corrected (*3)	$t(106) = 2.54$ $p = 0.01254$ Note: *p* value in paper is Bonferroni-corrected (*3)

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.3 (continued)

P-curve Disclosure Table No-Go P3

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Morie et al. (2014)	We hypothesized that **current cocaine abusers would demonstrate** reduction in task accuracy, reduced post error slowing, and **attenuation of ERP components related to inhibition**, performance monitoring and error awareness.	For the stimulus-locked N2 and P3, a 2 × 2 ANOVA was run for each with factors of Group (cocaine user versus controls) and Response Type (correct Go-response versus correct No-Go response).	Difference of means (SUDs vs. HC)	The ANOVA for the P3 revealed a main effect of response type ($F49 = 8.4$, $p < .01$) and a main effect of group ($F49 = 4.1$, $p < .05$), with an interaction of response type and group ($F49 = 5.0$, $p < .03$) . Pairwise comparisons revealed the P3 amplitude to be smaller in the addicts than in the controls (**$t49 = 6.2$, $p < .01$**).	$t(49) = 6.2$ $p = 1.15007e-07$	N/A
Sokhadze et al. (2008)	**Patient group differences in inhibition efficiency were predicted to be reflected in a lower amplitude of the fronto-central ERP indices of the inhibition** (so-called anterior NoGo-N2 and **NoGo-P3**).	Data for each EEG-based dependent variable are analysed using repeated measures ANOVA (SPSS 14.0). Within-subject factors are stimulus (Go-target, NoGo), congruence (congruent, incongruent), and error (commission, omission) type for the ERN. The between-subject factors is group (SUD, controls).	Difference of means (SUDs vs. HC)	Anterior frontal NoGo-P3 difference wave. This frontal difference wave showed between group difference being higher in controls compared to the SUD group (1.38 ± 1.30 vs. 0.41 ± 1.10 μV, **$F = 4.36$, $p = .047$**). *REQUESTED: df belonging to reported F.*	$F(1, 32) = 4.36$ $p = 0.04484$	N/A

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.4

P-curve Disclosure Table P3a

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Holguin et al. (1999)	The present study used a visual three-stimulus paradigm to assess the **P3a component in a population of abstinent, chronic alcoholics and a control group**.	ANCOVAs with group as a between-subjects factor and electrode as a within-subject factor using age as a covariate were performed to assess group differences in P3a amplitudes and latencies in each of the regions. (…) ANCOVAs were also used to analyse differences in the ERP measurements for individual electrodes between the two groups.	Difference of means (SUDs vs. HC)	The mixed-model ANCOVA analyses of the P3a characteristics for the infrequent nontarget condition demonstrated significant differences between the two groups in the central [$\boldsymbol{F(1, 69) = 6.81, p < .0111}$], parietal [$F(1, 69) = 8.21, p < .0055$], temporal [$F(1, 69) = 8.21, p < .0055$], and occipital [$F(1, 69) = 5.08, p < .0274$] regions, but not in the frontal region [$F(1, 69) = 3.58, p > .0626$]. ANCOVA analyses of each individual lead yielded significant reductions in the P3a amplitude in the alcoholic group for six central, five parietal, six temporal, and two occipital leads (see Table 3).	$F(1, 69) = 6.81$ $p = 0.01111$	$F(1, 69) = 5.73$ $p = 0.01940$
X. Liu et al. (2020)	The hypothesis of this study is that **cognitive control deficits in AD are** both **a trait-** and state-dependent **biomarker**, which is **reflected by P3a**/3b.	Using P3a and P3b as dependent variables, a 2 × 2 repeated-measures ANOVA on mean amplitudes and the mean latencies with group (AD group vs. HC group) as a between-subjects factor and time point (time 1 vs. time 2) as a within-subjects factor, was performed.	Difference of means (SUDs vs. HC)	The simple effect for group at time 1 level and time 2 level was significant ($\boldsymbol{F1, 58 =}$ 107.482, **18.006**, all $p <$ 0.001).	$F(1, 58) = 18.006$ $p = 0.00008$	N/A

Table 3. Mean P3a Amplitudes (μV) for the Infrequent Non-Target Stimulus in the Control and Alcoholic Groups

	Controls (*n* = 28)		Alcoholics (*n* = 44)			
	Mean	SD	Mean	SD	F	p‡
Fp1	6.61	5.16	6.16	4.21	1.55	ns
Fp2	6.63	4.79	6.33	4.40	0.93	ns
AF1	10.10	5.61	9.46	5.06	1.77	ns
AF2	10.75	6.30	9.46	4.94	2.75	ns
Fz	14.06	6.05	12.06	6.14	3.94	ns
F3	12.44	5.84	10.39	5.55	6.37	.0139*
F4	13.22	5.78	11.04	5.66	4.99	.0287*
F7	7.24	3.94	6.12	4.13	4.73	.0330*
F8	8.48	4.21	7.01	3.64	2.09	ns
FC1	16.91	6.83	13.70	6.60	7.35	.0085
FC2	17.45	7.10	13.95	6.51	5.66	.0201
FC5	11.36	5.08	9.43	5.21	4.96	.0292*
FC6	13.12	5.35	10.04	4.95	5.77	.0190
Cz	19.05	7.26	15.08	7.31	6.75	.0115
C3	16.72	6.18	12.72	6.61	8.11	.0058
C4	17.66	6.05	13.78	6.40	5.73	.0194
CP1	18.39	6.27	13.94	6.99	8.15	.0057
CP2	18.99	6.62	14.25	7.08	8.06	.0060
Pz	17.54	5.78	13.27	6.86	6.11	.0159
P3	15.88	6.04	11.25	6.42	7.56	.0076
P4	16.03	5.39	11.35	5.83	8.74	.0043
T7	9.00	4.62	6.73	4.54	4.80	.0318
T8	10.34	4.47	7.28	4.01	5.28	.0245
CP5	13.76	4.72	9.91	5.84	6.44	.0134
CP6	14.80	4.99	10.81	5.53	6.02	.0166
P7	10.43	5.07	6.62	5.24	6.72	.0116
P8	11.70	4.32	6.84	4.76	11.97	.0009
PO1	13.66	5.91	10.08	6.12	4.85	.0310
PO2	14.20	5.69	9.97	6.13	6.35	.0141
O1	9.58	5.50	6.47	4.91	3.67	ns
O2	9.36	5.24	6.18	4.84	3.74	ns

‡ df (1, 69); * no significant Bonferroni comparison; ns, not significant.

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.5

P-curve Disclosure Table P3b

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Boutros et al. (2000)	We compared the effects of **healthy aging to** those of **chronic heavy alcohol use** on cognitive evoked responses (**P300**).	The amplitudes and latencies of the P300, derived from the oddball paradigm, and the N100/P200 complex and the P50 components, derived from the paired-click paradigm, were examined in an analysis of variance procedure with groups as the between-subjects factor (alcoholic, young controls, aged controls).	Difference of means (SUDs vs. HC 1; SUDs vs. HC 2)	Table 2 shows the amplitudes of the P50 and N100/P200 complex to S1 stimuli of the paired clicks and the S2/S1 sensory gating ratios of both the P50 and N100/P200 components, as well as the P300 amplitudes and latencies derived from the oddball paradigm. The mean amplitude of the P300 was 4.2 ± 2.6 μV for alcohol-dependent subjects, 8.9 ± 3.3 μV for age-matched normal subjects, and 5.0 ± 2.6 μV for older control subjects (F[2, 281] = 3.64, p < .04 for an overall effect of group). **TABLE 2. Amplitudes of P50 and N100 Components to S1 Stimuli of Paired Clicks and S2/S1 Sensory Gating Ratios of Both P50 and N100/P200 Complexes as Well as P300 Amplitudes and Latencies (in Milliseconds)** Alcohol Dependent / Age-Matched Controls / Aged Controls P50 S1 μV: 2.8 ± 1.2 / 3.9 ± 2.2 / 3.0 ± 1.9 S2/S1 x 100: 69 ± 66 / 44 ± 23 / 56 ± 44 N1/P2 S1 μV: 12.1 ± 6.6 / 12.6 ± 5.3 / 8.7 ± 5.3 S2/S1 x 100: 52 ± 26 / 46 ± 10 / 81 ± 65 P300 μV*: 4.2 ± 2.6[a] / 8.9 ± 3.2[ab] / 5.0 ± 2.6[b] P300 Lat*: 391 ± 35 / 360 ± 28 / 408 ± 42 [a]t = 3.028, p < .007 for alcohol-dependent vs. age-matched comparison. [b]t = 2.1, p < .05 for age-matched vs. aged controls comparison. *An overall significant effect of variable.	$t(18) = 3.028$ $p = 0.00723$	n.s.
H. L. Cohen et al. (2002)	This study examines the **ERP differences between sober alcoholics and normal controls**, based on data recorded from 61 scalp electrode sites t**o both auditory and visual stimuli**.	MANCOVAs (SAS v6.11, Proc GLM, SAS, Cary, NC) (group + modality + age + modality x group + group x age) were used to perform between group comparisons for P3, N2, and P2 amplitudes and latencies in each of the five brain regions.	Difference of means (SUDs vs. HC)	P3 amplitudes were significantly different between alcoholics and controls at central [$F(14, 120) = 2.99$, $p < 0.001$]; parietal [$F(10, 124) = 3.17$, $p < 0.01$]; occipital [$F(8, 126) = 2.75$, $p < 0.01$]; and temporal [$F(12, 122) = 3.12$, $p < 0.001$] regions, but none of these effects was modality specific (group x modality interactions were not significant). In both modalities, P3 amplitudes were lower in alcoholics than in controls.	$F(14, 120) = 2.99$ $p = 0.00059$	$F(10, 124) = 3.17$ $p = 0.00119$
Fein & Andrew (2011)	In the current study, we examine the **visual P3b**, N2b and N2a components of the event-related potential (ERP) **response to target** and rare non-target **stimuli in a treatment-naïve alcohol dependent (TNAD) sample compared to age and gender comparable non-alcoholic controls (NAC)**.	Two-way group × gender analysis of variance (ANOVA) was conducted separately for the target and rare non-target conditions.	Difference of means (SUDs vs. HC)	For the target condition, P3b amplitude was lower in TNAD (**$F(1, 135) = 3.914$, $p = 0.050$**).	$F(1, 135) = 3.914$ $p = 0.04992$	N/A
Glenn et al. (1996)	Our first aim in this experiment, then, was to cross-validate previous findings that **sober alcoholics would manifest smaller visually evoked P300 amplitudes at the Pz electrode than peer controls**.	n.r.	Difference of means (SUDs vs. HC)	Univariate analyses indicated that alcoholics had significantly lower P300 amplitudes than controls, **$F(1, 73) = 5.00$, $p = 0.028$**.	$F(1, 73) = 5.00$ $p = 0.02840$	N/A

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.5 (continued)

P-curve Disclosure Table P3b

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Gooding et al. (2008)	Based on prior findings, we expected that the **cocaine-dependent patients would display reduced P300 amplitude**.	Independent-samples *T*-tests were performed to compare the two participant groups in terms of their P300 amplitude and latencies.	Difference of means (SUDs vs. HC)	Analyses revealed a significant group difference between cocaine-dependent subjects and healthy controls in P300 amplitude, with the cocaine-dependent subjects having reduced amplitude compared to the controls, ***t*(27) = 4.04, *P* < 0.001**.	$t(27) = 4.04$ $p = 0.00040$	N/A
Guney et al. (2009)	The aim of the study is to identify the chronic **effects of tobacco smoking on the P300**.	Independent sample *t*-test and Pearson's correlation coefficients were used.	Difference of means (SUDs vs. HC)	The P300 amplitude at Cz was lower in the smoking participants than in the control group (***t* = -3.202; *df* = 62; *p* = 0.002**).	$t(62) = -3.202$ $p = 0.00215$	N/A
K. A. Jones et al. (2006)	In the first stage of the analyses we are interested in **determining which** frequency band and/or **ERP components may act as useful predictors of the alcoholic status**.	The second and main part of the results section deals with between-group comparisons of the S-transform derived ERO data and ERP amplitudes using target condition trials of the visual oddball paradigm. Initial tests of group differences were applied to three midline electrodes (Fz, Cz and Pz) using multivariate analysis of variance (MANOVA). (…) Group differences were assessed using the Pillai-Bartlett trace to evaluate an F-value and corresponding significance level.	Difference of means (SUDs vs. HC)	Results of the MANOVA analyses are provided in Table 3; in these analyses the log transformed ERO power and ERP amplitude dependent variables (from the Fz, Cz and Pz electrodes) were modelled with group (alcoholic or control) as the independent variable. (…) In concordance with previously published work alcoholic individuals were observed to manifest lower P300 amplitude. Table 3 Results of MANOVA statistical tests of group significance using log transformed ERO_{Theta}, ERO_{Beta} and ERP amplitude data (log transformed P300 amplitude). Model: log(Fz), log(Cz), log(Pz) ~ Group (*N* = 200, *df* = 196) *df* — *Pillai-Bartlett trace* — *F-value* ERO_{Theta} f_1 (3 Hz, 300–700 ms): 1 — **0.017** — $3.34^{(*)}$ f_2 (3–7 Hz, 300–700 ms): 1 — 0.102 (0.074) — $7.2^{(**)}$ ($5.2^{(**)}$) f_3 (4–5 Hz, 200–500 ms): 1 — 0.133 (0.099) — $9.2^{(**)}$ ($7.2^{(**)}$) f_4 (6–7 Hz, 200–400 ms): 1 — 0.069 (0.047) — $4.8^{(*)}$ ($3.8^{(**)}$) α (8–12 Hz, 100–300 ms): 1 — 0.027 (0.037) — 1.8 (2.5) ERO_{Beta} f_1 (3 Hz, 300–700 ms): 1 — 0.035 — $2.5^{(**)}$ f_2 (3–7 Hz, 300–700 ms): 1 — 0.07 (0.047) — $5.4^{(**)}$ ($4.6^{(**)}$) β_1 (4–5 Hz, 200–500 ms): 1 — **0.146 (0.13)** — $11.8^{(**)}$ ($9.4^{(**)}$) β_2 (6–7 Hz, 200–400 ms): 1 — 0.063 (0.1) — $4.4^{(*)}$ ($7.36^{(**)}$) α (8–12 Hz, 100–300 ms): 1 — 0.002 (0.05) — 1.0 (1.9) *ERP amplitude* P300: **1** — 0.127 — **$9.8^{(**)}$** Model: Fz, Cz, Pz ~ Group (*N* = 200, *df* = 198) *ERP amplitude* N100: 1 — 0.021 — 1.6 P200: 1 — 0.047 — 3.8 N200: 1 — 0.027 — 1.8	$F(1, 196) = 9.6$ $p = 0.00223$	N/A
Kuntz et al. (2021)	Latency increase **and amplitude reduction of the P3b** are consistently reported in patients with chronic alcohol abuse. (…) The aim of the present study was to detect **differences in ERPs between ADP and controls** at inclusion, as well as a significant improvement after BZ treatment.	Infrequent tones, group (controls and patients), task condition (control and counting backward), session recording (D0 and D15), age, medication (benzodiazepine equivalent units), and a group × task interaction were included as independent variables in a repeated measure linear regression model to analyse their respective influence on each of the dependent variables (EEG measures).	Difference of means (SUDs vs. HC)	For the P3b component, the amplitude was significantly lower in ADP as compared to controls (***z* = -3.34, *p* = 0.001**).	$z = -3.34$ $p = 0.00084$	N/A
X. Liu et al. (2020)	The hypothesis of this study is that **cognitive control deficits in AD are** both **a trait-** and state-dependent **biomarker**, which is **reflected by** P3a/**3b**.	Using P3a and P3b as dependent variables, a 2 × 2 repeated-measures ANOVA on mean amplitudes and the mean latencies with group (AD group vs. HC group) as a between-subjects factor and time point (time 1 vs. time 2) as a within-subjects factor, was performed.	Difference of means (SUDs vs. HC)	The simple effect for group at time 1 level and time 2 level was significant (***F*1, 58 =** 56.161, **8.817**, $p <$ 0.001).	$F(1, 58) = 8.817$ $p = 0.00433$	N/A

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.5 (continued)

P-curve Disclosure Table P3b

<table>
<tr><th>Author (year)</th><th>Prediction of interest (quoted)</th><th>Study design (quoted)</th><th>Key statistical result</th><th>Statistical results (quoted)</th><th>Results</th><th>Robustness results</th></tr>
<tr>
<td>Moeller et al. (2004)</td>
<td>We further hypothesised that cocaine-dependent subjects would have reduced auditory P300 amplitude relative to controls.</td>
<td>Group comparisons on event-related potential measures were performed using analysis of covariance (ANCOVA), with age, gender, and education included as covariates.</td>
<td>Difference of means (SUDs vs. HC)</td>
<td>Amplitudes and latencies are summarized in table 2. ANCOVA showed a significant group difference between cocaine-dependent subjects and controls in P300 amplitude, with cocaine-dependent subjects having reduced P300 amplitudes compared to controls.

Table 2. Means ± standard deviations for P300 amplitude and latency at electrode Pz in cocaine-dependent subjects and controls
<table>
<tr><th></th><th>Cocaine dependent</th><th>Control</th><th>F</th><th>p</th></tr>
<tr><td>Amplitude</td><td>7.7 ± 3.2</td><td>13.1 ± 5.4</td><td>11.6</td><td>0.002</td></tr>
<tr><td>Latency</td><td>351.6 ± 50.4</td><td>357.4 ± 32.2</td><td>0.5</td><td>0.496</td></tr>
</table>
ANCOVA with age, gender, and education as covariates.</td>
<td>F(1, 26) = 11.6
p = 0.00215</td>
<td>N/A</td>
</tr>
<tr>
<td>Motlagh et al. (2017)</td>
<td>Investigating quantitative EEG (qEEG), MMN, P300, and P600 properties can lead to a better understanding of brain neurobiological feature alteration among heroin addicts.</td>
<td>Group comparisons for PSD and ERP features were performed using repeated-measure ANOVA, while the group of subjects was chosen as a between-subject factor.</td>
<td>Difference of means (SUDs vs. HC)</td>
<td>The post hoc analysis confirmed diminished P300 amplitude among the addicts during digit-span (P<0.0001) and auditory oddball (P = 0.0007) tasks, while it was higher for cue-reactivity condition (P = 0.009) compared with the healthy controls. (…) Table III shows the mean and standard deviations of electrophysiological measurements of the two groups.

TABLE III. Electrophysiological Measurements (Mean and SD) of Both Groups
<table>
<tr><th></th><th></th><th></th><th>Heroin addicts (n = 19)</th><th></th><th>Healthy controls (n = 19)</th><th></th><th></th></tr>
<tr><th>Variable</th><th></th><th></th><th>Mean</th><th>SD</th><th>Mean</th><th>SD</th><th>F statistics $F_{(1,36)}$</th></tr>
<tr><td>Delta</td><td>Frontal</td><td></td><td>4.0</td><td>2.1</td><td>9.1</td><td>2.2</td><td>51.68**</td></tr>
<tr><td></td><td>Central</td><td></td><td>3.3</td><td>2.1</td><td>7.8</td><td>2.3</td><td>40.35**</td></tr>
<tr><td></td><td>Temporal</td><td></td><td>2.9</td><td>1.7</td><td>5.5</td><td>1.1</td><td>29.1*</td></tr>
<tr><td></td><td>Occipital/parietal</td><td></td><td>3.9</td><td>2.6</td><td>8.1</td><td>2.2</td><td>27.5*</td></tr>
<tr><td>Theta</td><td>Frontal</td><td></td><td>1.9</td><td>0.9</td><td>3.4</td><td>0.7</td><td>31.48**</td></tr>
<tr><td></td><td>Central</td><td></td><td>1.8</td><td>0.8</td><td>3.2</td><td>0.8</td><td>26.66*</td></tr>
<tr><td></td><td>Temporal</td><td></td><td>1.3</td><td>0.6</td><td>1.8</td><td>0.3</td><td>8.78*</td></tr>
<tr><td></td><td>Occipital/parietal</td><td></td><td>1.6</td><td>0.8</td><td>2.5</td><td>0.5</td><td>17.57*</td></tr>
<tr><td>Alpha</td><td>Frontal</td><td></td><td>2.6</td><td>1.3</td><td>3.9</td><td>0.9</td><td>11.89*</td></tr>
<tr><td></td><td>Central</td><td></td><td>3.8</td><td>3.2</td><td>6.5</td><td>2.4</td><td>8.23*</td></tr>
<tr><td></td><td>Temporal</td><td></td><td>2.2</td><td>1.2</td><td>2.5</td><td>0.3</td><td>0.77</td></tr>
<tr><td></td><td>Occipital/parietal</td><td></td><td>3.3</td><td>2.6</td><td>4.7</td><td>1.3</td><td>4.35</td></tr>
<tr><td>Beta</td><td>Frontal</td><td></td><td>5.9</td><td>1.4</td><td>2.7</td><td>1.3</td><td>54.31**</td></tr>
<tr><td></td><td>Central</td><td></td><td>7.6</td><td>1.9</td><td>3.1</td><td>1.6</td><td>61.03**</td></tr>
<tr><td></td><td>Temporal</td><td></td><td>4.7</td><td>0.6</td><td>2.6</td><td>0.9</td><td>71.21**</td></tr>
<tr><td></td><td>Occipital/parietal</td><td></td><td>5.8</td><td>1.2</td><td>2.8</td><td>1.4</td><td>48.71**</td></tr>
<tr><td>ERP amplitude</td><td>MMN</td><td>Small</td><td>1.21</td><td>0.50</td><td>2.13</td><td>0.33</td><td>44.13**</td></tr>
<tr><td></td><td></td><td>Large</td><td>3.56</td><td>0.81</td><td>4.99</td><td>0.58</td><td>39.24**</td></tr>
<tr><td></td><td>P300</td><td>Task 1 attention phase</td><td>11.66</td><td>2.20</td><td>17.72</td><td>3.06</td><td>48.97**</td></tr>
<tr><td></td><td></td><td>Task 2 recall phase</td><td>13.89</td><td>1.84</td><td>17.64</td><td>2.93</td><td>22.39**</td></tr>
<tr><td></td><td></td><td>Auditory</td><td>11.98</td><td>2.73</td><td>16.17</td><td>4.09</td><td>13.83**</td></tr>
<tr><td></td><td></td><td>Visual cue reactivity</td><td>20.77</td><td>1.78</td><td>18.75</td><td>2.63</td><td>7.73*</td></tr>
<tr><td></td><td>P600</td><td></td><td>9.51</td><td>1.73</td><td>12.66</td><td>1.15</td><td>43.8**</td></tr>
<tr><td>ERP latency</td><td>MMN</td><td>Small</td><td>0.76</td><td>0.15</td><td>0.27</td><td>0.13</td><td>113**</td></tr>
<tr><td></td><td></td><td>Large</td><td>0.76</td><td>0.15</td><td>0.22</td><td>0.18</td><td>96.5**</td></tr>
<tr><td></td><td>P300</td><td>Task 1 attention phase</td><td>0.75</td><td>0.13</td><td>0.24</td><td>0.14</td><td>127***</td></tr>
<tr><td></td><td></td><td>Task 2 recall phase</td><td>0.77</td><td>0.15</td><td>0.24</td><td>0.15</td><td>155***</td></tr>
<tr><td></td><td></td><td>Auditory</td><td>0.81</td><td>0.12</td><td>0.22</td><td>0.15</td><td>172***</td></tr>
<tr><td></td><td></td><td>Visual cue reactivity</td><td>0.79</td><td>0.13</td><td>0.19</td><td>0.14</td><td>184***</td></tr>
<tr><td></td><td>P600</td><td></td><td>0.79</td><td>0.14</td><td>0.25</td><td>0.15</td><td>133***</td></tr>
</table>
Note: One-way ANOVA with continuous variable was used for calculating the F statistics. * P < 0.01; ** P < 0.001; *** P ≈ 0; SD indicates standard deviation.</td>
<td>F(1, 36) = 13.83
p = 0.00068</td>
<td>N/A</td>
</tr>
<tr>
<td>Olbrich et al. (2000)</td>
<td>ERP alterations and especially decreased P3 amplitudes have been reported both in adult alcoholics and in unaffected children of alcoholics. (…) Using well-established auditory and visual oddball paradigms, we performed topographical ERP recordings in detoxified alcoholics and control subjects.</td>
<td>As a main issue we compared for each component the amplitudes measured at midline recording sites with spatial parameters with respect to possible differences between alcoholics and controls.</td>
<td>Difference of means (SUDs vs. HC)</td>
<td>For P3 measurements at Cz, significantly smaller amplitudes were found in alcoholics for the visual (p = .026) but not for the auditory task.

REQUESTED: df and t belonging to reported p.</td>
<td>t(28.352) = 2.4386
p = 0.02126</td>
<td>N/A</td>
</tr>
</table>

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.5 (continued)

P-curve Disclosure Table P3b

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Park et al. (2016)	We hypothesized that the **P300 amplitudes of patients with IGD in response to target stimuli would be reduced compared with those of HCs**.	The amplitudes and latencies of the P300 component were separately analysed with repeated-measures ANOVAs with electrode sites (FCz, Cz, CPz and Pz) as within-subject factors and group as the between-subject factor.	Difference of means (SUDs vs. HC)	Patients with IGD showed significantly lower P300 amplitudes than HCs at CPz (**$F(1, 47) = 8.02$, $P < 0.01$**) but not at FCz, Cz and Pz.	$F(1, 47) = 8.02$ $p = 0.00679$	N/A
Park et al. (2017)	Based on previous studies of patients with IGD and SUD, it was hypothesised that **patients with IGD would exhibit decreased P300 amplitudes**.	A repeated-measures ANCOVA was performed for the P300 amplitudes and latencies with electrode site as a within-subject factor, group as the between-subject factor, and age, IQ, BDI, and BAI as covariates.	Difference of means (SUDs vs. HC)	Significant main effects of group (**$F[1, 40] = 5.458$, $P = .025$** for P300 amplitudes were found. There were no main effect of electrode site and no interaction with the P300 amplitudes. Post hoc tests revealed that the IGD group showed significantly lower P300 amplitudes than those of the HCs at CPz (**$F[1, 40] = 5.681$, $P = .022$**), but not at Pz.	$F(1, 40) = 5.458$ $p = 0.02458$	$F(1, 40) = 5.681$ $p = 0.02199$
Park et al. (2023)	We hypothesised that **patients with IGD would have abnormalities in** resting-state fMRI and **ERP features in brain regions associated with cognitive functioning and sensory processing**.	In the auditory oddball ERP task, repeated measures analysis of variance was performed for the amplitudes and latencies of the P3 component, with the electrode site as the within-subject factor and group as the between-subject factor.	Difference of means (SUDs vs. HC)	Analysis of P3 amplitudes revealed a significant main effect of group [**$F (1, 43) = 5.242$, $p = 0.027$**, $\eta_p^2 = 0.109$]. No main effect of electrode site was observed, and no interaction with P3 amplitudes was detected. Compared with HCs, individuals with IGD had significantly lower P3 amplitudes at CPz [**$F (1, 43) = 5.823$, $p = 0.020$**, $\eta_p^2 = 0.119$], but not at Pz.	$F(1, 43) = 5.242$ $p = 0.02702$	$F(1, 43) = 5.823$ $p = 0.02015$

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.5 (continued)

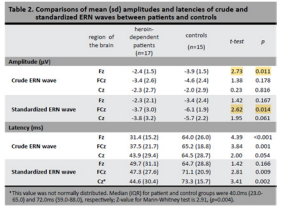
P-curve Disclosure Table P3b

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Parsons et al. (1990)	Accordingly, the present study was undertaken with the following aims: (...) 2. To cross-validate findings of **altered ERP responses in male alcoholics and** to determine whether similar altered responses are present in **female alcoholics**.	A groups [alcoholics (n = 128) versus controls (n = 77)] by FH+ (n = 120) versus FH- (n = 85) by gender [males (n = 93) versus females (n = 112)] ANOVA was performed on each of the ERP measures for the visual and auditory measures on the "attended" target trials at each electrode location.	Difference of means (SUDs vs. HC)	There was a significant main effect for groups at Pz for the visual P3 amplitude [$\boldsymbol{F}$ **(1, 196) = 4.45, p = 0.031** (…). In each instance alcoholics had lower amplitudes. In view of the many interactions involving gender, the male and female samples were separately analysed. In the female sample, applying ANOVAs similar to the above, there were no significant differences between alcoholics and controls or FH+ versus FH- or their interactions on the ERP measures. In contrast, in the male samples there were a number of significant findings (Tables 4 and 5). **Table 4.** Alcoholism and Family History of Alcoholism Effects in Male FH− and FH+ Alcoholics and Controls Variable / Effect / F / p (Pz) Vis P3A* — Alcoholism main effect — 5.88 — 0.017 (Oz) Vis N1A — Alcoholism main effect — 5.05 — 0.026 (Pz) Vis N1A — Alcoholism main effect — 5.00 — 0.027 (Cz) Vis N1A — Alcoholism main effect — 8.14 — 0.005 (Oz) Vis NdA — Alcoholism main effect — 6.39 — 0.010 (Pz) Vis NdA — Alcoholism main effect — 5.64 — 0.019 (Pz) Vis N2L — Alcoholism main effect — 4.59 — 0.030 (Cz) Vis N2L — Alcoholism main effect — 3.72 — 0.056 (Oz) Vis N1A — Family history main effect — 4.54 — 0.035 (Cz) Aud N1A — Family history main effect — 6.67 — 0.011 (Oz) Vis NdA — Family history main effect — 6.52 — 0.012 (Pz) Vis N2L — Family history main effect — 4.33 — 0.040 (Cz) Aud N2L — Family history main effect — 4.75 — 0.030 (Oz) Vis N1A — Alcoholism × family history interaction — 8.68 — 0.004 (Oz) Vis NdA — Alcoholism × family history interaction — 4.67 — 0.033 * Abbreviations: P3, midline parietal electrode placement; Oz, midline occipital placement; Cz, midline central placement; Vis, visual stimulus; Aud, auditory stimulus; "A", amplitude; "L", latency.	$F(1, 196) = 4.45$ $p = 0.03617$	$F(1, 89) = 5.88$ $p = 0.01733$
Polich & Ochoa (2004)	Given the variable findings from alcoholism-risk and smoking history **P300** studies outlined above, both task and chronic factors are likely to affect ERP outcomes. The present study was conducted to compare LR and HR young adult subjects who either had **never smoked tobacco or regularly smoked cigarettes** in order to characterize these effects more directly.	A 4-factor (2 risk groups [LR vs. HR] x 2 smoking status [non-smokers vs. smokers] x 3 anterior-to-posterior positions [frontal vs. central vs. parietal] x 5 coronal electrodes [F7, F3, Fz, F4, F8 vs. T7, C3, Cz, C4, T8 vs. P7, P3, Pz, P4, P8]) repeated measures ANOVA was applied to the P300 data from the target stimuli.	Difference of means (SUDs vs. HC)	Non-smokers yielded significantly larger P300 amplitudes overall than smokers, $\boldsymbol{F(1, 76) = 8.01, P < 0.01}$; η^2 = 0:089. P300 amplitude data from just the midline electrodes (…) was larger for LR compared to HR subjects, $F(1, 76) = 5.38$, $P < 0.05$, $\eta^2 = 0.056$; and non-smokers again produced larger P300 components than smokers, $\boldsymbol{F(1, 76) = 14.06, P < 0.001}$, $\eta^2 = 0.146$.	$F(1, 76) = 8.01$ $p = 0.00595$	$F(1, 76) = 14.06$ $p = 0.00034$
Whipple et al. (1991)	One of the goals of the present study was to **compare electrophysiological functioning in alcoholic fathers** and their sons **to non-alcoholic fathers** and their sons under similar task conditions.	An additional ANOVA directly tested the hypothesis of reduced P3 for the A+ group for the stimulus type (targets) and electrode site (Pz) where P3 is maximal. Since there were no observable P3 differences between the NA+ and NA- groups, the two groups were collapsed for this analysis.	Difference of means (SUDs vs. HC)	P3 amplitude was significantly reduced for A+ fathers, as compared to NA + and NA- fathers, both across all stimuli and recording sites, $F(2, 35) = 12.02$, $p < 0.001$, and for target stimuli at Pz, $\boldsymbol{F(1, 36) = 10.31, p < 0.003}$.	$F(1, 36) = 10.31$ $p = 0.00278$	N/A

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.6

P-curve Disclosure Table ERN

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
H. Chen, Jiang, et al. (2013)	This study uses high-density ERP techniques to compare the characteristics of **ERN** among **males undergoing rehabilitation for heroin dependence and non-dependent, male community members**.	Repeated measures analysis of variance was used to analyse the EEG data.	Difference of means (SUDs vs. HC)	The crude ERN waves at the three electrodes for the patient and control groups are shown in Figure 2 and the univariate comparison of the amplitudes and latencies of these waves is shown in Table 2.  **Table 2. Comparisons of mean (sd) amplitudes and latencies of crude and standardized ERN waves between patients and controls** region of the brain / heroin-dependent patients (n=17) / controls (n=15) / *t-test* / *p* **Amplitude (μV)** Crude ERN wave: Fz -2.4 (1.5) -3.9 (1.5) 2.73 0.011; FCz -3.4 (2.6) -4.6 (2.4) 1.38 0.178; Cz -2.3 (2.7) -2.0 (2.9) 0.23 0.816 Standardized ERN wave: Fz -2.3 (2.1) -3.4 (2.4) 1.42 0.167; FCz -3.7 (3.0) -6.1 (1.9) 2.62 0.014; Cz -3.8 (3.2) -5.7 (2.2) 1.95 0.061 **Latency (ms)** Crude ERN wave: Fz 31.4 (15.2) 64.0 (26.0) 4.39 <0.001; FCz 37.5 (21.7) 65.2 (18.8) 3.84 0.001; Cz 43.9 (29.4) 64.5 (28.7) 2.00 0.054 Standardized ERN wave: Fz 49.7 (31.1) 64.7 (28.8) 1.42 0.166; FCz 47.3 (27.6) 71.1 (20.9) 2.81 0.009; Cz* 44.6 (30.4) 73.3 (15.7) 3.41 0.002 *This value was not normally distributed. Median (IQR) for patient and control groups were 40.0ms (23.0-65.0) and 72.0ms (59.0-86.0), respectively. Z-value for Mann-Whitney test is 2.91, (p=0.004).	$t(30) = 2.73$ $p = 0.01050$	$t(30) = 2.62$ $p = 0.01366$
Franken et al. (2018)	The making of an error can clearly be observed in the EEG: An initial **Error-Related Negativity (ERN)** and a later P3-like Error-positivity (Pe) wave will emerge after committing an error. (…) We aimed to investigate whether **persons who can be considered as having a food addiction have impaired performance monitoring**.	For the ERN and Pe we conducted a 2 × 2 × 2 ANOVA with group (controls, food addiction) as between subjects factor, and response type (incorrect, correct), and region (for the ERN Fz, Cz; for the Pe Cz, Pz) as within subjects factors.	Difference of means (SUDs vs. HC)	There was a significant difference between the two groups for the ERN on errors only on Cz (***p* = 0.04**), with persons in the food addiction group having smaller ERN amplitude (less negative; M = -4.10 μV; SE = 0.97) than controls (M = -7.03 μV; SE = 0.95) on Cz. *REQUESTED: df and t belonging to reported p.*	$t(59) = 2.163866$ $p = 0.03453$	N/A
Littel, Van den Berg, et al. (2012)	We hypothesise **excessive gamers to show** reduced NoGo N2 amplitudes in response to NoGo trials and **reduced ERN** and Pe **amplitudes in response to errors**.	Repeated measures analyses of variance (RM ANOVAs) with Greenhouse-Geisser corrected P-values were applied to analyse the ERP indices of inhibition (N2 and P3), error processing (ERN and Pe) and behavioural performance measures (number of false alarms and RTs). This resulted in a (…) Group x Correctness (correct response, error) x Electrode ANOVAs for ERN and Pe amplitudes.	Difference of means (SUDs vs. HC)	A significant Group x Correctness interaction was found, $F(1, 50) = 10.93$, $P < 0.01$. Post hoc analyses revealed that excessive gamers relative to controls show a significantly reduced ERN in response to errors (***P* < 0.001**), but that both groups do not differ in ERP amplitude elicited by correct responses ($P = 0.08$). *REQUESTED: df and t belonging to reported p.*	$t(50) = 4.225626$ $p = 0.00010$	N/A
Marhe et al. (2013)	A previous study by our lab showed that **cocaine-dependent patients have diminished error processing compared with healthy control subjects**. To confirm these results, we compared the patient sample of the current study with a control group.	We first conducted an analysis of variance to compare the ERN of patients versus controls to confirm that patients in this sample had a reduced ERN.	Difference of means (SUDs vs. HC)	We confirmed that the cocaine-dependent patients in the current sample have a reduced ERN compared with nondependent control subjects (***F1, 70 = 10.03, p* < .01**).	$F(1, 70) = 10.03$ $p = 0.00228$	N/A
Morie et al. (2014)	We hypothesised that **current cocaine abusers would demonstrate** reduction in task accuracy, reduced post error slowing, and **attenuation of ERP components related to** inhibition, performance monitoring and **error awareness**.	For the response-locked ERN and Pe, a 2 × 2 ANOVA was run for each with factors of Group (cocaine user versus controls) and Response Type (correct Go-response versus incorrect No-Go-response).	Difference of means (SUDs vs. HC)	The ANOVA for the ERN revealed a main effect of response type ($F49 = 10.1$, $p < .01$) and a main effect of group ($F49 = 4.7$, $p < .03$), with an interaction of response type and group ($F49 = 4.7$, $p < .03$). Follow up t-tests revealed the ERN amplitude to be less robust in the addicts than in the controls (***t49 = 4.2, p* <.03**).	$t(49) = 4.2$ $p = 0.00011$	N/A

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.6 (continued)

P-curve Disclosure Table ERN

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Padilla et al. (2011)	We tested the hypothesis that **alcoholics engage in enhanced conflict monitoring marked by abnormally high ERN** and CRN **amplitudes** to maintain normal performance levels **relative to controls**.	Univariate ANOVAs were used to investigate error trial RT and ERN amplitudes with diagnosis as a between-group factor and years of education as a covariate.	Difference of means (SUDs vs. HC)	The results of the univariate analysis on the ERN showed a significant effect of diagnosis [$F(1, 24) = 5.83$; $p = 0.024$] such that alcoholics thus also developed larger negative ERN amplitudes following incorrect responses than the controls.	$F(1, 24) = 5.83$ $p = 0.02374$	N/A
Park et al. (2020)	We hypothesised that **IGD patients would demonstrate reduced ERN** and Pe amplitudes **after committing errors**.	Repeated measures ANOVA with IQ as a covariate was conducted to examine error processing in response-locked ERPs, with group as the between-subjects factor, and the lead as the within-subject factor in each analysis.	Difference of means (SUDs vs. HC)	Results for the ERN peak amplitudes revealed significant main effect of group ($F(1, 65) = 6.933$, $P = 0.011$, $\eta_p^2 = 0.096$). (…) Compared with HCs (M = -8.372 µV, SD=0.49), patients with IGD showed significantly lower ERN amplitudes (M = -6.480 µV, SD=0.49).	$F(1, 65) = 6.993$ $p = 0.01056$	N/A
Sokhadze et al. (2008)	**Patients with cocaine abuse were predicted to have** an attenuated amplitude and a prolonged latency of the anterior N200 and N450 ERP components in an incongruent flanker condition indicating a low reactivity to potential response conflict and **a smaller ERN on error trials compared to controls**.	Data for each EEG-based dependent variable are analysed using repeated measures ANOVA (SPSS 14.0). Within-subject factors are stimulus (Go-target, NoGo), congruence (congruent, incongruent), and error (commission, omission) type for the ERN. The between-subject factors is group (SUD, controls).	Difference of means (SUDs vs. HC)	The amplitude of ERN during commission errors was significantly more negative in controls compared to addicts (-5.71 ± 2.76 vs. -2.20 ± 1.52 µV, $F = 7.42$, $p = .021$), which shows higher effectiveness of error detection and error monitoring in control participants. *REQUESTED: df belonging to reported F.*	$F(1, 11) = 7.42$ $p = 0.01979$	N/A

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.7

P-curve Disclosure Table Pe

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Franken, Van Strien, & Kuijpers (2010)	Because we observed reduced ERN and **Pe** in cocaine dependent patients and it is suggested that error processing is reduced in externalising psychopathology, we expected to find **reduced error processing in smokers as compared to controls**.	For the ERN and Pe we conducted a 2 × 2 × 3 ANOVA with group (smokers and non-smokers) as between subjects factor, and response type (incorrect and correct), and region (Fz, Cz, and Pz) as within subjects factor. Follow up *t*-tests on the incorrect minus correct ERP difference scores were employed if appropriate.	Difference of means (SUDs vs. HC)	Post hoc analysis on the mean Pe difference wave showed that smokers had reduced Pe waves as compared to non-smokers, ***t*(44) = 2.09, *p* < .05**.	$t(44) = 2.09$ $p = 0.04243$	N/A
Morie et al. (2014)	We hypothesised that **current cocaine abusers would demonstrate** reduction in task accuracy, reduced post error slowing, and **attenuation of ERP components related to** inhibition, performance monitoring and **error awareness**.	For the response-locked ERN and Pe, a 2 × 2 ANOVA was run for each with factors of Group (cocaine user versus controls) and Response Type (correct Go-response versus incorrect No-Go-response).	Difference of means (SUDs vs. HC)	The ANOVA for the Pe revealed a main effect of response type ($F49 = 9.8$, $p < .01$) and a main effect of group ($F49 = 5.4$, $p < .02$), with an interaction of response type and group ($F49 = 4.3$, $p < .04$). Follow up t-tests once again revealed the Pe amplitude to be smaller in the addicts than in the controls (***t*49 = 4.1, *p* < .04**).	$t(49) = 4.1$ $p = 0.00016$	N/A
Park et al. (2020)	We hypothesised that **IGD patients would demonstrate reduced** ERN and **Pe** amplitudes **after committing errors**.	Repeated measures ANOVA with IQ as a covariate was conducted to examine error processing in response-locked ERPs, with group as the between-subjects factor, and the lead as the within-subject factor in each analysis.	Difference of means (SUDs vs. HC)	The Pe peak amplitudes exhibited a significant main effect of group (***F*(1, 65) = 26.546, *P* < 0.001**, $\eta_p^2 = 0.290$), with Pe amplitudes being lower for IGD patients.	$F(1, 65) = 26.546$ $p = 2.59641E-06$	N/A

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.8

P-curve Disclosure Table Cue-P3 to Drug Cues

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Cheng et al. (2016)	We hypothesised that the **P300** and/or SPW **amplitude of smoking-related pictures was larger than neutral pictures in young adult smokers and this effect was not present in non-smokers.**	ERP data were evaluated by performing repeated measurement analyses of variance (ANOVA) with 'Group', 'Cue-type' and 'Electrode site'. This results in a 2 (Group) × 2 (Cue-type) × 3 (Electrode site) repeated measures ANOVA.	2-way interaction (attenuated)	A 'Group' × 'Cue-type' interaction effect emerged [F **= 6.69, p = 0.01**].	$F(1, 34) = 6.69$ $p = 0.01415$	N/A
Heinze et al. (2007)	Compared to healthy controls higher visual drug stimulus-induced event-related potentials (ERPs) were found in heavy drinkers and addicts using EEG (…). The aim of this study was the assessment of the non-conscious and conscious emotional **responses to auditory alcohol-associated stimuli represented by the modulation of the EEG** (…) **in detoxified alcohol addicts and healthy controls**.	A 2 x 3 x 3 x 2 repeated measures analysis of variance (ANOVA) using as between-subject factor group (alcoholics and healthy controls) and three within-subject factors sagittal [electrodes at the left hemisphere (F3, C3 and P3), the midline (Fz, Cz and Pz) and the right hemisphere (F4, C4 and P4)] and coronal (electrodes at frontal (F3, Fz and F4), central (C3, Cz and C4) and parietal sites (P3, Pz, P4) as well as stimulus category (alcohol-related and neutral).	Difference of means (SUDs vs. HC)	Alcoholics showed at midline electrodes in frontal (Fz: $F(1, 18) = 14.34$, $p < .01$), central (Cz: **$F(1, 18) = 56.26$, $p < .001$**) and parietal sites (Pz: **$F(1, 18) = 13.67$, $p < .01$**) significantly larger P3b amplitudes than healthy controls responding to alcohol-related auditory material.	$F(1, 18) = 56.26$ $p = 6.06033e-07$	$F(1, 18) = 13.67$ $p = 0.00165$
C. Li et al. (2024)	The current study hypothesises that: (…) (3) **Compared to HC individuals, IGD individuals would exhibit enhanced P300 amplitudes when approaching game cues**, reflecting amplified motivational significance for compatible conditions.	2 (group) × 2 (response type) × 2 (stimuli type) mixed ANCOVAs were then respectively conducted for each of the components including amplitude of N100 and P300, with the BDI, BAI, and BIS scores included as covariates.	Difference of means (SUDs vs. HC)	Specially, the between-group simple effect analysis showed that the P300 induced by approaching game cues was significantly larger in the IGD group (7.03 ± 3.17) than in the HC group (2.72 ± 1.84) (**$F(1,40) = 25.73$, $p < 0.001$**).	$F(1, 40) = 25.73$ $p = 0.00001$	N/A

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.8 (continued)

P-curve Disclosure Table Cue-P3 to Drug Cues

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Littel & Franken (2007)	It is hypothesised that **ex-smokers, in response to smoking-related pictures, show less enhanced P300** and SPW **amplitudes than smokers**, whereas the P300 and SPW amplitudes of ex-smokers and never-smokers have approximately the same magnitude.	ERP effects were assessed by performing repeated-measurement analyses of variance (ANOVA) on the four midline electrode sites (Oz, Pz, Cz and Fz). Group (smokers, ex-smokers, and never-smokers) served as the between-subjects factor, and cue type (neutral versus smoking-related) and midline site (Oz, Pz, Cz and Fz) served as within-subjects factors. This resulted in a 4 (midline site) x 2 (cue) x 3 (group) repeated measures ANOVA.	Difference of means (SUDs vs. HC)	For the P300 peak, a G x C interaction effect was found, $F(3,171) = 3.83$, $p < 0.05$. Post-hoc comparisons revealed no significant differences between smokers, ex-smokers and never-smokers on neutral cues. On the smoking-related cues P300 amplitude was significantly larger for smokers than for never-smokers ($p < 0.005$) or ex-smokers (**$p < 0.05$**), whilst no significant differences were found between ex-smokers and never-smokers ($p = 1$). In response to smoking-related pictures several differences were found. At Fz, P300 amplitude was more enhanced for smokers than for never-smokers ($p < 0.05$) and ex-smokers ($p < 0.05$). At this site never-smokers' P300 amplitude did not differ from ex-smokers' P300 amplitude ($p = 1$). Almost the same differences were found at Cz: Smokers showed a more enhanced P300 amplitude than never-smokers ($p < 0.005$), ex-smokers showed a less enhanced P300 amplitude than smokers (**$p < 0.05$**), but ex-smokers' P300 amplitude did not differ from the never-smokers' amplitude ($p = 1$). At Pz, P300 amplitude differed between smokers and never-smokers ($p < 0.005$). No effects were found at Oz. *REQUESTED: df and t belonging to reported p.*	$t(57) = 2.5125938$ $p = 0.01484$ Note: *p* value in paper is Bonferroni-corrected (*3)	$t(57) = 2.9311073$ $p = 0.00485$ Note: *p* value in paper is Bonferroni-corrected (*3)
Littel & Franken (2012)	**In smokers**, a centro-frontally distributed **enlargement of P3** and LPP **amplitudes has been found in response to smoking cues relative to matched neutral cues, whereas non-smokers show no difference in P3** and LPP **amplitudes to both stimulus categories.**	Group (smokers versus non-smokers) served as the between-subjects factor. CS1 stimulus type (neutral versus smoking-related), block (first block, second block), and electrode site served as within-subjects factors in the ANOVA on CS1.	2-way interaction (attenuated)	A significant CS1 × Group effect, **$F(1, 59) = 6.57$, $p = 0.013$**, was found. Post-hoc comparisons revealed that smokers respond with significantly larger P3 amplitudes to $CS1_{smoke}$ than to $CS1_{neutral}$ ($p = 0.001$), whereas non-smokers show no amplitude difference between the two CS1 ($p = 0.239$).	$F(1, 59) = 6.57$ $p = 0.01294$	N/A
Lubman et al. (2007)	It was hypothesised that **heroin-related images** would **elicit larger P300s (compared with neutral stimuli) in opiate-dependent individuals (compared with non-opiate dependent individuals)**, despite attention being actively directed towards task-relevant oddball stimuli.	Statistical analyses were conducted on P300 amplitudes derived from nine central sites (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4) with a 3-stimulus category (neutral, opiate, oddball) by 3 anterior-posterior (frontal, central, parietal) by 3 laterality (right, central, left) by 2 group (MM, control) mixed model repeated measures analysis of variance.	2-way interaction (attenuated)	To test the hypotheses in this study, specific contrasts within the stimulus category effect, and their interaction with group, were examined. (…) The contrast between neutral and opiate stimuli was also significant [$F(1, 23) = 16.26$, $p < 0.001$] and was significantly modified by group [**$F(1, 23) = 4.39$, $p < 0.05$**].	$F(1, 23) = 4.39$ $p = 0.04736$	N/A

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.8 (continued)

P-curve Disclosure Table Cue-P3 to Drug Cues

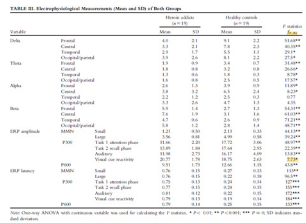
Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Lubman et al. (2008)	It was hypothesised that **heroin-related images** would **elicit larger P300s (compared with** both **neutral** and affective **stimuli) in opiate addicts when compared with drug treatment workers**.	A repeated measures analysis of variance (ANOVA) of the peak P300 amplitudes was performed with Group (addict versus control) as the between-subjects variable and Picture type (neutral, affective or opiate-related) and Electrode (Fz, Cz or Pz) as within-subjects variables.	2-way interaction (attenuated)	There was a main effect of Picture Type ($F[2,52] = 8.90$, $P < 0.001$) and a significant Group x Picture Type interaction ($F[2,52] = 12.4$, $P < 0.001$). Contrasts revealed that the groups significantly differed on the magnitude of the P300 elicited by neutral versus affective pictures ($F[1,26] = 6.56$, $P < 0.05$), neutral versus opiate pictures (**$F[1,26] = 6.37$, $P < 0.05$**) and affective versus opiate pictures ($F[1,26] = 23.1$, $P < 0.001$).	$F(1, 26) = 6.37$ $p = 0.01805$	N/A
Motlagh et al. (2017)	Investigating quantitative EEG (qEEG), MMN, **P300**, and P600 properties can lead to a better understanding of brain neurobiological feature **alteration among heroin addicts**.	Group comparisons for PSD and ERP features were performed using repeated-measure ANOVA, while the group of subjects was chosen as a between-subject factor.	Difference of means (SUDs vs. HC)	The post-hoc analysis confirmed diminished P300 amplitude among the addicts during digit-span ($P<0.0001$) and auditory oddball ($P = 0.0007$) tasks, while it was higher for cue-reactivity condition (**$P = 0.009$**) compared with the healthy controls. (…) Table III shows the mean and standard deviations of electrophysiological measurements of the two groups. 	$F(1, 36) = 7.73$ $p = 0.00858$	N/A
Thalemann et al. (2007)	An incentive salience of computer game-associated cues – **only in excessive computer game players** – should be indicated by a **cue-induced, conditioned emotional-motivational state** as it has been described in cue-reactivity studies **in alcohol- or drug-dependent subjects** confronted with alcohol- or drug-specific cues in a cue-reactivity paradigm.	ERPs of the LPC (…), expressed as a mean value of amplitudes in a special latency range between 350 and 750 ms after stimulus onset, were analysed in a 2 x 3 x 3 x 5 repeated measures ANOVA involving the between-subjects factor group (casual players, excessive players) and three within-subject factors: sagittal (left hemisphere F3, C3, P3; midline Fz, Cz, Pz; right hemisphere F4, C4, P4), coronal (frontal F3, Fz, F4; central C3, Cz, C4; parietal P3, Pz, P4), and stimulus category (game-related cues; alcohol-related cues; and positive, negative, and neutral visual cues).	Difference of means (SUDs vs. HC)	Post hoc between-groups analyses of ERPs representing game cue-induced cortical activity revealed that excessive players showed significantly more extended late positivity at electrode Pz, **$F(1, 28) = 7.84$, $p = .009$**, and trendwise at P4, $F(1, 28) = 3.86$, $p = .059$, whereas no significant group differences were found in the analysis of all other (computer game-irrelevant) categories at the parietal electrodes.	$F(1, 28) = 7.84$ $p = 0.00916$	N/A

TABLE III. Electrophysiological Measurements (Mean and SD) of Both Groups

Variable			Heroin addicts ($n = 19$) Mean	SD	Healthy controls ($n = 19$) Mean	SD	F statistics $f_{(1,36)}$
Delta	Frontal		4.0	2.1	9.1	2.2	51.68**
	Central		3.3	2.1	7.8	2.3	40.35**
	Temporal		2.9	1.7	5.5	1.1	29.1*
	Occipital/parietal		3.9	2.6	8.1	2.2	27.5*
Theta	Frontal		1.9	0.9	3.4	0.7	31.48**
	Central		1.8	0.8	3.2	0.8	26.66*
	Temporal		1.3	0.6	1.8	0.3	8.78*
	Occipital/parietal		1.6	0.8	2.5	0.5	17.57*
Alpha	Frontal		2.6	1.3	3.9	0.9	11.89*
	Central		3.8	3.2	6.5	2.4	8.23*
	Temporal		2.2	1.2	2.5	0.3	0.77
	Occipital/parietal		3.3	2.6	4.7	1.3	4.35
Beta	Frontal		5.9	1.4	2.7	1.3	54.31**
	Central		7.6	1.9	3.1	1.6	61.03**
	Temporal		4.7	0.6	2.6	0.9	71.21**
	Occipital/parietal		5.8	1.2	2.8	1.4	48.71**
ERP amplitude	MMN	Small	1.21	0.50	2.13	0.33	44.13**
		Large	3.56	0.81	4.99	0.58	39.24**
	P300	Task 1 attention phase	11.66	2.20	17.72	3.06	48.97**
		Task 2 recall phase	13.89	1.84	17.64	2.93	22.39**
		Auditory	11.98	2.73	16.17	4.09	13.83**
		Visual cue reactivity	20.77	1.78	18.75	2.63	7.73*
	P600		9.51	1.73	12.66	1.15	43.8**
ERP latency	MMN	Small	0.76	0.15	0.27	0.13	113**
		Large	0.76	0.15	0.22	0.18	96.5**
	P300	Task 1 attention phase	0.75	0.13	0.24	0.14	127***
		Task 2 recall phase	0.77	0.15	0.24	0.15	155***
		Auditory	0.81	0.12	0.22	0.15	172***
		Visual cue reactivity	0.79	0.13	0.19	0.14	184***
	P600		0.79	0.14	0.25	0.15	133***

Note: One-way ANOVA with continuous variable was used for calculating the F statistics. * $P < 0.01$; ** $P < 0.001$; *** $P \approx 0$; SD indicates standard deviation.

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.9

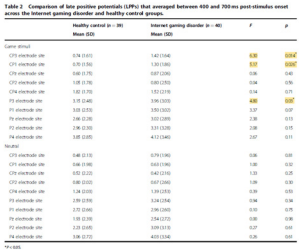
P-curve Disclosure Table LPP to Drug Cues

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Cheng et al. (2016)	We hypothesised that the P300 and/or **SPW amplitude of smoking-related pictures was larger than neutral pictures in young adult smokers and this effect was not present in non-smokers**.	We collapsed anterior (Fz), central (Cz) and posterior electrodes (Pz) to reduce the number of analyses and enhance the interpretation of the results. ERP data were evaluated by performing repeated measurement analyses of variance (ANOVA) with 'Group', 'Cue-type' and 'Electrode site'. This results in a 2 (Group) × 2 (Cue-type) × 3 (Electrode site) repeated measures ANOVA.	2-way interaction (attenuated)	A 'Group' × 'Cue-type' interaction was observed [$\boldsymbol{F = 26.02, p < 0.001}$].	$F(1, 34) = 26.02$ $p = 0.00001$	N/A
Franken et al. (2003)	One of the hypotheses of the present study is that **heroin dependent patients have larger slow positive voltage change on heroin pictures compared to neutral pictures**. If this represents an attentional bias of heroin dependent patients, **this effect should not be present in healthy control subjects**.	Repeated measurement ANOVA with group as between-subjects factor and picture category and electrode site and time-domain as within-subjects factor resulted in a 3 (Time-domain) x 11 (Electrode site) x 2 (Cue type) x 2 (Group) repeated measurement ANOVA, which was used to test for between- and within-group differences and interactions.	2-way interaction (attenuated)	Concerning the main hypothesis of this study, a Cue type x Group interaction effect was found [$\boldsymbol{F(1, 30) = 8.87, P = 0.006}$], indicating that overall positivity in the heroin group, but not in the control group, was more pronounced on heroin pictures than on the neutral pictures.	$F(1, 30) = 8.87$ $p = 0.00569$	N/A
Heinze et al. (2007)	Compared to healthy controls higher visual drug stimulus-induced event-related potentials (ERPs) were found in heavy drinkers and addicts using EEG (…). To take into consideration the complexity of different modalities (visual, olfactory, auditory) of alcohol-associated stimuli to which alcoholics are daily exposed, the aim of this study was the assessment of the non-conscious and conscious emotional **responses to auditory alcohol-associated stimuli represented by the modulation of the EEG** (…) **in detoxified alcohol addicts and healthy controls**.	A 2 x 3 x 3 x 2 repeated measures analysis of variance (ANOVA) using as between-subject factor group (alcoholics and healthy controls) and three within-subject factors sagittal [electrodes at the left hemisphere (F3, C3 and P3), the midline (Fz, Cz and Pz) and the right hemisphere (F4, C4 and P4)] and coronal (electrodes at frontal (F3, Fz and F4), central (C3, Cz and C4) and parietal sites (P3, Pz, P4) as well as stimulus category (alcohol-related and neutral).	Difference of means (SUDs vs. HC)	Post hoc analysis showed significant group differences between alcoholics and healthy controls concerning alcohol-stimulus-induced cortical activity at the midline electrodes in frontal (Fz: $F(1, 18) = 11.92$, $p < .01$), central (Cz: $\boldsymbol{F(1, 18) = 21.64, p < .001}$) and parietal sites (Pz: $\boldsymbol{F(1, 18) = 6.11, p < .05}$). Alcoholics showed a higher activation after the alcohol-related auditory stimulation.	$F(1, 18) = 21.64$ $p = 0.00020$	$F(1, 18) = 6.11$ $p = 0.02365$

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.9 (continued)

P-curve Disclosure Table LPP to Drug Cues

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
B. Kim et al. (2021)	The aim of this study was to examine deficits in cue-related reactivity using event-related potential (ERP) EEG in individuals with IGD. (…) We hypothesised that **individuals with IGD would exhibit increased LPP** and show more activation in visual attentional areas **in response to game-related stimuli**.	Group comparisons of the mean LPP amplitudes were performed using repeated-measures ANCOVAs with the ten centro-parietal electrode sites as the within-subject factors and the groups (HC and IGD) as between-subjects factors.	Difference of means (SUDs vs. HC)	Table 2 presents the means (SDs) and the group comparison results for the LPP amplitude at each electrode site. The game-related cues elicited higher LPP amplitudes in the patients with IGD than in the HC participants at the **CP3**, CP1, and **P3** electrode sites. 	$F(1, 77) = 6.30$ $p = 0.01417$	$F(1, 77) = 4.80$ $p = 0.03148$
Minnix et al. (2013)	In **drug users, drug-related cues may (…) receive preferential processing** in ways similar to emotional stimuli. (…) We investigated **cortical responses to** several categories of emotional cues (erotic, romance, pleasant objects, mutilation, sadness, unpleasant objects) as well as **two types of smoking-related cues** (people smoking and cigarette-related objects). (…) To obtain preliminary evidence about the specificity of our potential results to a non-smoking population, we also included an exploratory analysis of a small convenience sample of 40 **never-smokers**.	A mixed models regression approach was the primary analytic strategy for estimating the effects of stimuli content on the amplitude of the LPP. Subject was included as a random effect and race was included as a covariate since the two groups had significantly different racial compositions. This final model included both smokers and never-smokers and evaluated the effects of group, picture category, and the interaction of these factors on the amplitude of the LPP.	Difference of means (SUDs vs. HC)	Smokers evidenced significantly larger LPP amplitudes in response to both pleasant ($p < 0.01$) and cigarette (**$p < 0.002$**) stimuli compared to never-smokers. Smokers evidenced significantly larger LPP amplitudes than non-smokers for erotic stimuli (p's = .002) and cigarette objects (**p's = .02**), though surprisingly not for cigarette people. *REQUESTED: df and t belonging to reported p.*	$t(654) = 3.05$ $p = 0.00238$	$t(1962) = 2.35$ $p = 0.01887$
Parvaz et al. (2017)	**Increased LPP amplitude in response to drug-related compared with neutral cues** has consistently been shown across all substance use disorders.	The ERP time series in response to neutral pictures were subtracted from those in response to pleasant and drug-related pictures. (…) Healthy controls at baseline were compared with individuals with CUD at baseline and separately with individuals with CUD at follow-up on LPP amplitudes using pictures (pleasant v. drug) × group (CUD v. control) 2 × 2 mixed-model analyses of covariance (ANCOVAs), controlling for the effects of depression.	Difference of means (SUDs vs. HC)	t tests revealed that compared with healthy controls at baseline, individuals with CUD at baseline showed significantly higher LPPs in response to drug cues (**$t(35) = 2.92$, $p = 0.006$**) and significantly lower activity in response to pleasant cues ($t(35) = 4.33$, $p < 0.001$). A similar pattern of results emerged when we compared individuals with CUD at follow-up with healthy controls at baseline. (…). Individuals with CUD had higher LPPs in response to drug-related cues (**$t(35) = 2.73$, $p = 0.010$**) and lower LPPs in response to pleasant cues ($t(35) = 2.12$, $p = 0.040$) at follow-up than healthy controls at baseline.	$t(35) = 2.92$ $p = 0.00609$	$t(35) = 2.73$ $p = 0.00985$

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.9 (continued)

P-curve Disclosure Table LPP to Drug Cues

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Wölfling et al. (2008)	The **responses of the addicts to specific drug-associated stimuli resulted in significantly higher cortical activation as measured by ERPs compared with healthy controls**, which points to an increased incentive salience and a higher amount of elicited arousal of drug-associated stimuli. (…) The aim of the present study was thus to investigate the **emotional processing of drug-associated stimuli** and drug-unspecific stimuli **in users of the so-called 'soft dug' cannabis and cannabis-naive persons**.	ERPs of the LPC were submitted to a 2 x 3 x 3 x 5 repeated-measures ANOVA using group (cannabis vs. control group) as between-subject factor and three within-subject factors: sagittal [electrodes at the left hemisphere (F3, C3 and P3), the midline (Fz, Cz and Pz) and the right hemisphere (F4, C4 and P4)] and coronal (electrodes at frontal (F3, Fz and F4), central (C3, Cz and C4) and parietal sites (P3, Pz, P4) as well as stimulus category (cannabis-associated, alcohol-associated, negative, positive and neutral stimuli).	Difference of means (SUDs vs. HC)	Post-hoc analyses for the category of cannabis-associated stimuli were calculated for the midline electrodes and showed significant group differences between cannabis users and healthy controls at central (Cz: ***F*(1,28) = 17.33, *P* < 0.001**) and parietal (Pz: ***F*(1,28) = 24.19, *P* < 0.001**) but only a tendency towards significance at frontal (Fz: *F*(1,28) = 3.29, *P* = 0.08) sites. At all three sites cannabis users responded with higher cortical activation after exposure to cannabis-associated stimuli compared with the controls.	$F(1, 28) = 17.33$ $p = 0.00027$	$F(1, 28) = 24.19$ $p = 0.00003$
Wölfling et al. (2011)	The objective of the study was to investigate emotional processing of **gambling-relevant** and -irrelevant **stimuli in pathological gamblers and non-gambling controls** using an EEG cue-reactivity paradigm. (…) Based on prior findings in the research of motivated attention it can be **hypothesised that gambling stimuli attain motivational relevance, which is reflected in an enhanced late positivity in the ERP**.	The amplitudes of the LPP were submitted to a 2 × 3 × 3 × 4 repeated measures analysis of variance (ANOVA) using group (pathological gamblers and healthy controls) as between- and sagittal [electrodes at the left hemisphere (F3, C3 and P3), the midline (Fz, Cz and Pz) and the right hemisphere (F4, C4 and P4)], coronal [(electrodes at frontal (F3, Fz and F4), central (C3, Cz and C4) and parietal sites (P3, Pz, P4)] and stimulus category (gambling-related positive, negative and neutral) as within-subject factors.	Difference of means (SUDs vs. HC)	Post hoc analyses of the midline electrodes revealed significant group differences regarding the gambling stimulus-induced late positivity between pathological gamblers and healthy control subjects. Compared to healthy controls pathological gamblers responded with significantly increased amplitudes of the LPP. No significant differences were found for any other stimulus category (see Table 3).	$F(1, 28) = 17.23$ $p = 0.00028$	$F(1, 28) = 22.08$ $p = 0.00006$

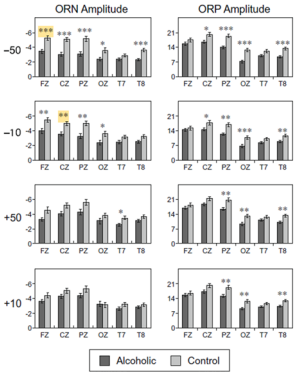
Table 3
Stimulus-induced LPC (450–750 ms, in µV), midline-electrodes (Fz, Cz, Pz) in pathological gamblers (PG) and healthy controls (HC).

	PG, M (SD)	HC, M (SD)	*F*	*df*	*p*
Eelectrode Fz					
LPC, gambling-related stimuli (µV)	4.68 (5.56)	−1.20 (3.81)	11.45	1, 28	.002
LPC, negative stimuli (µV)	5.56 (5.80)	6.77 (8.69)	.20	1, 28	.655
LPC, positive stimuli (µV)	6.60 (6.09)	9.31 (6.54)	1.38	1, 28	.250
LPC, neutral stimuli (µV)	−.05 (3.45)	−.13 (2.68)	.01	1, 28	.949
Electrode Cz					
LPC, gambling-related stimuli (µV)	7.43 (6.50)	−.46 (3.47)	17.23	1, 28	< .001
LPC, negative stimuli (µV)	8.51 (6.50)	8.68 (7.34)	.01	1, 28	.946
LPC, positive stimuli (µV)	8.69 (6.38)	9.61 (5.65)	.17	1, 28	.680
LPC, neutral stimuli (µV)	.50 (3.70)	−.02 (1.69)	.24	1, 28	.626
Electrode Pz					
LPC, gambling-related stimuli (µV)	9.39 (4.62)	2.12 (3.80)	22.08	1, 28	<.001
LPC, negative stimuli (µV)	11.61 (4.14)	14.12 (9.54)	.88	1, 28	.358
LPC, positive stimuli (µV)	10.87 (5.82)	13.48 (7.71)	1.20	1, 28	.283
LPC, neutral stimuli (µV)	2.16 (2.66)	1.50 (2.12)	.57	1, 28	.456

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.10

P-curve Disclosure Table FRN / RewP

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Kamarajan et al. (2010)	The main goal of the present study is to examine reward/outcome processing in alcohol dependent individuals as compared to healthy controls while they subjectively experience monetary loss and gain during the performance of a gambling task. (…) Our hypotheses were the following: (1) **alcoholics will show decreased amplitude in** both **ORN** and ORP **components**.	The model included five factors as fixed effects: valence (loss and gain), amount (50¢ and 10¢), region (frontal, central, parietal, occipital, left-temporal, and right-temporal), electrode (six electrodes) as within-subjects factors, and group (control and alcoholic) as a between-subjects factor.	Difference of means (SUDs vs. HC)	Comparisons of ORN and ORP amplitudes across groups in each outcome condition at different electrode sites are illustrated in Fig. 7 along with Bonferroni adjusted significance levels. Alcoholics showed significantly lower ORP amplitude during all outcome conditions and decreased ORN amplitude during loss conditions (-50 and -10). (…) Bonferroni adjusted significance level is marked with asterisks (*p < 0.05, **p < 0.01, and ***p < 0.001).  *REQUESTED: df and t belonging to reported p.*	$t(78) = 4.073$ $p = 0.00011$ Note: *p* value in paper is Bonferroni-corrected (*6)	$t(78) = 3.914$ $p = 0.00019$ Note: *p* value in paper is Bonferroni-corrected (*6)
Lole et al. (2015)	In order to test the assumptions of several theoretical accounts of problem gambling (…), we examined whether these neural correlates of incentive value processing **indicate reduced sensitivity to reward, as indicated by attenuated FRP amplitudes following wins, and/or to non-reward/punishment, as indicated by reduced FRN amplitudes following Losses, in individuals with this disorder**.	Two separate 2 Group (PG, HC) x 4 Outcome (Large win, Small win, Near-win, Loss) mixed model repeated measures analysis of variance (ANOVA) were performed, with group as a between-subjects factor and outcome as a within-subjects factor. Two orthogonal planned contrasts were conducted on the outcome factor to test: (i) valence differences (Small and Large wins combined versus Losses and Near-wins combined). (…) The Outcome x Group interactions were also examined for each of these planned contrasts.	2-way interaction (attenuated)	Planned contrasts revealed a significant difference in factor scores for the component corresponding to the FRN following Losses and the factor scores corresponding to the FRP following Wins, as indicated by a significant main effect of outcome for this comparison, $F(1, 34) = 93.27$, $p < .001$, $\eta_p^2 = .73$. This effect was more pronounced in the HC group, as indicated by a significant Group x Outcome interaction, $\mathbf{F(1, 34) = 4.89}$, $\mathbf{p = .034}$, $\eta_p^2 = .13$.	$F(1, 34) = 4.89$ $p = 0.03384$	N/A

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.10 (continued)

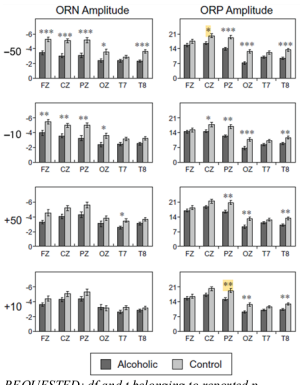
P-curve Disclosure Table FRN / RewP

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Sehrig et al. (2019)	Specific hypotheses were (…) 3. **FRN and feedback P3 are larger for loss than for cash-out feedback, but AUD show less differentiation in response to such feedback**, indicating indifference to negative outcome.	The variation of FRN and feedback P3 with group and outcome condition was examined with a 2 × 2 Group × Outcome ANOVA.	2-way interaction (attenuated)	In support of Hypothesis 3, peak-to-peak FRN was larger for loss than for cash-out (outcome, $F(1, 72) = 192.65$, $p < .001$, $\eta_p^2 = .73$). In further support of Hypothesis 3, the enhancement of FRN for loss compared to cash-out feedback was more prominent in HC than in AUD (group, $F(1, 72) = 2.78$, $p = .100$, $\eta_p^2 = .04$; Group × Outcome, $\mathbf{F(1, 72) = 6.90, p = .011}$, $\eta_p^2 = .09$). Groups differed in FRN after loss feedback, $F(1, 72) = 4.96$, $p = .029$, $\eta_p^2 = .06$, but not after cash-out feedback, $F(1, 72) = 1.73$, $p = .193$, $\eta_p^2 = .02$.	$F(1, 72) = 6.90$ $p = 0.01052$	N/A
Wei et al. (2018)	According to incentive sensitization theory, we hypothesized that **MA use disordered subjects would show** increased amplitude of SPN, but **no difference in the amplitude of FRN** and P300 **compared to HC**.	The FRN data were analysed with a group (MA vs. HC) × magnitude (9 vs. 99) RM-ANOVA, with group as between-subject factor and magnitude as within-subject factor. (…) In order to separate the FRN component, we calculated the difference waveform.	Difference of means (SUDs vs. HC)	A group (MA vs. HC) × magnitude (9 vs. 99) RM-ANOVA of the FRN showed a main effect of group ($\mathbf{F (1, 41) = 4.26, p < 0.05}$, $\eta_p^2 = 0.09$) and magnitude ($F (1, 41) = 8.02$, $p < 0.01$, $\eta_p^2 = 0.2$), while there was no significant interaction effect. The MA group (−4.78 ± 0.63 μV) displayed larger FRN than the HC group (−2.95 ± 0.62 μV).	$F(1, 41) = 4.26$ $p = 0.04539$	N/A
L. Yang et al. (2022)	We hypothesised that there would be **no significant difference in the FRN elicited by positive and negative outcomes in heroin abstainers**. Notably, there was also no significant difference in the P3 elicited by neutral, positive, and negative outcomes in heroin abstainers **compared to controls**.	For the ERP data, mean amplitudes for FRN were analysed using a repeated-measures ANOVA with group as a between-subjects factor and outcome (i.e., positive, negative) as within-subject factors.	2-way interaction (attenuated)	We found a significant interaction between Group*Outcome ($\mathbf{F(1, 53) = 4.58, p = 0.037}$, $\eta_p^2 = 0.080$). The FRN elicited by positive outcomes was significantly more positive than that elicited by negative outcomes for the controls ($p = 0.005$), while there was no significant difference in the FRN at negative and positive outcomes for the heroin abstainers ($p > 0.05$).	$F(1, 53) = 4.58$ $p = 0.03697$	N/A
L. Yang et al. (2024)	We hypothesised that **the FRN amplitude in the OUD would not significantly differ under conditions of monetary gain and loss. However, for the control group, we expected the FRN amplitude elicited under conditions of monetary loss to be significantly higher than that elicited under conditions of monetary gain**.	We analysed the FRN data using a 2 (group: OUD, healthy controls) × 2 (outcome: gain, loss) × 3 (electrode position: Fz, FCz, Cz) repeated-measures ANOVA, with group as a between-subjects factor and outcome and electrode position as within-subjects factors.	2-way interaction (attenuated)	There was a significant interaction between feedback for group and outcome ($\mathbf{F(1, 55) = 7.20, p = .01}$, $\eta_p^2 = 0.116$). A simple effect analysis indicated that FRN elicited by loss-related outcomes was significantly more negative than that elicited by gain-related comes in the control group ($p = .013$). However, there was no significant difference in FRN between gain and loss outcomes in the OUD ($p = .234$).	$F(1, 55) = 7.20$ $p = 0.00961$	N/A
Zhao et al. (2017)	We expect a decreased pre-feedback SPN amplitude in the right hemisphere and an **enhanced FRN difference wave in AHAs compared with HCs**.	For the FRN component, a 2 × 2 repeated measures ANOVA was performed with group (AHA vs. HC) and difference waveforms in amplitude (loss-gain 9/99) as between- and within-subjects factors, respectively.	Difference of means (SUDs vs. HC)	The amplitude of difference waveforms of FRN had a significant main group effect [$\mathbf{F(1, 38) = 6.806}$, $p = 0.013$], indicating that more negative FRN was elicited for AHAs vs. HCs (AHAs: $M = -2.664$, $SE = 0.417$; HCs: $M = -1.125$, $SE = 0.417$].	$F(1, 38) = 6.806$ $p = 0.01292$	N/A
Zhong et al. (2020)	We hypothesised that the **chronic MA users** showed poor risky decision-making ability, and **display neural insensitivity to unfavourable feedback represented by smaller FRN** and P300 **potentials**, which is associated with high impulsive traits **compared to healthy controls**.	Difference waves were produced by subtracting the positive feedback waves from the negative feedback waves. (…) 2 x 3 repeated ANOVAs with groups as a between-subjects factor and electrodes as a within-subjects factor were used.	Difference of means (SUDs vs. HC)	The significant group effect ($\mathbf{F (1, 56) = 4.56, p = 0.037}$) and the electrode effect ($F (2, 112) = 5.86$, $p = 0.012$) were found. (…) The FRN amplitudes in FCz electrode (-2.73 mV Vs -5.77 mV) and Cz electrode (-0.10 mV Vs -4.54 mV) were significant less negative in the MA group than in the HC group ($P < 0.05$).	$F(1, 56) = 4.56$ $p = 0.03711$	N/A

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.11

P-curve Disclosure Table Feedback-P3

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Duven et al. (2015)	We want to explore **whether pathological computer gamers show tolerance effects as evidenced by a reduced P300 in the ERP** or whether motivational attention is enhanced **in reaction to finding a reward** as evidenced by increased amplitude of the P300 and later components.	For the ERP-data separate 2 x 3 x 3 repeated-measures ANOVAs with group as between subject factor (PCG vs. CG) and sagittal (left, central, right) and coronal (frontal, central, parietal) electrode position as within-subject factors were conducted. Where the sphericity assumption was violated, Greenhouse-Geisser-Corrections were applied; otherwise Pillai-Spur is reported. Significant interactions are followed up by Bonferroni-corrected post-hoc analyses.	Difference of means (SUDs vs. HC)	For the P300 peak there was a significant interaction between sagittal electrode position and group ($F(2) = 5.576$, $P = 0.01$). Post-hoc analyses showed that PCG had a significantly less positive P300 amplitude in comparison to CG on the left hemisphere (**$F(1) = 7.403$, $P = 0.012$**) and on midline (**$F(1) = 6.738$, $P = 0.016$**).	$F(1, 25) = 7.403$ $p = 0.01168$	$F(1, 25) = 6.738$ $p = 0.01557$
Kamarajan et al. (2010)	The main goal of the present study is to examine reward/outcome processing in alcohol dependent individuals as compared to healthy controls while they subjectively experience monetary loss and gain during the performance of a gambling task. (…) Our hypotheses were the following: (1) **alcoholics will show decreased amplitude in** both ORN and **ORP components**.	The model included five factors as fixed effects: valence (loss and gain), amount (50¢ and 10¢), region (frontal, central, parietal, occipital, left-temporal, and right-temporal), electrode (six electrodes) as within-subjects factors, and group (control and alcoholic) as a between-subjects factor.	Difference of means (SUDs vs. HC)	Comparisons of ORN and ORP amplitudes across groups in each outcome condition at different electrode sites are illustrated in Fig. 7 along with Bonferroni adjusted significance levels. Alcoholics showed significantly lower ORP amplitude during all outcome conditions and decreased ORN amplitude during loss conditions (-50 and -10). (…) Bonferroni adjusted significance level is marked with asterisks (*p < 0.05, **p < 0.01, and ***p < 0.001).  *REQUESTED: df and t belonging to reported p.*	$t(78) = -3.143$ $p = 0.00236$ Note: *p* value in paper is Bonferroni-corrected (*6)	$t(78) = -3.306$ $p = 0.00143$ Note: *p* value in paper is Bonferroni-corrected (*6)

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.11 (continued)

P-curve Disclosure Table Feedback-P3

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Miedl et al. (2014)	During later time windows, especially **PG were expected to produce enhanced ERPs related to centro-parietal generators triggered by enhanced cognitive processing of motivationally relevant win situations**.	A four-way repeated measures ANOVA was calculated, including the within-subject factors ANTERIOR-POSTERIOR (AP, 3 levels: Frontal, central, and posterior electrodes), LATERALITY (LAT, 5 levels: From right to left scalp electrodes), RISK (high-risk, low-risk) and the between-subject factor GROUP (PG, OG). Analysing the ERP time interval and factor composition during the phase of reward processing was identical as compared to the analysis of risk assessment, but with the factor REWARD (win, lose) instead of RISK.	Difference of means (SUDs vs. HC)	For the 390-440 ms time window a three-way interaction (AP x REWARD x GROUP; F[1.4, 31.6] = 2.99, p = 0.085, GG-adjusted; η^2 = 0.11) revealed a statistical trend. Applied post-hoc tests indicated significantly lower mean amplitude values (difference: win vs. lose) in PG compared to OG at electrode P7 (**t[1,22] = 2.35, p = 0.028**), and higher amplitude values (difference: win vs. lose) in PG compared to OG at C4 electrode (**t[1,22] = 2.18, p = 0.041**).	$t(22) = 2.35$ $p = 0.02815$	$t(22) = 2.18$ $p = 0.04024$
Wei et al. (2018)	We tested reward processing in MA use disordered individuals using the neural correlates in the anticipatory stage as indexed by the SPN, and in the outcome stage as indexed by the FRN and P300. According to incentive sensitization theory, we hypothesized that **MA use disordered subjects would show** increased amplitude of SPN, but **no difference in the amplitude of** FRN and **P300 compared to HC**.	A group (MA vs. HC) × valence (gain vs. loss) × magnitude (9 vs. 99) RM-ANOVA was applied to the P300 data, with group as between-subject factor, valence and magnitude as within-subject factors. When significant interaction effects were indicated, further simple effect analyses followed.	Difference of means (SUDs vs. HC)	Importantly, a significant between-group effect was observed, **$F(1, 41) = 7.65$, $p < 0.01$**, $\eta_p^2 = 0.16$, indicating that the P300 was enhanced in the MA group (17.79 ± 1.21 µV) as compared to the HC group (13.12 ± 1.18 µV). A valence × group interaction effect was found, $F(1, 41) = 10.34$, $p < 0.01$, $\eta_p^2 = 0.20$. Further simple effect analysis revealed that the group effect was significant under the gain condition (**$p < .001$**, MA: 17.25 ± 1.15 µV, HC: 12.29 ± 0.76 µV), but not under the loss condition ($p = 0.11$).	$F(1, 41) = 7.65$ $p = 0.00847$	n.a.

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.12

P-curve Disclosure Table N170 / VPP

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Maurage, Philippot, et al. (2008)	We hypothesised that **people with alcoholism would present** behavioural (i.e., higher error level and longer RTs) and electrophysiological (i.e., delayed latencies and **reduced amplitudes) deficits in unimodal conditions**. Because alcoholism is linked to visuospatial impairments, we hypothesised that the deficit would **begin at the N170 (visual)** and N2 (auditory) stage.	For each component of interest, 2 × 3 × 3 × 5 (2,3) ANOVAs were computed separately for latencies and amplitudes, with group (alcoholism patients and control subjects) as between-factor and emotion (angry, happy, neutral), modality (visual, auditory, auditory-visual) and location (Oz, O1, O2, T5 and T6 for P100; T5 and T6 for N170; Pz, P3 and P4 for P3b) as within-factors.	Difference of means (SUDs vs. HC)	In the second interaction, between group and emotion (F2,56 = 3.32, $p < 0.05$), control subjects had higher N170–N2 amplitudes than alcoholism patients only for happy (**t14 = 2.69, $p < 0.05$**) and neutral (t14 = 2.37, $p < 0.05$) stimuli.	$t(14) = 2.69$ $p = 0.01760$	N/A

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.13

P-curve Disclosure Table Cue-P3 to EFE

Author (year)	Prediction of interest (quoted)	Study design (quoted)	Key statistical result	Statistical results (quoted)	Results	Robustness results
Hoffman et al. (2019)	We obtained both neurophysiological and behavioural measures during the conduct of an EFE task with a sample of treatment-seeking men and women with **AUD and a community comparison (CC) group**. (…) We predicted that the **AUD group would exhibit altered neurophysiological responses as reflected in the** N170 and/or **P3**.	Generalized linear mixed models (GLMM) were used to analyse SJT and EJT outcomes. (…) For each of the tasks, the ERP parameters and behavioural measures were considered in separate analyses. (…) Morph level was designated as the within-participant factor in SJT analyses. Emotion and morph were within-participant factors in EJT models.	Difference of means (SUDs vs. HC)	EJT ERP analyses revealed diminished P3 amplitudes among AUD individuals relative to CCs (**$F(1, 59) = 15.30$, $p < .001$**) with no group by emotion interaction ($F(2, 118) = 0.01$, $p = .986$).	$F(1, 59) = 15.30$ $p = 0.00024$	N/A
Maurage, Campanella, et al. (2008)	ERP were used to define, if a behavioural deficit is found, where this deficit originates from. Indeed, as suggested above, **alcoholism has been related to early perceptive impairments** (P100 and N170) **in the processing of EFE (associated with the classical P3b alteration)**.	For each component of interest, 4 X 3 X 5 (3) ANOVAs were computed separately for latencies and amplitudes, with group (C, A, D, AD) as between-factor, morphing (65, 80 and 95%) and location (Oz, O1, 02, T5, T6 for P100-N170; C3, Cz, C4 for N100; P3, Pz, P4 for P3b) as within-factors.	Difference of means (SUDs vs. HC)	A main effect of group ($F[3, 44] = 3.08$, $P < 0.05$, $\eta^2 = 0.17$) was found – Scheffe post-hoc tests showed that C subjects had higher P3b amplitudes than A (**$P = 0.037$**), D ($P = 0.031$) and AD subjects ($P = 0.011$), which did not differ. Moreover, an interaction was found between group and emotion ($F[6, 88] = 2.38$, $P < 0.05$, $\eta^2 = 0.14$) – for sadness stimuli, C had higher P3b amplitudes than A (**$P = 0.023$**), D ($P = 0.045$) and AD subjects ($P = 0.016$), while for anger and happiness stimuli, controls had higher P3b amplitudes than D (anger: $P = 0.023$; happiness: $P = 0.048$) and AD subjects (anger: $P = 0.0029$; happiness: $P = 0.045$), but not than A subjects. *REQUESTED: df and t belonging to reported p.*	$t(11) = 2.37$ $p = 0.03715$	$t(11) = 2.76$ $p = 0.01856$
Maurage, Philippot, et al. (2008)	We hypothesised that **people with alcoholism would present** behavioural (i.e., higher error level and longer RTs) and electrophysiological (i.e., delayed latencies and **reduced amplitudes) deficits in unimodal conditions**. Because alcoholism is linked to visuospatial impairments, we hypothesised that the deficit would begin at the N170 (visual) and N2 (auditory) stage.	For each component of interest, 2 × 3 × 3 × 5 (2,3) ANOVAs were computed separately for latencies and amplitudes, with group (alcoholism patients and control subjects) As between-factor and emotion (angry, happy, neutral), modality (visual, auditory, auditory-visual) and location (Oz, O1, O2, T5 and T6 for P100; T5 and T6 for N170; Pz, P3 and P4 for P3b) as within-factors.	Difference of means (SUDs vs. HC)	The second interaction was between group, modality and emotion ($F(2,56) = 3.93$, $p < 0.05$). Control subjects had higher P3b amplitudes than alcoholism patients only for auditory-visual happiness ($t(14) = 2.41$, $p < 0.05$), visual anger (**$t(14) = 2.19$, $p < 0.05$**) and visual neutral ($t(14) = 2.63$, $p < 0.05$) stimuli.	$t(14) = 2.19$ $p = 0.04595$	N/A

Note. N/A = not applicable; n.s. = not significant; n.r. = not reported; n.a. = not available.

Table S7.14

Experimental Characteristics N2b Studies

Author (year)	*n* SUDs group	*n* HC group	Mean age (*SD*)	% women	Addiction type	Comorbidity SUDs	Use status	Paradigm
J. Chen et al. (2016)	16	16	19.5 (1.3) SUDs 19.7 (1.3) HC	56% SUDs 44% HC	Smartphone	No history of seizures; periods of unconsciousness; psychiatric illness	Current	Go/No-Go task (0.2 no-go)
Y. Chen et al. (2022)	26 (25 for analysis)	24	20.2 (1.7) SUDs 20.2 (1.4) HC	46% SUDs 50% HC	Internet gaming	FTND $M = 0.7$ ($SD = 2.2$)	Current	Go/No-Go task (0.3 no-go)
Dong et al. (2010)	12	12	20.5 (4.1) SUDs 20.2 (4.5) HC	0% SUDs 0% HC	Internet	No neurological or psychiatric disorders	Current	Go/No-Go task (0.25 no-go)
Fathi et al. (2022)	30	30	20.4 (3.0) SUDs 19.9 (1.9) HC	0% SUDs 0% HC	Internet gaming	No drug or alcohol abuse (except smoking); traumatic brain injury; mental or neurological disorders; history of memory disorders	Current	Selective stop-signal task (0.4 stop)
Fathi et al. (2024)	25	25	20.4 (3.0) SUDs 19.9 (1.9) HC	0% SUDs 0% HC	Internet gaming	No substance abuse (except smoking); traumatic brain injury; psychological or neurological disorders; history of severe memory problems	Current	Cued Go/No-Go task (0.33 no-go)
Morie et al. (2014)	23	27	44.0 (6.6) SUDs 41.0 (8.5) HC	30% SUDs 26% HC	Cocaine	No other Axis I diagnosis (except dependence, and depression or dysthymia caused by drug use); head trauma resulting in loss of consciousness for 30+ minutes; past or current brain pathology; HIV	Current	Go/No-Go task (0.15 no-go)
Pandey et al. (2012)	78	58	40.7 (6.4) SUDs 21.1 (2.5) HC	0% SUDs 0% HC	Alcohol	No liver disease or head injury; history of other SUDs, ASPD, CD, ADHD; ODD was allowed	Abstinent for 30+ days	Go/No-Go task (0.5 no-go)
Sokhadze et al. (2008)	19	15	42.1 (5.5) SUDs 37.0 (9.4) HC	37% SUDs 53% HC	Cocaine	No other Axis I disorder (except PTSD); current psychiatric symptoms requiring medication; neurological disorders that can affect the EEG; severe medical/psychiatric impairments that interfere with partaking	16 current; 3 abstinent for <60 days	Flanker task with no-go aspects (0.2 no-go)
B. Yang et al. (2009)	14	14	41.0 (7.1) SUDs 41.0 (10.5) HC	0% SUDs n.r. HC	Heroin	No history of head trauma or other conditions that can cause cognitive impairment	Abstinent for $M = 4.7$ ($SD = 6.4$) months	Go/No-Go task (0.5 no-go)

Note. SUDs = substance use disorders; HC = healthy controls; SD = standard deviation; n.r. = not reported.

Table S7.15

Experimental Characteristics No-Go P3 Studies

Author (year)	*n* SUDs group	*n* HC group	Mean age (*SD*)	% women	Addiction type	Comorbidity SUDs	Use status	Paradigm
Y. Chen et al. (2022)	26 (25 for analysis)	24	20.2 (1.7) SUDs 20.2 (1.4) HC	46% SUDs 50% HC	Internet gaming	FTND $M = 0.7$ ($SD = 2.2$)	Current	Go/No-Go task (0.3 no-go)
Colrain et al. (2011)	10	25	53.1 (9.1) SUDs 56.5 (12.8) HC	50% SUDs 52% HC	Alcohol	No other Axis I disorder	Abstinent for 2 days to 2.5 years	Go/No-Go task (0.12 no-go)
Dong et al. (2010)	12	12	20.5 (4.1) SUDs 20.2 (4.5) HC	0% SUDs 0% HC	Internet	No neurological or psychiatric disorders	Current	Go/No-Go task (0.25 no-go)
Fathi et al. (2022)	30	30	20.4 (3.0) SUDs 19.9 (1.9) HC	0% SUDs 0% HC	Internet gaming	No drug or alcohol abuse (except smoking); traumatic brain injury; mental or neurological disorders; history of memory disorders	Current	Selective stop-signal task (0.4 stop)
Ghin et al. (2022)	59	64	27.1 (6.1) SUDs 27.1 (5.6) HC	48% SUDs 56% HC	Alcohol	No medical, neurological, or mental illness (except alcohol addiction)	Current	Simon task with no-go aspects (0.3 no-go)
Maij et al. (2017)	37 (35 for analysis)	39	21.7 (2.1) SUDs 22.1 (2.1) HC	16.2% SUDs 30.8% HC	Cannabis	No other drug or alcohol addiction (except smoking); no history of head trauma; no severe current psychiatric symptoms; ADD and ADHD were allowed and were diagnosed in 4 participants	Abstinent for 2-6 weeks	Go/No-Go task (0.25 no-go)
Morie et al. (2014)	23	27	44.0 (6.6) SUDs 41.0 (8.5) HC	30% SUDs 26% HC	Cocaine	No other Axis I diagnosis (except dependence, and depression or dysthymia caused by drug use); head trauma resulting in loss of consciousness for 30+ minutes; past or current brain pathology; HIV	Current	Go/No-Go task (0.15 no-go)
Sokhadze et al. (2008)	19	15	42.1 (5.5) SUDs 37.0 (9.4) HC	37% SUDs 53% HC	Cocaine	No other Axis I disorder (except PTSD); current psychiatric symptoms requiring medication; neurological disorders that can affect the EEG; severe medical/psychiatric impairments that interfere with partaking	16 current; 3 abstinent for < 60 days	Flanker task with no-go aspects (0.2 no-go)

Note. SUDs = substance use disorders; HC = healthy controls; SD = standard deviation; n.r. = not reported.

Table S7.16

Experimental Characteristics P3a Studies

Author (year)	*n* SUDs group	*n* HC group	Mean age (*SD*)	% women	Addiction type	Comorbidity SUDs	Use status	Paradigm
Holguin et al. (1999)	44	28	37.0 (6.8) SUDs 27.7 (7.8) HC	0% SUDs 0% HC	Alcohol	No history of psychiatric illness or drug abuse (except secondary to alcoholism); major medical problems; liver damage; seizures unrelated to withdrawal	Abstinent for ± 28 days	3-stimulus visual oddball task (0.10 distractor)
X. Liu et al. (2020)	30	30	43.2 (7.4) SUDs 43.9 (7.6) HC	0% SUDs 0% HC	Alcohol	No neurological illness or comorbid psychiatric illness; other substance dependence including smoking	Abstinent for 4 weeks	3-stimulus auditory oddball task (0.12 distractor)

Note. SUDs = substance use disorders; HC = healthy controls; SD = standard deviation; n.r. = not reported.

Table S7.17

Experimental Characteristics P3b Studies

Author (year)	*n* SUDs group	*n* HC group	Mean age (*SD*)	% women	Addiction type	Comorbidity SUDs	Use status	Paradigm
Boutros et al. (2000)	10	10	48.8 (5.5) SUDs 48.9 (4.6) HC	40% SUDs 50% HC	Alcohol	No comorbid Axis I or II diagnosis	n.r.	2-stimulus auditory oddball task (0.15 target)
H. L. Cohen et al. (2002)	30	39	30.3 (4.3) SUDs 24.7 (3.3) HC	0% SUDs 0% HC	Alcohol	No major medical problems; history of psychiatric problems	Abstinent for 30+ days	2-stimulus auditory oddball task & 3-stimulus visual oddball task (both 0.125 target)
Fein & Andrew (2011)	76	65 (63 for analysis)	31.7 (8.1) $SUDs_{male}$ 30.8 (7.9) $SUDs_{female}$ 32.1 (8.4) HC_{male} 32.4 (8.4) HC_{female}	43% SUDs 43% HC	Alcohol	No lifetime or current schizophrenia or schizophreniform disorder; lifetime or current drug dependence or abuse (except nicotine or caffeine); history of significant neurological disease, head trauma or cranial surgery; diabetes, stroke, or hypertension requiring medical intervention; Wernicke-Korsakoff syndrome	Current	3-stimulus visual oddball task (0.125 target)
Glenn et al. (1996)	51	24	37.1 (7.4) SUDs 34.0 (8.9) HC	0% SUDs n.r. HC	Alcohol	No acute medical conditions; history of neurological disease or trauma; major psychiatric disorder, substantial abuse of drugs other than alcohol	Abstinent for 21-45 days	2-stimulus visual oddball task with simultaneous 2-stimulus auditory oddball task (0.20 target)
Gooding et al. (2008)	14	15	39.0 (5.5) SUDs 34.3 (7.2) HC	50% SUDs 53% HC	Cocaine	No other Axis I disorder that was independent of cocaine use; 71% reported cocaine-induced psychotic symptoms	Abstinent for 3+ weeks	2-stimulus auditory oddball task (0.15 target)
Guney et al. (2009)	32	32	40.6 (13.1) SUDs 37.1 (10.0) HC	44% SUDs 50% HC	Smoking	No lifetime history of a major medical disorder (neurological, hepatic or cardiovascular); head injury resulting in a loss of consciousness; seizures; DSM-defined sedative or barbiturate, alcohol or cocaine abuse or dependence	Current	2-stimulus auditory oddball task (0.20 target)
K. A. Jones et al. (2006)	100	100	30.1 (5.3) SUDs 29.6 (5.7) HC	0% SUDs 0% HC	Alcohol	No history of major medical and neurological conditions including head injury	Abstinent for 30+ days	3-stimulus visual oddball task (0.125 target)
Kuntz et al. (2021)	18	18	44.9 (10.5) SUDs 44.8 (9.3) HC	22% SUDs 33% HC	Alcohol	89% had psychiatric comorbidities; 28% had a comorbid addiction	One recording at current use, and one after 2 weeks abstinence	2-stimulus auditory oddball task with button-pushing and another with backward counting (0.20 target)
X. Liu et al. (2020)	30	30	43.2 (7.4) SUDs 43.9 (7.6) HC	0% SUDs 0% HC	Alcohol	No neurological illness or comorbid psychiatric illness; other substance dependence including smoking	Abstinent for 4 weeks	3-stimulus auditory oddball task (0.12 distractor)

Note. SUDs = substance use disorders; HC = healthy controls; SD = standard deviation; n.r. = not reported.

Table S7.17 (continued)

Experimental Characteristics P3b Studies

Author (year)	*n* SUDs group	*n* HC group	Mean age (*SD*)	% women	Addiction type	Comorbidity SUDs	Use status	Paradigm
Moeller et al. (2004)	17	14	37.3 (6.2) SUDs 32.0 (8.2) HC	29% SUDs 50% HC	Cocaine	No Axis I disorders (except cocaine abuse or dependence); current or past medical disorders that affect the central nervous system	Current	2-stimulus auditory oddball task (0.20 target)
Motlagh et al. (2017)	19	19	39.9 (7.7) SUDs 33.2 (8.1) HC	0% SUDs n.r. HC	Heroin	No acute psychiatric or neurological disorders (except heroin abuse or dependence); lifetime history of a major medical disorder; HIV; head injury resulting in loss of consciousness; seizures; FTND $M = 6.2$ ($SD = 1.4$)	Current	4-stimulus auditory oddball task (0.068 target)
Olbrich et al. (2000)	27	15	42.8 (7.7) SUDs 42.8 (6.7) HC	n.r.	Alcohol	No major medical problems; history of neurological disease or trauma; major psychiatric disorders, substantial abuse or use of drugs other than alcohol	Abstinent for 13-34 days ($M = 22$ days)	2-stimulus visual oddball task (0.15 target)
Park et al. (2016)	26	23	23.0 (4.2) SUDs 25.0 (4.3) HC	23% SUDs 13% HC	Internet gaming	No history of significant head injury; seizure disorder; mental retardation; psychotic disorder; substance use disorder (except nicotine); 4 SUDs participants fulfilled DSM-IV criteria for depressive disorder and 3 fulfilled DSM-IV criteria for anxiety disorder	Current	2-stimulus auditory oddball task (0.15 target)
Park et al. (2017)	18	29	22.6 (5.1) SUDs 24.7 (3.8) HC	0% SUDs 0% HC	Internet gaming	No history of significant head injury; seizure disorder; intellectual disability; psychotic disorder; substance use disorder (except nicotine)	Current	2-stimulus auditory oddball task (0.15 target)
Park et al. (2023)	23	25	23.1 (5.2) SUDs 25.0 (3.9) HC	0% SUDs 0% HC	Internet gaming	No history of head injury; neurological disease; other psychiatric or psychotic disorder	Current	2-stimulus auditory oddball task (0.15 target)
Parsons et al. (1990)	143 (128 for analysis)	97 (77 for analysis)	n.r.	47% SUDs 51% HC	Alcohol	No history or presence of severe mental disease (schizophrenia or manic-depressive disorder); positive neurological history; positive medical history for diseases (e.g., chronic pulmonary or heart disease) that could affect central nervous system functioning	Abstinent for 3-6 weeks	2-stimulus visual oddball task with simultaneous 2-stimulus auditory oddball task (0.20 target)
Polich & Ochoa (2004)	40	40	21.9 (3.9) full group	50% full group	Smoking	No major medical problems; history of psychiatric problems; history of drug abuse (except nicotine)	Current	3-stimulus visual oddball task (0.125 target)
Whipple et al. (1991)	15 (11 for analysis)	30 (27 for analysis)	41.0 (1.3) SUDs n.r. HC	0% SUDs 0% HC	Alcohol	No history or current status of medical or psychiatric disorder (except alcoholism); history of major head trauma	Abstinent for 2+ years	Visual continuous performance task (0.10 target)

Note. SUDs = substance use disorders; HC = healthy controls; SD = standard deviation; n.r. = not reported.

Table S7.18

Experimental Characteristics ERN Studies

Author (year)	*n* SUDs group	*n* HC group	Mean age (*SD*)	% women	Addiction type	Comorbidity SUDs	Use status	Paradigm
H. Chen, Jiang, et al. (2013)	20 (17 for analysis)	15	37.1 (9.5) SUDs 32.5 (9.9) HC	0% SUDs 0% HC	Heroin	No history of mental retardation; serious physical or mental illness	Abstinent for *M* = 8.2 (*SD* = 2.0) months	Flanker task
Franken et al. (2018)	34 (30 for analysis)	34 (31 for analysis)	19.9 (1.7) SUDs 20.8 (3.0) HC	n.r.	Food	No history of drug abuse; current self-reported psychiatric or physical illness	Current	Flanker task
Littel, Van den Berg, et al. (2012)	25	27	20.5 (3.0) SUDs 21.4 (2.6) HC	8% SUDs 63% HC	Internet gaming	n.r.	Current	Go/No-Go task (0.116 no-go)
Marhe et al. (2013)	49	23	39.6 (8.4) SUDs 39.9 (9.4) HC	11% SUDs 26% HC	Cocaine	No indications of severe psychopathology (i.e., psychosis, severe mood disorder) as assessed by a physician	Abstinent for 3+ days	Flanker task
Morie et al. (2014)	23	27	44.0 (6.6) SUDs 41.0 (8.5) HC	30% SUDs 26% HC	Cocaine	No other Axis I diagnosis (except dependence, and depression or dysthymia caused by drug use); head trauma resulting in loss of consciousness for 30+ minutes; past or current brain pathology; HIV	Current	Go/No-Go task (0.15 no-go)
Padilla et al. (2011)	14	14	37.9 (9.3) SUDs 43.5 (14.5) HC	0% 0%	Alcohol	n.r.	Abstinent for 1-4 months	Flanker task
Park et al. (2020)	34	34	25.9 (6.0) SUDs 25.5 (4.3) HC	12% SUDs 12% HC	Internet gaming	No psychiatric diagnoses (except internet gaming disorder); history of head injury; cognitive delay	Current	Go/No-Go task (0.293 no-go)
Sokhadze et al. (2008)	19 (6 for analysis)	15 (6 for analysis)	42.1 (5.5) SUDs 37.0 (9.4) HC	37% SUDs 53% HC	Cocaine	No other Axis I disorder (except PTSD); current psychiatric symptoms requiring medication; neurological disorders that can affect the EEG; severe medical/psychiatric impairments that interfere with partaking	16 current; 3 abstinent for <60 days	Flanker task with no-go aspects (0.2 no-go)

Note. SUDs = substance use disorders; HC = healthy controls; SD = standard deviation; n.r. = not reported.

Table S7.19

Experimental Characteristics Pe Studies

Author (year)	*n* SUDs group	*n* HC group	Mean age (*SD*)	% women	Addiction type	Comorbidity SUDs	Use status	Paradigm
Franken, Van Strien, & Kuijpers (2010)	23 (21 for analysis)	28 (25 for analysis)	21.7 (2.7) SUDs 21.3 (2.8) HC	48% SUDs 56% HC	Smoking	n.r.	Current	Flanker task
Morie et al. (2014)	23	27	44.0 (6.6) SUDs 41.0 (8.5) HC	30% SUDs 26% HC	Cocaine	No other Axis I diagnosis (except dependence, and depression or dysthymia caused by drug use); head trauma resulting in loss of consciousness for 30+ minutes; past or current brain pathology; HIV	Current	Go/No-Go task (0.15 no-go)
Park et al. (2020)	34	34	25.9 (6.0) SUDs 25.5 (4.3) HC	12% SUDs 12% HC	Internet gaming	No psychiatric diagnoses (except internet gaming disorder); history of head injury; cognitive delay	Current	Go/No-Go task (0.293 no-go)

Note. SUDs = substance use disorders; HC = healthy controls; SD = standard deviation; n.r. = not reported.

Table S7.20

Experimental Characteristics Cue-P3 to Drug Cues Studies

Author (year)	*n* SUDs group	*n* HC group	Mean age (*SD*)	% women	Addiction type	Comorbidity SUDs	Use status	Paradigm
Cheng et al. (2016)	19	19	21.0 (1.4) SUDs 20.5 (1.2) HC	0% SUDs 0% HC	Smoking	No neurological or psychopathological diseases; epilepsy; substance abuse (except nicotine)	Current	Passive viewing task with SUDs and neutral images
Heinze et al. (2007)	10	10	40.2 (6.1) SUDs 34.0 (4.0) HC	0% SUDs 0% HC	Alcohol	No substance dependence or abuse (except alcohol and smoking); 80% smoked cigarettes	Abstinent for 6 weeks	Passive listening task with SUDs and neutral sounds
C. Li et al. (2024)	22	23	19.5 (1.3) SUDs 19.3 (1.8) HC	59% SUDs 61% HC	Internet gaming	No history of potential mental illness; substance use disorder; neurological disease; severe head injuries with loss of consciousness	Current	Stimulus-response compatibility task with SUDs and neutral images
Littel & Franken (2007)	21	18	21.6 (2.5) SUDs 23.1 (4.1) HC	n.r.	Smoking	n.r.	Current	Passive viewing task with SUDs and neutral images
Littel & Franken (2012)	30	31	21.9 (3.0) SUDs 20.5 (1.9) HC	83% SUDs 84% HC	Smoking	n.r.	Current	Conditioning task with SUDs and neutral images
Lubman et al. (2007)	14 (13 for analysis)	14 (12 for analysis)	29.6 (7.4) SUDs 28.5 (5.1) HC	14% SUDs 21% HC	Heroin	No history of major mental illness; significant head injury; neurological disorder	Methadone-maintained; abstinent for 2.5-296 days (median 83.5)	Oddball task with SUDs and neutral images
Lubman et al. (2008)	20 (16 for analysis)	13 (12 for analysis)	33.3 (5.6) SUDs 31.8 (5.4) HC	0% SUDs 0% HC	Heroin	No history of major mental illness; significant head injury; neurological disorder	Abstinent for 2 weeks; half methadone-, cyclomorphine-, or dihydrocodeine-maintained and half drug-free	Passive viewing task with SUDs, positive, negative, and neutral images
Motlagh et al. (2017)	19	19	39.9 (7.7) SUDs 33.2 (8.1) HC	0% SUDs n.r. HC	Heroin	No acute psychiatric or neurological disorders (except heroin abuse or dependence); lifetime history of a major medical disorder; HIV; head injury resulting in loss of consciousness; seizures; FTND $M = 6.2$ ($SD = 1.4$)	Current	Oddball task with SUDs, positive, and neutral images
Thalemann et al. (2007)	15	15	28.8 (6.1) SUDs 25.7 (8.1) HC	0% SUDs 0% HC	Internet gaming	$M = 5.6$ ($SD = 8.74$) cigarettes a day	Current	Passive viewing task with SUDs, alcohol, positive, negative, and neutral images

Note. SUDs = substance use disorders; HC = healthy controls; SD = standard deviation; n.r. = not reported.

Table S7.21

Experimental Characteristics LPP to Drug Cues Studies

Author (year)	*n* SUDs group	*n* HC group	Mean age (*SD*)	% women	Addiction type	Comorbidity SUDs	Use status	Paradigm
Cheng et al. (2016)	19	19	21.0 (1.4) SUDs 20.5 (1.2) HC	0% SUDs 0% HC	Smoking	No neurological or psychopathological diseases; epilepsy; substance abuse (except nicotine)	Abstinent for 12+ hours	Passive viewing task with SUDs and neutral images
Franken et al. (2003)	19	14	33.5 (7.7) SUDs 33.7 (9.0) HC	0% SUDs 0% HC	Heroin	No schizophrenia; affective disorder; mental retardation; significant somatic disorders such as Parkinson disease or symptoms thereof	Abstinent for 2+ weeks	Passive viewing task with SUDs, positive, negative, and neutral images
Heinze et al. (2007)	10	10	40.2 (6.1) SUDs 34.0 (4.0) HC	0% SUDs 0% HC	Alcohol	No substance dependence or abuse (except alcohol and smoking); 80% smoked cigarettes	Abstinent for 6 weeks	Passive listening task with SUDs and neutral sounds
B. Kim et al. (2021)	40	39	25.3 (5.6) SUDs 25.1 (3.3) HC	10% SUDs 26% HC	Internet gaming	No lifetime diagnosis of substance abuse or dependence (except nicotine); psychiatric or neurological disorders; significant head injuries with loss of consciousness; medical illness with documented cognitive sequelae; intellectual disability; 23% of SUDs participants smoked, *M* = 18.9 (*SD* = 9.3) cigarettes a day	Current	Passive viewing task with SUDs and neutral images
Minnix et al. (2013)	180	40	45.1 (10.6) SUDs 46.2 (11.0) HC	35% SUDs 50% HC	Smoking	No current psychiatric disorder; uncontrolled medical illness	Current	Passive viewing task with SUDs, positive, negative, and neutral images
Parvaz et al. (2017)	19	18	41.7 (7.4) SUDs 43.2 (6.2) HC	42% SUDs 28% HC	Cocaine	No history of head trauma or loss of consciousness; neurologic diseases; current medical diseases that required hospitalisation or monitoring; positive urine screens for psychoactive drugs or metabolites other than cocaine; comorbidities included marijuana use disorder (n = 4), ecstasy abuse (n = 1), alcohol use disorder (n = 5), opiate use disorder (n = 1), antisocial personality disorder (n = 4) and posttraumatic stress disorder (n = 1); 58% of SUDs participants smoked, *M* = 5.3 (*SD* = 3.6) cigarettes a day	Time 1: abstinent for 3+ weeks (median 75 days) Time 2: median 210 days	Passive viewing task with SUDs, positive, negative, and neutral images
Wölfling et al. (2008)	15	15	29.0 (6.3) SUDs 26.8 (3.5) HC	53% SUDs 53% HC	Cannabis	No diagnosis of dependence (except cannabis or tobacco) or any other mental disorder; 93% of SUDs participants smoked, *M* = 12.1 (*SD* = 6.7) cigarettes a day	Current	Passive viewing task with SUDs, alcohol, positive, negative, and neutral images
Wölfling et al. (2011)	15	15	34.9 (9.8) SUDs 34.3 (5.3) HC	20% SUDs 13% HC	Gambling	No addiction to a psychotropic substance (except tobacco); 80% of SUDs participants smoked, *M* = 15.1 (*SD* = 9.3) cigarettes a day	Current	Passive viewing task with SUDs, alcohol, positive, negative, and neutral images

Note. SUDs = substance use disorders; HC = healthy controls; SD = standard deviation; n.r. = not reported.

Table S7.22

Experimental Characteristics FRN / RewP Studies

Author (year)	*n* SUDs group	*n* HC group	Mean age (*SD*)	% women	Addiction type	Comorbidity SUDs	Use status	Paradigm
Kamarajan et al. (2010)	40	40	38.3 (6.4) SUDs 21.1 (3.4) HC	0% SUDs 0% HC	Alcohol	No history of major medical or psychiatric disorders and substance-related addictive illnesses; liver disease; head injury	Abstinent for 28+ days	Simple gambling task
Lole et al. (2015)	16	20	34.8 (16.8) SUDs 28.8 (11.2) HC	31% SUDs 55% HC	Gambling	No history of seizures or severe head injury	Current	Slot machine task
Sehrig et al. (2019)	39	35	43.2 (8.4) SUDs 42.7 (9.9) HC	28% SUDs 34% HC	Alcohol	No history of neurological condition or disorder, including epilepsy or head trauma with loss of consciousness	Abstinent	Balloon Analogue Risk Task
Wei et al. (2018)	21	22	24.2 (3.1) SUDs 24.4 (3.5) HC	100% SUDs 100% HC	Methamphetamine	No other DSM-5 diagnosis (including other substance use disorders); brain injuries resulting in loss of consciousness for longer than 30 minutes; current or past brain pathology	Abstinent for 2-21 months	Simple gambling task
L. Yang et al. (2022)	28	27	49.2 (3.7) SUDs 51.3 (7.4) HC	0% SUDs 0% HC	Heroin	No history of psychiatric or neurological disorders; severe head injury	Abstinent for *M* = 10.5 (*SD* = 8.0) months	Social Incentive Delay Task
L. Yang et al. (2024)	29	28	49.1 (4.9) SUDs 51.5 (7.4) HC	0% SUDs 0% HC	Heroin	No history of psychiatric or neurological disorders; severe head injury	Abstinent for *M* = 1.3 (*SD* = 6.7) months	Monetary Incentive Delay Task
Zhao et al. (2017)	20	20	35.4 (8.2) SUDs 31.8 (5.7) HC	50% SUDs 50% HC	Heroin	No history of other addictions or any neurological or psychological disorders	Abstinent for 4+ months (*M* = 9.0 months, *SD* = 3.4 months)	Simple gambling task
Zhong et al. (2020)	31	27	33.5 (6.9) SUDs 31.0 (7.2) HC	48% SUDs 33% HC	Methamphetamine	No severe cognitive impairments; other substance abuse or dependence (except nicotine); other present or history DSM-IV axis I disorders (except past methamphetamine induced psychosis)	Abstinent for *M* = 76.4 (*SD* = 43.4) days	Balloon Analogue Risk Task

Note. SUDs = substance use disorders; HC = healthy controls; SD = standard deviation; n.r. = not reported.

Table S7.23

Experimental Characteristics Feedback-P3 Studies

Author (year)	*n* SUDs group	*n* HC group	Mean age (*SD*)	% women	Addiction type	Comorbidity SUDs	Use status	Paradigm
Duven et al. (2015)	14	13	24.3 (5.8) SUDs 23.3 (3.0) HC	0% SUDs 0% HC	Internet gaming	n.r.	Current	Token search task
Kamarajan et al. (2010)	40	40	38.3 (6.4) SUDs 21.1 (3.4) HC	0% SUDs 0% HC	Alcohol	No history of major medical or psychiatric disorders and substance-related addictive illnesses; liver disease; head injury	Abstinent for 28+ days	Simple gambling task
Miedl et al. (2014)	12	12	33.8 (7.8) SUDs 35.8 (9.5) HC	0% SUDs 0% HC	Gambling	No history of psychiatric or neurological illness	Current	Blackjack task
Wei et al. (2018)	21	22	24.2 (3.1) SUDs 24.4 (3.5) HC	100% SUDs 100% HC	Methamphetamine	No other DSM-5 diagnosis (including other substance use disorders); brain injuries resulting in loss of consciousness for longer than 30 minutes; current or past brain pathology	Abstinent for 2-21 months	Simple gambling task

Note. SUDs = substance use disorders; HC = healthy controls; SD = standard deviation; n.r. = not reported.

Table S7.24

Experimental Characteristics N170 / VPP Studies

Author (year)	*n* SUDs group	*n* HC group	Mean age (*SD*)	% women	Addiction type	Comorbidity SUDs	Use status	Paradigm
Maurage, Philippot, et al. (2008)	15	15	46.7 (12.0) SUDs 42.1 (14.2) HC	33% SUDs 33% HC	Alcohol	No major medical problems; central nervous system disease (including epilepsy); other psychiatric diagnosis; polysubstance abuse	Abstinent for 2+ weeks	Emotion discrimination task with happy, angry, and neutral faces and voices (unimodal and cross-modal)

Note. SUDs = substance use disorders; HC = healthy controls; SD = standard deviation; n.r. = not reported.

Table S7.25

Experimental Characteristics Cue-P3 to EFE Studies

Author (year)	*n* SUDs group	*n* HC group	Mean age (*SD*)	% women	Addiction type	Comorbidity SUDs	Use status	Paradigm
Hoffman et al. (2019)	34 (28 for analysis)	39 (33 for analysis)	40.8 (8.8) SUDs 43.8 (12.1) HC	18% SUDs 62% HC	Alcohol	No significant neurologic disorder/insult; medical conditions that challenged interpretations of neurobehavioral function; lifetime psychotic or bipolar disorder; current major depression; significant anxiety-related disorders (e.g., PTSD); 76% of SUDs participants smoked	Abstinent for 21-90 days (M = 41.0, SD = 12.6)	Emotion discrimination task with happy, angry, and sad faces morphed from neutral faces at 35%, 65%, and 95%
Maurage, Campanella, et al. (2008)	12	12	41.8 (6.7) SUDs 41.8 (9.1) HC	42% SUDs 42% HC	Alcohol	No major medical problems; neurological disease (including epilepsy); other psychiatric diagnosis; polysubstance abuse	Abstinent for M = 17.4 (SD = 3.7) days	Emotion discrimination task with happy, angry, and sad faces morphed from neutral faces at 35%, 65%, and 95%
Maurage, Philippot, et al. (2008)	15	15	46.7 (12.0) SUDs 42.1 (14.2) HC	33% SUDs 33% HC	Alcohol	No major medical problems; central nervous system disease (including epilepsy); other psychiatric diagnosis; polysubstance abuse	Abstinent for 2+ weeks	Emotion discrimination task with happy, angry, and neutral faces and voices (unimodal and cross-modal)

Note. SUDs = substance use disorders; HC = healthy controls; SD = standard deviation; n.r. = not reported.

Table S7.26

Inputted Values for the P-Curve Analyses

P-curve	Main Analysis	Robustness Analysis
N2b	F(1, 47) = 5.35 F(1, 22) = 6.92 F(1, 60) = 11.324 F(1, 46) = 6.133 t(49) = 5.8 t(134) = -7.21 F(1, 32) = 4.90 F(1, 126) = 7.81	-
No-Go P3	F(1, 30) = 4.73 t(121) = 2.298 t(106) = 2.48 t(49) = 6.2 F(1, 32) = 4.36	F(1, 30) = 4.73 t(121) = 2.298 t(106) = 2.54 t(49) = 6.2 F(1, 32) = 4.36
P3b	t(18) = 3.028 F(14, 120) = 2.99 F(1, 135) = 3.914 F(1, 73) = 5.00 t(27) = 4.04 t(62) = -3.202 F(1, 196) = 9.6 z = -3.34 F(1, 58) = 8.817 F(1, 26) = 11.6 F(1, 36) = 13.83 t(28.352) = 2.4386 F(1, 47) = 8.02 F(1, 40) = 5.458 F(1, 43) = 5.242 F(1, 196) = 4.45 F(1, 76) = 8.01 F(1, 36) = 10.31	F(10, 124) = 3.17 F(1, 135) = 3.914 F(1, 73) = 5.00 t(27) = 4.04 t(62) = -3.202 F(1, 196) = 9.6 z = -3.34 F(1, 58) = 8.817 F(1, 26) = 11.6 F(1, 36) = 13.83 t(28.352) = 2.4386 F(1, 47) = 8.02 F(1, 40) = 5.681 F(1, 43) = 5.823 F(1, 89) = 5.88 F(1, 76) = 14.06 F(1, 36) = 10.31
ERN	t(30) = 2.73 t(59) = 2.163866 t(50) = 4.225626 F(1, 70) = 10.03 t(49) = 4.2 F(1, 65) = 6.993 F(1, 11) = 7.42	t(30) = 2.62 t(59) = 2.163866 t(50) = 4.225626 F(1, 70) = 10.03 t(49) = 4.2 F(1, 65) = 6.993 F(1, 11) = 7.42
Drug-P3	F(1, 34) = 6.69 F(1, 18) = 56.26 F(1, 40) = 25.73 t(57) = 2.5125938 F(1, 59) = 6.57 F(1, 23) = 4.39 F(1, 26) = 6.37 F(1, 36) = 7.73 F(1, 28) = 7.84	F(1, 34) = 6.69 F(1, 18) = 13.67 F(1, 40) = 25.73 t(57) = 2.9311073 F(1, 59) = 6.57 F(1, 23) = 4.39 F(1, 26) = 6.37 F(1, 36) = 7.73 F(1, 28) = 7.84
Drug-LPP	F(1, 34) = 26.02 F(1, 30) = 8.87 F(1, 18) = 21.64 F(1, 77) = 6.30 t(654) = 3.05 t(35) = 2.92 F(1, 28) = 17.33 F(1, 28) = 17.23	F(1, 34) = 26.02 F(1, 30) = 8.87 F(1, 18) = 6.11 F(1, 77) = 4.80 t(1962) = 2.35 t(35) = 2.73 F(1, 28) = 24.19 F(1, 28) = 22.08
FRN / RewP	t(78) = 4.073 F(1, 34) = 4.89 F(1, 72) = 6.90 F(1, 53) = 4.58 F(1, 55) = 7.20 F(1, 56) = 4.56	t(78) = 3.914 F(1, 34) = 4.89 F(1, 72) = 6.90 F(1, 53) = 4.58 F(1, 55) = 7.20 F(1, 56) = 4.56

Chapter 8

General Discussion

8.1 Summary of the Research Findings and Conclusions

The use of behavioural and (neuro)physiological measures for operationalising *risk-taking propensity (RTP)* and related concepts like reward responsiveness and impulsivity has increased in recent years. In part, this seems to be due to the assumption that behavioural and (neuro)physiological measures provide more accurate operationalisations than self-reports. However, critical evaluation of these measures is in short supply. For this reason, the present dissertation posed the following question: to what extent are behaviour and brain activity recorded during computer tasks in a laboratory setting successful in operationalising concepts such as risk-taking propensity?

8.1.1 Part I: Behavioural Tasks

The suitability of behavioural tasks for measuring risk-taking propensity was examined in Chapters 2, 3, and 4. First, we focused our attention on the Balloon Analogue Risk Task (BART; Lejuez et al., 2002). This is the most popular instrument for measuring RTP in psychology (C. Wang et al., 2025), and is also often used as a benchmark against which new risk-taking tasks are evaluated. As such, the BART has had a profound impact on the field. At face value, the BART seems suited as an operationalisation of risk-taking propensity: participants engage in an intuitive risk-taking activity where their decisions have tangible consequences. However, Chapters 2 and 3 of this dissertation demonstrated that the BART suffers from methodological shortcomings. First, it is unclear in the BART which decisions are made under uncertainty and which under risk. In addition, the ratio between the two types of decision differs between participants based on their learning abilities. Second, a subset of observations in the BART is censored, with participants who take more risk having a larger number of censored trials. Third, the risk of decisions taken in the BART is confounded with their expected value, which cannot be decomposed. Finally, participants take relatively little risk in the BART, regardless of whether they are risk-seeking or risk-averse, which makes it hard to measure individual differences and distinguish adaptive from maladaptive behaviour. Taken together, these problems make the BART – in its current form – an unfit measure for operationalising risk-taking propensity. Therefore, researchers who have used the BART should critically reflect on its limitations when interpreting their findings; and researchers intending to measure RTP are advised to use a different instrument, such as the Columbia Card Task (CCT; Figner et al., 2009).

In this dissertation, the Columbia Card Task was presented as an alternative to the Balloon Analogue Risk Task. Like the BART, the CCT is a sequential risk-taking task in which participants' decisions have tangible (monetary) consequences – they can win or they

can lose. However, its design gets around the methodological problems that the BART suffers from. Specifically, in the CCT, participants make decisions under clear conditions of risk – although conditions of uncertainty can also easily be constructed. This risk condition is not confounded by expected value, and it is easy in the CCT to distinguish adaptive from maladaptive risk-taking. Various versions of the CCT exist, with the "warm" version that was applied in Chapter 4 of the dissertation being free from statistical censoring. As a bonus, the design of the CCT allows researchers to extract data on participants' information use, that is, the impact that information on gains, losses, and the probability of losing has on how much risk they take. This makes the CCT not only interesting to researchers from the field of psychology, but to those from economics as well, as these parameters relate to the main components of cumulative prospect theory. The benefits of the characteristics of the CCT were illustrated in Chapter 4, where the task was used to study students' RTP in relation to their entrepreneurial orientation. In addition, this Chapter demonstrated how methodological problems specific to the CCT can readily be solved with statistical modelling. Specifically, because the CCT presents participants with a grid of four by eight cards that they can turn, the likelihood of participants selecting multiples of four cards (4, 8, 12, …) is increased. The Censored Mixture Model (Dijkstra et al., 2022) used in Chapter 4 therefore assigns extra probability mass to such outcomes. This model, which was designed specifically for the CCT, can also address censoring – both in versions of the CCT that suffer from this, as well as in the BART. To sum up, while the findings from Part I of the dissertation show that the BART is unsuitable for operationalising risk-taking propensity, they also demonstrate that the CCT has great potential for doing so. However, it is important to be attentive to the idiosyncrasies of behavioural tasks, including the CCT, and to make well-informed and well-motivated decisions about which task to use and in what way.

8.1.2 Part II: Brain Activity

Having evaluated the BART and the CCT, Chapters 5, 6, and 7 focused on the possibilities of operationalising RTP with the use of electroencephalography (EEG). Here, findings were mixed. First, Chapter 5 measured Event-Related Potentials (ERPs) in the EEG signal in response to participants winning or losing in the CCT, as such ERPs are often taken to reflect people's risk-taking propensity and related constructs, such as reward responsiveness. Although the ERPs were clearly identified, correlations between these ERPs on the one hand and self-reports and behavioural measures of similar constructs on the other were non-significant and small. Thus, convergent validity was low, indicating that these ERPs may not be suitable for operationalising risk-taking propensity and related constructs.

Chapter 6 followed up on the findings from Chapter 5 by examining the correlations between self-reported, behavioural, and EEG operationalisations of impulsivity and related concepts (Study 1) and reward responsiveness and related concepts (Study 2). Here, once more, few significant correlations between the EEG measurements and the behavioural and self-report measurements were found. However, two additional findings stood out. First, self-reports and behavioural measures did not correlate well with each other either. Second, self-reports tended to correlate well with other self-reports that were designed to measure comparable constructs, while accounting for common method bias. The same was found for behavioural and EEG measures. Together, these findings provide an alternative explanation for the low convergent validity observed in Chapter 5. Specifically, it opens up the possibility that EEG *can* measure RTP and related constructs, but that it measures a *different dimension or aspect* of it than do self-report measurements or behavioural ones. Support for this conjecture that different (types of) instruments capture different dimensions of a construct is already provided within the existing RTP literature (e.g., Frey et al., 2017), albeit with the important caveat that this research is limited to self-reports and behaviour.

Finally, the findings from Chapter 7 suggest yet another reason why EEG is suitable for operationalising RTP and related constructs. Here, we examined if the statistically significant differences in ERPs between persons with and without addiction, as reported in the existing literature, reflect genuine differences or if they could be explained by selective reporting. Despite most of the examined literature being underpowered, the findings pointed towards genuine effects. In other words, ERPs do differ between clinical and non-clinical populations. This could mean that EEG *is suitable* for operationalising the constructs examined in Chapter 7, including reward responsiveness and impulsivity, but only when differentiating between groups of people on the more extreme ends of a spectrum. It may not be sensitive enough to detect individual differences among healthy persons, especially when benchmarked against self-reports and behavioural measures like in Chapters 5 and 6. Taken together, the findings from Part II cannot unequivocally establish whether EEG is suitable for operationalising RTP, but they do provide pointers as to what it may or may not measure.

8.2 Strengths and Limitations of the Dissertation

Across the dissertation, the strengths and limitations specific to each chapter were discussed. The present section highlights two general strengths and two general limitations.

A *primary strength of this dissertation* is the fact that it takes a step back from empirical enquiries and looks at the broader picture of measurement. Too often, when planning a study and deciding on the operationalisation and (type of) measurement, researchers rely on convenience and conventions. From a personal perspective, back when I was planning data collection for this dissertation, I did so as well. The BART was readily available in the task repository of our lab, and was widely used both by other researchers at the lab and by the broader research community, giving it a legitimacy that I initially did not question. Only through using the task (Chapters 5 and 6) did I recognise its shortcomings, driving me to investigate alternatives like the "warm" (Chapter 4) and "hot" (Chapter 5) versions of the CCT. In the end, this resulted in a dissertation that not only thoroughly reflects on methodological choices, but also provides applications in which the relevant trade-offs between different operationalisations have been carefully taken into account. Likewise focusing both on theory and practice, *the second strength of the dissertation* is that it provides a critical perspective on the importance of sample size (Chapters 6 and 7) and that it "practices what it preaches" by consistently using well-powered samples of at least n = 120. Especially in EEG research, where data collection is time-consuming and expensive compared to research based on self-reports or behavioural tasks, samples are typically 20-40 participants per group, which is insufficient for detecting anything other than large effects (Clayson et al., 2019; Kissel & Friedman, 2023). Because of the large samples, one can be more confident in the estimates provided across the present dissertation.

As for *general limitations, first*, it is important to note the scope of the dissertation. Specifically, while various ERPs were examined to determine their suitability for operationalising RTP and associated constructs, with regard to behavioural tasks the dissertation was heavily concentrated around the BART and CCT. This choice was deliberate: the BART is the most popular instrument for measuring RTP and related constructs in psychology (C. Wang et al., 2025) and is often used as a benchmark for evaluating new tasks. The CCT is in many ways similar to the BART, but circumvents its limitations, and has rapidly gained popularity since its publication 15 years ago. Nonetheless, other risk-taking propensity tasks would also have been worth evaluating, including the Cambridge Gambling Task (Rogers et al., 1999) and the Game of Dice Task (Brand et al., 2005) that were mentioned in Chapter 2. Likewise related to scope, *a second drawback* of the dissertation is that it is mainly concerned with validity, and does not address reliability. This lack of attention to reliability is a shortcoming of the existing empirical literature as well: in studies using self-report scales, estimates of reliability are frequently reported, whereas this is much less so the case for scores obtained with behavioural tasks (Green et al., 2016) and EEG or other (neuro)physiological measures, arguably because

estimating reliability for such scores is complicated (Ellis et al., 2024; Green et al., 2016). The impact of low reliability, however, is large, as it determines the upper limit of the validity: if a measurement is not reliable, by definition it cannot correlate strongly with a criterion due to the high measurement error. According to Hedge et al. (2018), this is precisely why data from behavioural tasks correlates poorly with that from other measures, such as self-reports, although these authors do not examine tasks that measure risk-taking propensity. Research on RTP specifically, however, shows moderate to good reliabilities both for behavioural (Buelow & Barnhart, 2018; White et al., 2008) and EEG measures (Jin et al., 2025; Marco-Pallares et al., 2011), making low reliability an unlikely cause of the small between-measure correlations observed in this dissertation. Further supporting this conjecture, reliabilities of the here presented behavioural and EEG data were computed, showing (Guttman split-half, or in case of unequal numbers of observations theorem 2 by Ellis et al., 2024) reliabilities in the >.80 range for the behavioural data, and reliabilities in the >.50 range for the EEG data.

8.3 Moving Forward: What Should the Next Decade Bring?

In addition to the recommendations for future research made in each individual chapter, the findings from the studies and the general trends that were identified together point to three ambitions that should receive priority in the upcoming years.

First, it is critical that the field of psychology attains a (renewed) appreciation of basic research. Over the past decade, especially in the social sciences, a clear shift towards emphasising the societal impact of research has been observed (Dotti & Walczyk, 2022). Increasingly, policymakers, funding agencies, and universities expect or demand that researchers' academic output has clear positive implications for societal challenges like the UN Sustainable Development Goals (no poverty, reduced inequality, …). In part, this shift represents progress: historically, emphasis was placed solely on the quality and intellectual significance of research, while the broader societal implications of science were neglected. Acknowledging the impact research has outside of academia is therefore justified (Lauronen, 2022). However, the importance that is now placed on societal impact comes at the risk of devaluing basic research (Laing et al., 2017; Penfield et al., 2014). Therefore, it is vital to recognise the importance of basic research for applied research. After all, the value of the conclusions drawn in any empirical study depends on by how well the constructs central to that study are measured. In other words, thorough basic research is a prerequisite for impactful applied research. To give an example: if your measure of risk-taking propensity is

not valid, then your conclusion as to whether persons with ADHD have a lot of it cannot be valid either, and neither can the intervention that you develop based on this conclusion.

Second, basic research on operationalisation and on the characteristics of measurement instruments should be conducted with the user in mind. That is, applied researchers in search of information about the validity and reliability of behavioural or (neuro)physiological measures should be able to interpret this research without specialist knowledge of psychometrics – just like information on the validity and reliability of self-reports is usually straightforward to interpret. This user-friendliness is especially important for studies presenting solutions to problems, which should be supported by working examples and code packages to bridge the gap between method and application. Relatedly, basic researchers should consider publishing (some of) their work in applied journals, rather than focusing purely on specialised method or psychometric journals (like *Behaviour Research Methods* and *Psychometrika*), to more effectively reach their intended users. Frequently, basic research on behavioural or (neuro)physiological measures fails to gain momentum among the users of these instruments. For example, the study by Young and McCoy (2019) that was referred to in Chapter 3 makes excellent points on the limitations of the BART and provides solutions, but citations show that only a handful of studies implement them. Similarly, Huizenga et al. (2002) proposes an elegant method to reduce standard errors in EEG data, yet is only cited by other statistic papers. Hence, for basic research to benefit applied research, stronger links between them are needed.

Finally, it is key that applied researchers in the social sciences actively seek out information about the instruments they intend to use. This is standard practice when using self-reports, but is less common for behavioural and (neuro)physiological measures. The findings presented in this dissertation show the risks this brings: if what you measure is not what you intended to measure, the research question cannot be answered and you may arrive at incorrect conclusions. Hence, when planning a study, one should consider different operationalisations, and carefully evaluate the characteristics, qualities, and fit of possible instruments. Importantly, researchers should not rely on conventions, nor on which instrument is most cited. What worked in one study is not necessarily suitable for another that aims to answer a different question. Furthermore, new research may have been published in the meantime, offering new insights. Thus, each individual study requires a well-informed and well-motivated decision about which measurement to use. Chapter 1 quoted Landis and Cortina (2015, p. 15): “the fact that we must define and measure unobservable constructs adds an element of difficulty to our sciences that does not exist in others”. Unfortunately, we sometimes seem to forget this difficulty exists. It is therefore time we renew our appreciation of this challenge.

References

Abdellaoui, M., Baillon, A., Placido, L., & Wakker, P. (2011). The rich domain of uncertainty: Source functions and their experimental implementation. *The American Economic Review, 101*(2), 695–723. https://doi.org/10.1257/aer.101.2.695

Abdellaoui, M., Vossmann, F., & Weber, M. (2005). Choice-based elicitation and decomposition of decision weights for gains and losses under uncertainty. *Management Science, 51*(9), 1384–1399. https://doi.org/10.1287/mnsc.1050.0388

Ait Oumeziane, B., & Foti, D. (2016). Reward-related neural dysfunction across depression and impulsivity: A dimensional approach. *Psychophysiology, 53*(8), 1174–1184. https://doi.org/10.1111/psyp.12672

Alsharif, A. H., & Mohd Isa, S. (2024). Revolutionizing consumer insights: The impact of fMRI in neuromarketing research. *Future Business Journal, 10*, Article 79. https://doi.org/10.1186/s43093-024-00371-z

Anderson, L. R., & Mellor, J. M. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of Health Economics, 27*(5), 1260–1274. https://doi.org/10.1016/j.jhealeco.2008.05.011

Angner, E. (2012). *A course in behavioral economics*. Palgrave Macmillan Ltd.

Anokhin, A. P., Vedeniapin, A. B., Sirevaag, E. J., Bauer, L. O., O'connor, S. J., Kuperman, S., Porjesz, B., Reich, T., Begleiter, H., & Polich, J. (2000). The P300 brain potential is reduced in smokers. *Psychopharmacology, 149*, 409–413. https://doi.org/10.1007/s002130000387

Arvanitakis, Z. (2010). Update on frontotemporal dementia. *The Neurologist, 16*(1), 16–22. https://doi.org/10.1097/NRL.0b013e3181b1d5c6

Asmaro, D., Carolan, P. L., & Liotti, M. (2014). Electrophysiological evidence of early attentional bias to drug-related pictures in chronic cannabis users. *Addictive Behaviors, 39*(1), 114–121. https://doi.org/10.1016/j.addbeh.2013.09.012

Åstebro, T., Herz, H., Nanda, R., & Weber, R. A. (2014). Seeking the roots of entrepreneurship: Insights from behavioral economics. *Journal of Economic Perspectives, 28*(3), 49–70. https://doi.org/10.1257/jep.28.3.49

Bach, D. R., Seymour, B., & Dolan, R. J. (2009). Neural activity associated with the passive prediction of ambiguity and risk for aversive events. *The Journal of Neuroscience, 29*(6), 1648–1656. https://doi.org/10.1523/JNEUROSCI.4578-08.2009

Baillon, A. (2015). *Subjective truths*. Inaugural Speech, Rotterdam, the Netherlands.

Baillon, A., Bleichrodt, H., Keskin, U., L'Haridon, O., & Li, C. (2013). Learning under ambiguity: An experiment using initial public offerings on a stock market. *Economics Working Paper Archive, Center for Research in Economics and Management (CREM)*

Baillon, A., Bleichrodt, H., Keskin, U., L'Haridon, O., & Li, C. (2017). The effect of learning on ambiguity attitudes. *Management Science, 64*(5), 2181–2198. https://doi.org/10.1287/mnsc.2016.2700

Baillon, A., Cabantous, L., & Wakker, P. (2012). Aggregating imprecise or conflicting beliefs: An experimental investigation using modern ambiguity theories. *Journal of Risk and Uncertainty, 44*, 115–147. https://doi.org/10.1007/s11166-012-9140-x

Bakker, M., & Wicherts, J. M. (2011). The (mis) reporting of statistical results in psychology journals. *Behavior Research Methods, 43*, 666–678. https://doi.org/10.3758/s13428-011-0089-5

Balaz, V., & Williams, A. M. (2011). Risk attitudes and migration experience. *Journal of Risk Research, 14*(5), 583–596. https://doi.org/10.1080/13669877.2010.547256

Bauer, L. O. (2001). CNS recovery from cocaine, cocaine and alcohol, or opioid dependence: A P300 study. *Clinical Neurophysiology, 112*(8), 1508–1515. https://doi.org/10.1016/S1388-2457(01)00583-1

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*(1-3), 7–15. https://doi.org/10.1016/0010-0277(94)90018-3

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex, 10*(3), 295–307. https://doi.org/10.1093/cercor/10.3.295

Becker, M. P. I., Nitsch, A. M., Miltner, W. H. R., & Straube, T. (2014). A single-trial estimation of the feedback-related negativity and its relation to BOLD responses in a time-estimation task. *Journal of Neuroscience, 34*(8), 3005–3012. https://doi.org/10.1523/JNEUROSCI.3684-13.2014

Bekker, E. M., Kenemans, J. L., & Verbaten, M. N. (2005). Source analysis of the N2 in a cued Go/NoGo task. *Cognitive Brain Research, 22*(2), 221–231. https://doi.org/10.1016/j.cogbrainres.2004.08.011

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience, 8*(6), 551–565. https://doi.org/10.1162/jocn.1996.8.6.551

Bernoster, I., De Groot, K., Wieser, M. J., Thurik, R., & Franken, I. H. A. (2019). Birds of a feather flock together: Evidence of prominent correlations within but not between self-

report, behavioral, and electrophysiological measures of impulsivity. *Biological Psychology, 145*, 112–123. https://doi.org/10.1016/j.biopsycho.2019.04.008

Bernoster, I., Rietveld, C. A., Thurik, A. R., & Torrès, O. (2018). Overconfidence, optimism and entrepreneurship. *Sustainability, 10*(7), Article 2233. https://doi.org/10.3390/su10072233

Berridge, K. C., & Robinson, T. E. (2016). Liking, wanting, and the incentive-sensitization theory of addiction. *American Psychologist, 71*(8), 670–679. https://doi.org/10.1037/amp0000059

Bishara, A. J., Pleskac, T. J., Fridberg, D. J., Yechiam, E., Lucas, J., Busemeyer, J. R., Finn, P. R., & Stout, J. C. (2009). Similar processes despite divergent behavior in two commonly used measures of risky decision making. *Journal of Behavioral Decision Making, 22*(4), 435–454. https://doi.org/10.1002/bdm.641

Blais, A., & Weber, E. U. (2006). A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. *Judgment and Decision Making, 1*(1), 33–47. https://doi.org/10.1017/S1930297500000334

Blum, K., Cull, J. G., Braverman, E. R., & Comings, D. E. (1996). Reward deficiency syndrome. *American Scientist, 84*(2), 132–145.

Bohan, R., Lewis, B., Garcia, C., & Nixon, S. J. (2020). Emotional facial expression perception and processing in AUD subgroups: A behavioral and electroencephalographic investigation of polysubstance use. *UF Journal of Undergraduate Research, 22.* https://doi.org/10.32473/ufjur.v22i0.120248

Bolton, D. L. (2012). Individual entrepreneurial orientation: Further investigation of a measurement instrument. *Academy of Entrepreneurship Journal, 18*(1), 91–98.

Bolton, D. L., & Lane, M. D. (2012). Individual entrepreneurial orientation: Development of a measurement instrument. *Education + Training, 54*(2), 219–233. https://doi.org/10.1108/00400911211210314

Bonatti, E., Kuchukhidze, G., Zamarian, L., Trinka, E., Bodner, T., Benke, T., & Delazer, M. (2009). Decision making in ambiguous and risky situations after unilateral temporal lobe epilepsy surgery. *Epilepsy & Behavior, 14*(4), 665–673. https://doi.org/10.1016/j.yebeh.2009.02.015

Borawska, A. (2017). Cognitive neuroscience tools in economic experiments investigating the decision making process. *Folia Oeconomica Stetinensia, 17*(1), 159–169.

Bötzel, K., & Grüsser, O. J. (1989). Electric brain potentials evoked by pictures of faces and non-faces: A search for "face-specific" EEG-potentials. *Experimental Brain Research, 77*, 349–360. https://doi.org/10.1007/BF00274992

Boutros, N. N., Reid, M. C., Petrakis, I., Campbell, D., Torello, M., & Krystal, J. (2000). Similarities in the disturbances in cortical information processing in alcoholism and aging: A pilot evoked potential study. *International Psychogeriatrics, 12*(4), 513–525. https://doi.org/10.1017/S1041610200006621

Brand, M., Fujiwara, E., Borsutzky, S., Kalbe, E., Kessler, J., & Markowitsch, H. J. (2005). Decision-making deficits of Korsakoff patients in a new gambling task with explicit rules: Associations with executive functions. *Neuropsychology, 19*(3), 267–277. https://doi.org/0.1037/0894-4105.19.3.267

Brand, M., Labudda, K., & Markowitsch, H. J. (2006). Neuropsychological correlates of decision-making in ambiguous and risky situations. *Neural Networks, 19*(8), 1266–1276. https://doi.org/10.1016/j.neunet.2006.03.001

Brand, M., Recknor, E. C., Grabenhorst, F., & Bechara, A. (2007). Decisions under ambiguity and decisions under risk: Correlations with executive functions and comparisons of two different gambling tasks with implicit and explicit rules. *Journal of Clinical and Experimental Neuropsychology, 29*(1), 86–99. https://doi.org/10.1080/13803390500507196

Bress, J. N., & Hajcak, G. (2013). Self-report and behavioral measures of reward sensitivity predict the feedback negativity. *Psychophysiology, 50*(7), 610–616. https://doi.org/10.1111/psyp.12053

Brevers, D., Cleeremans, A., Goudriaan, A. E., Bechara, A., Kornreich, C., Verbanck, P., & Noël, X. (2012). Decision making under ambiguity but not under risk is related to problem gambling severity. *Psychiatry Research, 200*(2-3), 568–574. https://doi.org/10.1016/j.psychres.2012.03.053

Brunell, A. B., & Buelow, M. T. (2017). Narcissism and performance on behavioral decision-making tasks. *Journal of Behavioral Decision Making, 30*(1), 3–14. https://doi.org/10.1002/bdm.1900

Brunner, J., & Schimmack, U. (2020). Estimating population mean power under conditions of heterogeneity and selection for significance. *Meta-Psychology, 4*, Article 2018.874. https://doi.org/10.15626/MP.2018.874

Buckert, M., Schwieren, C., Kudielka, B. M., & Fiebach, C. J. (2014). Acute stress affects risk taking but not ambiguity aversion. *Frontiers in Neuroscience, 8*, Article 82. https://doi.org/10.3389/fnins.2014.00082

Buelow, M. T. (2015). Predicting performance on the Columbia Card Task: Effects of personality characteristics, mood, and executive functions. *Assessment, 22*(2), 178–187. https://doi.org/10.1177/1073191114539383

Buelow, M. T., & Barnhart, W. R. (2018). Test-retest reliability of common behavioral decision making tasks. *Archives of Clinical Neuropsychology, 33*(1), 125–129. https://doi.org/10.1093/arclin/acx038

Buelow, M. T., & Blaine, A. L. (2015). The assessment of risky decision making: A factor analysis of performance on the Iowa Gambling Task, Balloon Analogue Risk Task, and Columbia Card Task. *Psychological Assessment, 27*(3), 777–785. https://doi.org/10.1037/a0038622

Buelow, M. T., & Suhr, J. A. (2009). Construct validity of the Iowa Gambling Task. *Neuropsychology Review, 19*, 102–114. https://doi.org/10.1007/s11065-009-9083-4

Burns, E. J., Arnold, T., & Bukach, C. M. (2019). P-curving the fusiform face area: Meta-analyses support the expertise hypothesis. *Neuroscience & Biobehavioral Reviews, 104*, 209–221. https://doi.org/10.1016/j.neubiorev.2019.07.003

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara Gambling Task. *Psychological Assessment, 14*(3), 253–262. https://doi.org/10.1037/1040-3590.14.3.253

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376. https://doi.org/10.1038/nrn3475

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin, 125*(3), 367–383. https://doi.org/10.1037/0033-2909.125.3.367

Carbine, K. A., Lindsey, H. M., Rodeback, R. E., & Larson, M. J. (2019). Quantifying evidential value and selective reporting in recent and 10-year past psychophysiological literature: A pre-registered P-curve analysis. *International Journal of Psychophysiology, 142*, 33–49. https://doi.org/10.1016/j.ijpsycho.2019.06.004

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science, 280*(5364), 747–749. https://doi.org/10.1126/science.280.5364.747

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science, 2*(2), 115–144. https://doi.org/10.1177/2515245919847196

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. https://doi.org/10.1037/0022-3514.67.2.319

Cassidy, C. M., Brodeur, M. B., Lepage, M., & Malla, A. (2014). Do reward-processing deficits in schizophrenia-spectrum disorders promote cannabis use? An investigation of physiological response to natural rewards and drug cues. *Journal of Psychiatry & Neuroscience: JPN, 39*(5), 339–347. https://doi.org/10.1503/jpn.130207

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. https://doi.org/10.1017/S0033291717002185

Chandrakumar, D., Feuerriegel, D., Bode, S., Grech, M., & Keage, H. A. D. (2018). Event-related potentials in relation to risk-taking: A systematic review. *Frontiers in Behavioral Neuroscience, 12*, Article 111. https://doi.org/10.3389/fnbeh.2018.00111

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. https://doi.org/10.1016/j.parkreldis.2011.02.018

Chen, H., Jiang, H., Guo, Q., Du, J., Wang, J., & Zhao, M. (2013). Case-control study of error-related negativity among males with heroin dependence undergoing rehabilitation. *Shanghai Archives of Psychiatry, 25*(3), 141–148. https://doi.org/10.3969/j.issn.1002-0829.2013.03.003

Chen, J., Liang, Y., Mai, C., Zhong, X., & Qu, C. (2016). General deficit in inhibitory control of excessive smartphone users: Evidence from an event-related potential study. *Frontiers in Psychology, 7*, Article 511. https://doi.org/10.3389/fpsyg.2016.00511

Chen, X., Zhu, C., Li, J., Qiu, L., Zhang, L., Yu, F., Ye, R., Zhang, J., & Wang, K. (2013). Dissociation of decision making under ambiguity and decision making under risk in breast cancer patients receiving adjuvant chemotherapy: A neuropsychological study. *Brain Research, 1533*, 63–72. https://doi.org/10.1016/j.brainres.2013.08.015

Chen, Y., Yu, H., & Gao, X. (2022). Influences of emotional information on response inhibition in gaming disorder: Behavioral and ERP evidence from Go/NoGo task. *International Journal of Environmental Research and Public Health, 19*(23), Article 16264. https://doi.org/10.3390/ijerph192316264

Cheng, J., Guan, Y., Zhang, Y., Bi, Y., Bu, L., Li, Y., Shi, S., Liu, P., Lu, X., & Yu, D. (2016). Electrophysiological mechanisms of biased response to smoking-related cues in young

smokers. *Neuroscience Letters, 629*, 85–91. https://doi.org/10.1016/j.neulet.2016.06.062

Clayson, P. E., Baldwin, S. A., Rocha, H. A., & Larson, M. J. (2021). The data-processing multiverse of event-related potentials (ERPs): A roadmap for the optimization and standardization of ERP processing and reduction pipelines. *NeuroImage, 245*, Article 118712. https://doi.org/10.1016/j.neuroimage.2021.118712

Clayson, P. E., Carbine, K. A., Baldwin, S. A., & Larson, M. J. (2019). Methodological reporting behavior, sample sizes, and statistical power in studies of event-related potentials: Barriers to reproducibility and replicability. *Psychophysiology, 56*(11), Article e13437. https://doi.org/10.1111/psyp.13437

Clayson, P. E., Carbine, K. A., & Larson, M. J. (2020). A registered report of error-related negativity and reward positivity as biomarkers of depression: P-curving the evidence. *International Journal of Psychophysiology, 150*, 50–72. https://doi.org/10.1016/j.ijpsycho.2020.01.005

Cohen, H. L., Ji, J., Chorlian, D. B., Begleiter, H., & Porjesz, B. (2002). Alcohol-related ERP changes recorded from different modalities: A topographic analysis. *Alcoholism: Clinical and Experimental Research, 26*(3), 303–317. https://doi.org/10.1111/j.1530-0277.2002.tb02539.x

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45*(12), 1304–1312. https://doi.org/10.1037/0003-066X.45.12.1304

Collado, A., Felton, J. W., MacPherson, L., & Lejuez, C. W. (2014). Longitudinal trajectories of sensation seeking, risk taking propensity, and impulsivity across early to middle adolescence. *Addictive Behaviors, 39*(11), 1580–1588. https://doi.org/10.1016/j.addbeh.2014.01.024

Colrain, I. M., Sullivan, E. V., Ford, J. M., Mathalon, D. H., McPherson, S., Roach, B. J., Crowley, K. E., & Pfefferbaum, A. (2011). Frontally mediated inhibitory processing and white matter microstructure: Age and alcoholism effects. *Psychopharmacology, 213*, 669–679. https://doi.org/10.1007/s00213-010-2073-7

Corneille, O., & Gawronski, B. (2024). Self-reports are better measurement instruments than implicit measures. *Nature Reviews Psychology, 3*, 835–846. https://doi.org/10.1038/s44159-024-00376-z

Courchesne, E., Hillyard, S. A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and Clinical Neurophysiology, 39*(2), 131–143. https://doi.org/10.1016/0013-4694(75)90003-6

Craig, J. B., Pohjola, M., Kraus, S., & Jensen, S. H. (2014). Exploring relationships among proactiveness, risk-taking and innovation output in family and non-family firms. *Creativity and Innovation Management, 23*(2), 199–210. https://doi.org/10.1111/caim.12052

Criado, J. R., & Ehlers, C. L. (2007). Electrophysiological responses to affective stimuli in southwest California Indians: Relationship to alcohol dependence. *Journal of Studies on Alcohol and Drugs, 68*(6), 813–823. https://doi.org/10.15288/jsad.2007.68.813

Crosetto, P., & Filippin, A. (2013). The "bomb" risk elicitation task. *Journal of Risk and Uncertainty, 47*, 31–65. https://doi.org/10.1007/s11166-013-9170-z

Crowley, M. J., Wu, J., Crutcher, C., Bailey, C. A., Lejuez, C. W., & Mayes, L. C. (2009). Risk-taking and the feedback negativity response to loss among at-risk adolescents. *Developmental Neuroscience, 31*(1-2), 137–148. https://doi.org/10.1159/000207501

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective picture processing: Covariation with autonomic arousal and affective report. *Biological Psychology, 52*(2), 95–111. https://doi.org/10.1016/S0301-0511(99)00044-7

Dai, L., Maksimov, V., Gilbert, B. A., & Fernhaber, S. A. (2014). Entrepreneurial orientation and international scope: The differential roles of innovativeness, proactiveness, and risk-taking. *Journal of Business Venturing, 29*(4), 511–524. https://doi.org/10.1016/j.jbusvent.2013.07.004

Danso, A., Adomako, S., Damoah, J. O., & Uddin, M. (2016). Risk-taking propensity, managerial network ties and firm performance in an emerging economy. *The Journal of Entrepreneurship, 25*(2), 155–183. https://doi.org/10.1177/0971355716650367

Davies, P. L., Segalowitz, S. J., & Gavin, W. J. (2004). Development of error-monitoring event-related potentials in adolescents. *Annals of the New York Academy of Sciences, 1021*(1), 324–328. https://doi.org/10.1196/annals.1308.039

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. https://doi.org/10.1016/j.addbeh.2004.06.004

De Groot, K. (2020). Burst beliefs – methodological problems in the Balloon Analogue Risk Task and implications for its use. *Journal of Trial and Error, 1*(1), 43–51. https://doi.org/10.36850/mr1

De Groot, K., & Thurik, R. (2018). Disentangling risk and uncertainty: When risk-taking measures are not about risk. *Frontiers in Psychology, 9*, Article 2194. https://doi.org/10.3389/fpsyg.2018.02194

De Groot, K., & Van Strien, J. W. (2019). Event-related potentials in response to feedback following risk-taking in the hot version of the Columbia Card Task. *Psychophysiology, 56*(9), Article e13390. https://doi.org/10.1111/psyp.13390

De Pascalis, V., Varriale, V., & D'Antuono, L. (2010). Event-related components of the punishment and reward sensitivity. *Clinical Neurophysiology, 121*(1), 60–76. https://doi.org/10.1016/j.clinph.2009.10.004

Deweese, M. M., Codispoti, M., Robinson, J. D., Cinciripini, P. M., & Versace, F. (2018). Cigarette cues capture attention of smokers and never-smokers, but for different reasons. *Drug and Alcohol Dependence, 185*, 50–57. https://doi.org/10.1016/j.drugalcdep.2017.12.010

Di Stefano, G., & Gutierrez, C. (2019). Under a magnifying glass: On the use of experiments in strategy research. *Strategic Organization, 17*(4), 497–507. https://doi.org/10.1177/1476127018803840

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research, 61*(12), 1203–1218. https://doi.org/10.1016/j.jbusres.2008.01.009

Dijkstra, N. F. S., De Groot, K., & Rietveld, C. A. (2023). Entrepreneurial orientation and decision-making under risk and uncertainty: Experimental evidence from the Columbia Card Task. *Applied Psychology, 72*(4), 1577–1592. https://doi.org/10.1111/apps.12436

Dijkstra, N. F. S., Tiemeier, H., Figner, B. C., & Groenen, P. J. F. (2020). *A censored mixture model for modeling risk taking*. ArXiv. https://doi.org/10.48550/arXiv.2002.08146

Dijkstra, N. F. S., Tiemeier, H., Figner, B., & Groenen, P. J. F. (2022). A censored mixture model for modeling risk taking. *Psychometrika, 87*(3), 1103–1129. https://doi.org/10.1007/s11336-021-09839-1

Ding, Y., Wang, E., Zou, Y., Song, Y., Xiao, X., Huang, W., & Li, Y. (2017). Gender differences in reward and punishment for monetary and social feedback in children: An ERP study. *PloS One, 12*, Article e0174100. https://doi.org/10.1371/journal.pone.0174100

Dittmar, O., Krehl, R., & Lautenbacher, S. (2011). Interrelation of self-report, behavioural and electrophysiological measures assessing pain-related information processing. *Pain Research and Management, 16*(1), 33–40. https://doi.org/10.1155/2011/675785

Donchin, E. (1981). Surprise!… Surprise? *Psychophysiology, 18*(5), 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Donders, F. C. (1969). On the speed of mental processes (W. G. Koster, Trans.). *Acta Psychologica, 30*, 412–431. (Original work published 1869)

Dong, G., Lu, Q., Zhou, H., & Zhao, X. (2010). Impulse inhibition in people with internet addiction disorder: Electrophysiological evidence from a Go/NoGo study. *Neuroscience Letters, 485*(2), 138–142. https://doi.org/10.1016/j.neulet.2010.09.002

Donkers, F. C. L., & Van Boxtel, G.,J.M. (2004). The N2 in Go/No-Go tasks reflects conflict monitoring not response inhibition. *Brain and Cognition, 56*(2), 165–176. https://doi.org/10.1016/j.bandc.2004.04.005

Dotti, N. F., & Walczyk, J. (2022). What is the societal impact of university research? A policy-oriented review to map approaches, identify monitoring methods and success factors. *Evaluation and Program Planning, 95*, Article 102157. https://doi.org/10.1016/j.evalprogplan.2022.102157

Dunning, J. P., & Hajcak, G. (2007). Error-related negativities elicited by monetary loss and cues that predict loss. *Neuroreport, 18*(17), 1875–1878. https://doi.org/10.1097/WNR.0b013e3282f0d50b

Dunning, J. P., Parvaz, M. A., Hajcak, G., Maloney, T., Alia-Klein, N., Woicik, P. A., Telang, F., Wang, G., Volkow, N. D., & Goldstein, R. Z. (2011). Motivated attention to cocaine and emotional cues in abstinent and current cocaine users – an ERP study. *European Journal of Neuroscience, 33*(9), 1716–1723. https://doi.org/10.1111/j.1460-9568.2011.07663.x

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Duven, E. C. P., Müller, K. W., Beutel, M. E., & Wölfling, K. (2015). Altered reward processing in pathological computer gamers – ERP-results from a semi-natural gaming-design. *Brain and Behavior, 5*(1), Article e00293. https://doi.org/10.1002/brb3.293

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51*(4), 380–417. https://doi.org/10.1037/h0053870

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Ellis, J. L., Sijtsma, K., De Groot, K., & Groenen, P. J. F. (2024). Reliability theory for measurements with variable test length, illustrated with ERN and Pe collected in the

Flanker Task. *Psychometrika, 89*(4), 1280–1303. https://doi.org/10.1007/s11336-024-09982-5

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics, 75*(4), 643–669. https://doi.org/10.2307/1884324

Endrass, T., Schuermann, B., Roepke, S., Kessler-Scheil, S., & Kathmann, N. (2016). Reduced risk avoidance and altered neural correlates of feedback processing in patients with borderline personality disorder. *Psychiatry Research, 243*, 14–22. https://doi.org/10.1016/j.psychres.2016.06.016

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a non-search task. *Perception & Psychophysics, 16*(1), 143–149. https://doi.org/10.3758/bf03203267

Euser, A. S., Arends, L. R., Evans, B. E., Greaves-Lord, K., Huizink, A. C., & Franken, I. H. A. (2012). The P300 event-related brain potential as a neurobiological endophenotype for substance use disorders: A meta-analytic investigation. *Neuroscience & Biobehavioral Reviews, 36*(1), 572–603. https://doi.org/10.1016/j.neubiorev.2011.09.002

Euser, A. S., Evans, B. E., Greaves-Lord, K., Huizink, A. C., & Franken, I. H. A. (2013). Parental rearing behavior prospectively predicts adolescents' risky decision-making and feedback-related electrical brain activity. *Developmental Science, 16*(3), 409–427. https://doi.org/10.1111/desc.12026

Euser, A. S., Greaves-Lord, K., Crowley, M. J., Evans, B. E., Huizink, A. C., & Franken, I. H. A. (2013). Blunted feedback processing during risky decision making in adolescents with a parental history of substance use disorders. *Development and Psychopathology, 25*(4), 1119–1136. https://doi.org/10.1017/S0954579413000412

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*, 111–125. https://doi.org/10.1007/s00213-011-2264-x

Euteneuer, F., Schaefer, F., Stuermer, R., Boucsein, W., Timmermann, L., Barbe, M. T., Ebersbach, G., Otto, J. ö, Kessler, J., & Kalbe, E. (2009). Dissociation of decision-making under ambiguity and decision-making under risk in patients with Parkinson's disease: A neuropsychological and psychophysiological study. *Neuropsychologia, 47*(13), 2882–2890. https://doi.org/10.1016/j.neuropsychologia.2009.06.014

Eysenck, M. W. (1992). *Anxiety: The cognitive perspective*. Psychology Press.

Eysenck, S. B., Syed, I. A., & Eysenck, H. J. (1966). Desirability response set in children. *British Journal of Educational Psychology, 36*(1), 87–90. https://doi.org/10.1111/j.2044-8279.1966.tb01843.x

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology, 78*(6), 447–455. https://doi.org/10.1016/0013-4694(91)90062-9

Fathi, M., Mazhari, S., Pourrahimi, A. M., Poormohammad, A., & Sardari, S. (2022). Proactive and reactive inhibitory control are differently affected by video game addiction: An event-related potential study. *Brain and Behavior, 12*(6), Article e2584. https://doi.org/10.1002/brb3.2584

Fathi, M., Pourrahimi, A. M., Poormohammad, A., Sardari, S., Rajizadeh, M. A., Mazhari, S., & Pourkand, D. (2024). Video game addiction is associated with early stage of inhibitory control problems: An event-related potential study using cued Go/NoGo task. *Addiction Biology, 29*(4), Article e13391. https://doi.org/10.1111/adb.13391

Fein, G., & Andrew, C. (2011). Event-related potentials during visual target detection in treatment-naïve active alcoholics. *Alcoholism: Clinical and Experimental Research, 35*(6), 1171–1179. https://doi.org/10.1111/j.1530-0277.2011.01450.x

Fein, G., & Chang, M. (2008). Smaller feedback ERN amplitudes during the BART are associated with a greater family history density of alcohol problems in treatment-naive alcoholics. *Drug and Alcohol Dependence, 92*(1-3), 141–148. https://doi.org/10.1016/j.drugalcdep.2007.07.017

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. https://doi.org/10.1177/1745691612459059

Feuerriegel, D., Churches, O., Hofmann, J., & Keage, H. A. D. (2015). The N170 and face perception in psychiatric and neurological disorders: A systematic review. *Clinical Neurophysiology, 126*(6), 1141–1158. https://doi.org/10.1016/j.clinph.2014.09.015

Field, A. P. (2013). *Discovering statistics using SPSS: And sex and drugs and rock 'n' roll* (4th ed.). Sage publications.

Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: Age differences in risk taking in the Columbia Card Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(3), 709–730. https://doi.org/10.1037/a0014983

Forstmeier, W., Wagenmakers, E., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings – a practical guide. *Biological Reviews, 92*(4), 1941–1968. https://doi.org/10.1111/brv.12315

Foti, D., Hajcak, G., & Dien, J. (2009). Differentiating neural responses to emotional pictures: Evidence from temporal-spatial PCA. *Psychophysiology, 46*(3), 521–530. https://doi.org/10.1111/j.1469-8986.2009.00796.x

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. https://doi.org/10.1016/j.paid.2005.11.016

Franken, I. H. A., Nijs, I. M. T., Toes, A., & Van der Veen, F. M. (2018). Food addiction is associated with impaired performance monitoring. *Biological Psychology, 131*, 49–53. https://doi.org/10.1016/j.biopsycho.2016.07.005

Franken, I. H. A., Stam, C. J., Hendriks, V. M., & Van den Brink, W. (2003). Neurophysiological evidence for abnormal cognitive processing of drug cues in heroin dependence. *Psychopharmacology, 170*, 205–212. https://doi.org/10.1007/s00213-003-1542-7

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. https://doi.org/10.1111/j.1530-0277.2009.01139.x

Franken, I. H. A., Van Strien, J. W., Franzek, E. J., & Van de Wetering, B. J. (2007). Error-processing deficits in patients with cocaine dependence. *Biological Psychology, 75*(1), 45–51. https://doi.org/10.1016/j.biopsycho.2006.11.003

Franken, I. H. A., Van Strien, J. W., & Kuijpers, I. (2010). Evidence for a deficit in the salience attribution to errors in smokers. *Drug and Alcohol Dependence, 106*(2-3), 181–185. https://doi.org/10.1016/j.drugalcdep.2009.08.014

Frey, R., Pedroni, A., Mata, R., Rieskamp, J., & Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. *Science Advances, 3*(10), Article e1701381. https://doi.org/10.1126/sciadv.1701381

Funke, G., Greenlee, E., Carter, M., Dukes, A., Brown, R., & Menke, L. (2016). Which eye tracker is right for your research? Performance evaluation of several cost variant eye trackers. *Proceedings of the Human Factors and Ergonomics Society, 60*(1) 1240–1244. https://doi.org/10.1177/1541931213601289

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. https://doi.org/10.1016/j.ijpsycho.2015.11.015

Geburek, A. J., Rist, F., Gediga, G., 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. https://doi.org/10.1016/j.ijpsycho.2012.08.006

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4*(6), 385–390. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science, 295*(5563), 2279–2282. https://doi.org/10.1126/science.1066893

Gelman, A., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time*. Open Science Framework. https://osf.io/n3axs/download

Gerbing, 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. https://doi.org/10.1207/s15327906mbr2203_6

Ghin, F., Beste, C., & Stock, A. (2022). On the role of stimulus-response context in inhibitory control in alcohol use disorder. *Journal of Clinical Medicine, 11*(21), Article 6557. https://doi.org/10.3390/jcm11216557

Glenn, S. W., Parsons, O. A., & Smith, L. T. (1996). ERP responses to target and nontarget visual stimuli in alcoholics from VA and community treatment programs. *Alcohol, 13*(1), 85–92. https://doi.org/10.1016/0741-8329(95)02018-7

Goldstein, R. Z., & Volkow, N. D. (2011). Dysfunction of the prefrontal cortex in addiction: Neuroimaging findings and clinical implications. *Nature Reviews Neuroscience, 12*, 652–669. https://doi.org/10.1038/nrn3119

Gooding, D. C., Burroughs, S., & Boutros, N. N. (2008). Attentional deficits in cocaine-dependent patients: Converging behavioral and electrophysiological evidence. *Psychiatry Research, 160*(2), 145–154. https://doi.org/10.1016/j.psychres.2007.11.019

Grable, J., & Lytton, R. H. (1999). Financial risk tolerance revisited: The development of a risk assessment instrument. *Financial Services Review, 8*(3), 163–181. https://doi.org/10.1016/S1057-0810(99)00041-4

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. https://doi.org/10.1016/0013-4694(83)90135-9

Gray, H. M., Ambady, N., Lowenthal, W. T., & Deldin, P. (2004). P300 as an index of attention to self-relevant stimuli. *Journal of Experimental Social Psychology, 40*(2), 216–224. https://doi.org/10.1016/S0022-1031(03)00092-1

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

Green, S. B., Yang, Y., Alt, M., Brinkley, S., Gray, S., Hogan, T., & Cowan, N. (2016). Use of internal consistency coefficients for estimating reliability of experimental task scores. *Psychonomic Bulletin & Review, 23*, 750–763. https://doi.org/10.3758/s13423-015-0968-3

Groen, Y., Wijers, A. A., Mulder, L. J. M., 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 Neurophysiology, 119*(11), 2476–2493. https://doi.org/10.1016/j.clinph.2008.08.004

Gu, R., Zhang, D., Luo, Y., Wang, H., & Broster, L. S. (2018). Predicting risk decisions in a modified Balloon Analogue Risk Task: Conventional and single-trial ERP analyses. *Cognitive, Affective, & Behavioral Neuroscience, 18*, 99–116. https://doi.org/10.3758/s13415-017-0555-3

Gunessee, S., & Lane, T. (2023). Changing perceptions about experimentation in economics: 50 years of evidence from principles textbooks. *Journal of Behavioral and Experimental Economics, 107*, Article 102086. https://doi.org/10.1016/j.socec.2023.102086

Guney, F., Genc, B. O., Kutlu, R., & Ilhan, B. C. (2009). Auditory P300 event-related potential in tobacco smokers. *Journal of Clinical Neuroscience, 16*(10), 1311–1315. https://doi.org/10.1016/j.jocn.2008.11.025

Hada, M., Porjesz, B., Begleiter, H., & Polich, J. (2000). Auditory P3a assessment of male alcoholics. *Biological Psychiatry, 48*(4), 276–286. https://doi.org/10.1016/S0006-3223(00)00236-5

Hair, J. F., Anderson, R. E., Babin, B. J., & Black, W. C. (2010). *Multivariate data analysis: A global perspective* (7th ed.). Pearson Education.

Hajcak Proudfit, G. (2015). The reward positivity: From basic research on reward to a biomarker for depression. *Psychophysiology, 52*(4), 449–459. https://doi.org/10.1111/psyp.12370

Hajcak, G., Dunning, J. P., & Foti, D. (2007). Neural response to emotional pictures is unaffected by concurrent task difficulty: An event-related potential study. *Behavioral Neuroscience, 121*(6), 1156–1162. https://doi.org/10.1037/0735-7044.121.6.1156

Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: An integrative review. *Developmental Neuropsychology, 35*(2), 129–155. https://doi.org/10.1080/87565640903526504

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2006). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biological Psychology, 71*(2), 148–154. https://doi.org/10.1016/j.biopsycho.2005.04.001

Haslam, N., Vylomova, E., Murphy, S. C., & Wilson, S. J. (2022). The neuroscientification of psychology: The rising prevalence of neuroscientific concepts in psychology from 1965 to 2016. *Perspectives on Psychological Science, 17*(2), 519–529. https://doi.org/10.1177/1745691621991864

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description–experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making, 21*(5), 493–518. https://doi.org/10.1002/bdm.598

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*, 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Hedges, L. V. (1984). Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics, 9*(1), 61–85. https://doi.org/10.3102/10769986009001061

Heinze, M., Wölfling, K., & Grüsser, S. M. (2007). Cue-induced auditory evoked potentials in alcoholism. *Clinical Neurophysiology, 118*(4), 856–862. https://doi.org/10.1016/j.clinph.2006.12.003

Henry, E. A., Kaye, J. T., Bryan, A. D., Hutchison, K. E., & Ito, T. A. (2014). Cannabis cue reactivity and craving among never, infrequent and heavy cannabis users. *Neuropsychopharmacology, 39*, 1214–1221. https://doi.org/10.1038/npp.2013.324

Herrmann, M. J., Saathoff, C., Schreppel, T. J., Ehlis, A., 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. https://doi.org/10.1016/j.psychres.2008.06.015

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. https://doi.org/10.1111/psyp.12671

Hinojosa, J. A., Mercado, F., & Carretié, L. (2015). N170 sensitivity to facial expression: A meta-analysis. *Neuroscience & Biobehavioral Reviews, 55*, 498–509. https://doi.org/10.1016/j.neubiorev.2015.06.002

Hirayasu, Y., Samura, M., Ohta, H., & Ogura, C. (2000). Sex effects on rate of change of P300 latency with age. *Clinical Neurophysiology, 111*(2), 187–194. https://doi.org/10.1016/S1388-2457(99)00233-3

Hoffman, L. A., Lewis, B., & Nixon, S. J. (2019). Neurophysiological and interpersonal correlates of emotional face processing in alcohol use disorder. *Alcoholism: Clinical and Experimental Research, 43*(9), 1928–1936. https://doi.org/10.1111/acer.14152

Holguin, S. R., Porjesz, B., Chorlian, D. B., Polich, J., & Begleiter, H. (1999). Visual P3a in male alcoholics and controls. *Alcoholism: Clinical and Experimental Research, 23*(4), 582–591. https://doi.org/10.1111/j.1530-0277.1999.tb04158.x

Holm, H. J., Opper, S., & Nee, V. (2013). Entrepreneurs under uncertainty: An economic experiment in China. *Management Science, 59*(7), 1671–1687. https://doi.org/10.1287/mnsc.1120.1670

Holper, L., & Murphy, R. O. (2014). Hemodynamic and affective correlates assessed during performance on the Columbia Card Task (CCT). *Brain Imaging and Behavior, 8*, 517–530. https://doi.org/10.1007/s11682-013-9265-9

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109*(4), 679–709. https://doi.org/10.1037//0033-295X.109.4.679

Holroyd, C. B., Pakzad-Vaezi, K. L., & Krigolson, O. E. (2008). The feedback correct-related positivity: Sensitivity of the event-related brain potential to unexpected positive feedback. *Psychophysiology, 45*(5), 688–697. https://doi.org/10.1111/j.1469-8986.2008.00668.x

Hosseini-Kamkar, N., Lowe, C., & Morton, J. B. (2021). The differential calibration of the HPA axis as a function of trauma versus adversity: A systematic review and p-curve meta-analyses. *Neuroscience & Biobehavioral Reviews, 127*, 54–135. https://doi.org/10.1016/j.neubiorev.2021.04.006

Houston, R. J., & Schlienz, N. J. (2018). Event-related potentials as biomarkers of behavior change mechanisms in substance use disorder treatment. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3*(1), 30–40. https://doi.org/10.1016/j.bpsc.2017.09.006

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. https://doi.org/10.1016/S0191-8869(01)00032-0

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., & Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science, 310*(5754), 1680–1683. https://doi.org/10.1126/science.1115327

Huang, Y., Wood, S., Berger, D. E., & Hanoch, Y. (2013). Risky choice in younger versus older adults: Affective context matters. *Judgment and Decision Making, 8*(2), 179–187. https://doi.org/10.1017/S1930297500005106

Huang, Y., Wood, S., Berger, D. E., & Hanoch, Y. (2015). Age differences in experiential and deliberative processes in unambiguous and ambiguous decision making. *Psychology and Aging, 30*(3), 675–687. https://doi.org/10.1037/pag0000038

Huettel, S. A., Stowe, C. J., Gordon, E. M., Warner, B. T., & Platt, M. L. (2006). Neural signatures of economic preferences for risk and ambiguity. *Neuron, 49*(5), 765–775. https://doi.org/10.1016/j.neuron.2006.01.024

Huizenga, H. M., Heslenfeld, D. J., & Molenaar, P. C. M. (2002). Optimal measurement conditions for spatiotemporal EEG/MEG source analysis. *Psychometrika, 67*(2), 299–313. https://doi.org/10.1007/BF02294848

Huster, R. J., Westerhausen, R., Pantev, C., & Konrad, C. (2010). The role of the cingulate cortex as neural generator of the N200 and P300 in a tactile response inhibition task. *Human Brain Mapping, 31*(8), 1260–1271. https://doi.org/10.1002/hbm.20933

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry, 167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), Article e124. https://doi.org/10.1371/journal.pmed.1004085

Jaeger, D. A., Dohmen, T., Falk, A., Huffman, D., Sunde, U., & Bonin, H. (2010). Direct evidence on risk attitudes and migration. *The Review of Economics and Statistics, 92*(3), 684–689. https://doi.org/10.1162/REST_a_00020

Jeffreys, D. A. (1989). A face-responsive potential recorded from the human scalp. *Experimental Brain Research, 78*, 193–202. https://doi.org/10.1007/BF00230699

Jin, J., Xiao, Q., Liu, Y., Xu, T., & Shen, Q. (2025). Test-retest reliability of decisions under risk with outcome evaluation: Evidence from behavioral and event-related potentials (ERPs) measures in 2 monetary gambling tasks. *Cerebral Cortex, 35*(3), Article bhaf058. https://doi.org/10.1093/cercor/bhaf058

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. https://doi.org/10.1037/1064-1297.13.3.259

Jones, K. A., Porjesz, B., Chorlian, D., Rangaswamy, M., Kamarajan, C., Padmanabhapillai, A., Stimus, A., & Begleiter, H. (2006). S-transform time-frequency analysis of P300 reveals deficits in individuals diagnosed with alcoholism. *Clinical Neurophysiology, 117*(10), 2128–2143. https://doi.org/10.1016/j.clinph.2006.02.028

Jonkman, L. M., van Melis, J.,J.M., 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. https://doi.org/10.1016/j.biopsycho.2007.08.004

Josef, A. K., Richter, D., Samanez-Larkin, G., Wagner, G. G., Hertwig, R., & Mata, R. (2016). Stability and change in risk-taking propensity across the adult life span. *Journal of Personality and Social Psychology, 111*(3), 430–450. https://doi.org/10.1037/pspp0000090

Joyce, C., & Rossion, B. (2005). The face-sensitive N170 and VPP components manifest the same brain processes: The effect of reference electrode site. *Clinical Neurophysiology, 116*(11), 2613–2631. https://doi.org/10.1016/j.clinph.2005.07.005

Kahn, B. E., & Sarin, R. K. (1988). Modeling ambiguity in decisions under uncertainty. *Journal of Consumer Research, 15*(2), 265–272. https://doi.org/10.1086/209163

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–291. https://doi.org/10.2307/1914185

Kalkhoven, C., Sennef, C., Peeters, A., & Van Den Bos, R. (2014). Risk-taking and pathological gambling behavior in Huntington's disease. *Frontiers in Behavioral Neuroscience, 8*, Article 103. https://doi.org/10.3389/fnbeh.2014.00103

Kam, J. W. Y., 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. https://doi.org/10.1016/j.ijpsycho.2012.05.006

Kamarajan, C., Rangaswamy, M., Tang, Y., Chorlian, D. B., Pandey, A. K., Roopesh, B. N., Manz, N., Saunders, R., Stimus, A. T., & Porjesz, B. (2010). Dysfunctional reward processing in male alcoholics: An ERP study during a gambling task. *Journal of Psychiatric Research, 44*(9), 576–590. https://doi.org/10.1016/j.jpsychires.2009.11.019

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*, Article 1368. https://doi.org/10.3389/fpsyg.2014.01368

Kardos, Z., Kóbor, A., Takács, Á, Tóth, B., Boha, R., File, B., & Molnár, M. (2016). Age-related characteristics of risky decision-making and progressive expectation formation. *Behavioural Brain Research, 312*, 405–414. https://doi.org/10.1016/j.bbr.2016.07.003

Keil, A., Bradley, M. M., Hauk, O., Rockstroh, B., Elbert, T., & Lang, P. J. (2002). Large-scale neural correlates of affective picture processing. *Psychophysiology, 39*(5), 641–649. https://doi.org/10.1017.S0048577202394162

Kessler, L., Hewig, J., Weichold, K., Silbereisen, R. K., & Miltner, W. H. R. (2017). Feedback negativity and decision-making behavior in the Balloon Analogue Risk Task (BART) in adolescents is modulated by peer presence. *Psychophysiology, 54*(2), 260–269. https://doi.org/10.1111/psyp.12783

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., Howes, M. J., Jin, R., Secnik, K., & Spencer, T. (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. https://doi.org/10.1017/S0033291704002892

Khadka, S., Stevens, M. C., Aslanzadeh, F., Narayanan, B., Hawkins, K. A., Austad, C. S., Raskin, S. A., Tennen, H., Wood, R. M., & Fallahi, C. (2017). Composite impulsivity-related domains in college students. *Journal of Psychiatric Research, 90*, 118–125. https://doi.org/10.1016/j.jpsychires.2017.02.016

Kiefer, M., Marzinzik, F., Weisbrod, M., Scherg, M., & Spitzer, M. (1998). The time course of brain activations during response inhibition: Evidence from event-related potentials in a Go/No Go task. *Neuroreport, 9*(4), 765–770. https://doi.org/10.1097/00001756-199803090-00037

Kilka, M., & Weber, M. (2001). What determines the shape of the probability weighting function under uncertainty? *Management Science, 47*(12), 1712–1726. https://doi.org/10.1287/mnsc.47.12.1712.10239

Kilner, J. M. (2013). Bias in a common EEG and MEG statistical analysis and how to avoid it. *Clinical Neurophysiology, 10*(124), 2062–2063. https://doi.org/10.1016/j.clinph.2013.03.024

Kim, B., Lee, J., Choi, A. R., Chung, S. J., Park, M., Koo, J. W., Kang, U. G., & Choi, J. (2021). Event-related brain response to visual cues in individuals with internet gaming

disorder: Relevance to attentional bias and decision-making. *Translational Psychiatry, 11*, Article 258. https://doi.org/10.1038/s41398-021-01375-x

Kim, H. W., Kang, J. I., Namkoong, K., Jhung, K., Ha, R. Y., & Kim, S. J. (2015). Further evidence of a dissociation between decision-making under ambiguity and decision-making under risk in obsessive-compulsive disorder. *Journal of Affective Disorders, 176*, 118–124. https://doi.org/10.1016/j.jad.2015.01.060

Kissel, H. A., & Friedman, B. H. (2023). Participant diversity in psychophysiology. *Psychophysiology, 60*(11), Article e14369. https://doi.org/10.1111/psyp.14369

Kleih, S. C., Nijboer, F., Halder, S., & Kübler, A. (2010). Motivation modulates the P300 amplitude during brain-computer interface use. *Clinical Neurophysiology, 121*, 1023–1031. https://doi.org/10.1016/j.clinph.2010.01.034

Kluwe-Schiavon, B., Sanvicente-Vieira, B., Viola, T. W., Veiga, E., Bortolotto, V., & Grassi-Oliveira, R. (2015). Assessing affective and deliberative decision-making: Adaptation of the Columbia Card Task to Brazilian Portuguese. *The Spanish Journal of Psychology, 18*, Article e89. https://doi.org/10.1017/sjp.2015.81

Kluwe-Schiavon, B., Viola, T. W., Sanvicente-Vieira, B., Pezzi, J. C., & Grassi-Oliveira, R. (2016). Similarities between adult female crack cocaine users and adolescents in risky decision-making scenarios. *Journal of Clinical and Experimental Neuropsychology, 38*(7), 795–810. https://doi.org/10.1080/13803395.2016.1167171

Knight, F. H. (1921). *Risk, uncertainty and profit*. Sentry Press.

Knight, R. T. (1984). Decreased response to novel stimuli after prefrontal lesions in man. *Electroencephalography and Clinical Neurophysiology, Evoked Potentials Section, 59*(1), 9–20. https://doi.org/10.1016/0168-5597(84)90016-9

Kóbor, A., Takács, Á, Janacsek, K., Németh, D., Honbolygó, F., & Csépe, V. (2015). Different strategies underlying uncertain decision making: Higher executive performance is associated with enhanced feedback-related negativity. *Psychophysiology, 52*(3), 367–377. https://doi.org/10.1111/psyp.12331

Koe, W. (2016). The relationship between individual entrepreneurial orientation (IEO) and entrepreneurial intention. *Journal of Global Entrepreneurship Research, 6*, 1–11. https://doi.org/10.1186/s40497-016-0057-8

Koenig, S., & Mecklinger, A. (2008). Electrophysiological correlates of encoding and retrieving emotional events. *Emotion, 8*(2), 162–173. https://doi.org/10.1037/1528-3542.8.2.162

Koob, G. F., Buck, C. L., Cohen, A., Edwards, S., Park, P. E., Schlosburg, J. E., Schmeichel, B., Vendruscolo, L. F., Wade, C. L., & Whitfield Jr, T. W. (2014). Addiction as a stress

surfeit disorder. *Neuropharmacology, 76*, 370–382. https://doi.org/10.1016/j.neuropharm.2013.05.024

Koob, G. F., & Le Moal, M. (2001). Drug addiction, dysregulation of reward, and allostasis. *Neuropsychopharmacology, 24*(2), 97–129. https://doi.org/10.1016/S0893-133X(00)00195-0

Kopp, B., Mattler, U., Goertz, R., & Rist, F. (1996). N2, P3 and the lateralized readiness potential in a NoGo task involving selective response priming. *Electroencephalography and Clinical Neurophysiology, 99*(1), 19–27. https://doi.org/10.1016/0921-884X(96)95617-9

Kopp, B., & Wolff, M. (2000). Brain mechanisms of selective learning: Event-related potentials provide evidence for error-driven learning in humans. *Biological Psychology, 51*(2-3), 223–246. https://doi.org/10.1016/S0301-0511(99)00039-3

Koudstaal, M., Sloof, R., & Van Praag, M. (2016). Risk, uncertainty, and entrepreneurship: Evidence from a lab-in-the-field experiment. *Management Science, 62*(10), 2897–2915. https://doi.org/10.1287/mnsc.2015.2249

Krain, A. L., Wilson, A. M., Arbuckle, R., Castellanos, F. X., & Milham, M. P. (2006). Distinct neural mechanisms of risk and ambiguity: A meta-analysis of decision-making. *NeuroImage, 32*(1), 477–484. https://doi.org/10.1016/j.neuroimage.2006.02.047

Kreiser, P. M., & Davis, J. (2010). Entrepreneurial orientation and firm performance: The unique impact of innovativeness, proactiveness, and risk-taking. *Journal of Small Business & Entrepreneurship, 23*(1), 39–51. https://doi.org/10.1080/08276331.2010.10593472

Krigolson, O. E., Williams, C. C., Norton, A., Hassall, C. D., & Colino, F. L. (2017). Choosing MUSE: Validation of a low-cost, portable EEG system for ERP research. *Frontiers in Neuroscience, 11*, Article 109. https://doi.org/10.3389/fnins.2017.00109

Kunasegaran, K., Ismail, A. M. H., Ramasamy, S., Gnanou, J. V., Caszo, B. A., & Chen, P. L. (2023). Understanding mental fatigue and its detection: A comparative analysis of assessments and tools. *PeerJ, 11*, Article e15744. https://doi.org/10.7717/peerj.15744

Kuntz, A., Missonnier, P., Prévot, A., Favre, G., Herrmann, F. R., Debatisse, D., Merlo, M. C. G., & Gothuey, I. (2021). Persistence of neuronal alterations in alcohol-dependent patients at conclusion of the gold standard withdrawal treatment: Evidence from ERPs. *Frontiers in Psychiatry, 12*, Article 666063. https://doi.org/10.3389/fpsyt.2021.666063

Kwako, L. E., Bickel, W. K., & Goldman, D. (2018). Addiction biomarkers: Dimensional approaches to understanding addiction. *Trends in Molecular Medicine, 24*(2), 121–128. https://doi.org/10.1016/j.molmed.2017.12.007

Kwako, L. E., Momenan, R., Litten, R. Z., Koob, G. F., & Goldman, D. (2016). Addictions neuroclinical assessment: A neuroscience-based framework for addictive disorders. *Biological Psychiatry, 80*(3), 179–189. https://doi.org/10.1016/j.biopsych.2015.10.024

Laing, K., Mazzoli Smith, L., & Todd, L. (2018). The impact agenda and critical social research in education: Hitting the target but missing the spot? *Policy Futures in Education, 16*(2), 169–184. https://doi.org/10.1177/1478210317742214

Lakens, D., & Evers, E. R. K. (2014). Sailing from the seas of chaos into the corridor of stability: Practical recommendations to increase the informational value of studies. *Perspectives on Psychological Science, 9*(3), 278–292. https://doi.org/10.1177/1745691614528520

Landis, R. S., & Cortina, J. M. (2014). Is ours a hard science (and do we care)? In C. E. Lance & R. J. Vandenberg (Eds.), *More statistical and methodological myths and urban legends* (pp. 9–35). Routledge.

Lansbergen, M. M., Böcker, K.,B.E., Bekker, E. M., & Kenemans, J. L. (2007). Neural correlates of stopping and self-reported impulsivity. *Clinical Neurophysiology, 118*(9), 2089–2103. https://doi.org/10.1016/j.clinph.2007.06.011

Larson, M. J., & Carbine, K. A. (2017). Sample size calculations in human electrophysiology (EEG and ERP) studies: A systematic review and recommendations for increased rigor. *International Journal of Psychophysiology, 111*, 33–41. https://doi.org/10.1016/j.ijpsycho.2016.06.015

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. https://doi.org/10.1002/bdm.1784

Lauriola, M., & Weller, J. (2018). Personality and risk: Beyond daredevils – risk taking from a temperament perspective. In M. Raue, E. Lermer & B. Streicher (Eds.), *Psychological perspectives on risk and risk analysis* (pp. 3–36). Springer. https://doi.org/10.1007/978-3-319-92478-6_1

Lauronen, J. (2022). Tension in interpretations of the social impact of the social sciences: Walking a tightrope between divergent conceptualizations of research utilization.

SAGE Open, 12(2), Article 21582440221089967. https://doi.org/10.1177/21582440221089967

Lawyer, S. R. (2013). Risk taking for sexual versus monetary outcomes using the Balloon Analogue Risk Task. *The Psychological Record, 63*, 803–820. https://doi.org/10.11133/j.tpr.2013.63.4.007

Lechner, C., & Gudmundsson, S. V. (2014). Entrepreneurial orientation, firm strategy and small firm performance. *International Small Business Journal, 32*(1), 36–60. https://doi.org/10.1177/0266242612455034

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 behaviours. *Journal of Adolescence, 26*(4), 475–479. https://doi.org/10.1016/S0140-1971(03)00036-8

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & 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. https://doi.org/10.1037/1076-898X.8.2.75

Levy, I., Snell, J., Nelson, A. J., Rustichini, A., & Glimcher, P. W. (2010). Neural representation of subjective value under risk and ambiguity. *Journal of Neurophysiology, 103*(2), 1036–1047. https://doi.org/10.1152/jn.00853.2009

Li, C., Pan, T., He, J., Zheng, Y., Fan, L., & Meng, Y. (2024). Approach bias in individuals with internet gaming disorder: Evidence from an event-related potential-based approach-avoid task. *International Journal of Psychophysiology, 202*, Article 112376. https://doi.org/10.1016/j.ijpsycho.2024.112376

Li, J., & Yao, H. (2023). Aberrant reward anticipating and processing in abstinent heroin addicts. *IEEE Transactions on Computational Social Systems, 11*(2), 2234–2244. https://doi.org/10.1109/TCSS.2023.3290969

Limbrick-Oldfield, E., Van Holst, R. J., & Clark, L. (2013). Fronto-striatal dysregulation in drug addiction and pathological gambling: Consistent inconsistencies? *NeuroImage: Clinical, 2*, 385–393. https://doi.org/10.1016/j.nicl.2013.02.005

Linden, D. E. J. (2005). The P300: Where in the brain is it produced and what does it tell us? *The Neuroscientist, 11*(6), 563–576. https://doi.org/10.1177/1073858405280524

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. https://doi.org/10.1016/S0010-9452(08)70274-0

Littel, M., Euser, A. S., Munafò, M. R., & Franken, I. H. A. (2012). Electrophysiological indices of biased cognitive processing of substance-related cues: A meta-analysis. *Neuroscience & Biobehavioral Reviews, 36*(8), 1803–1816. https://doi.org/10.1016/j.neubiorev.2012.05.001

Littel, M., & Franken, I. H. A. (2007). The effects of prolonged abstinence on the processing of smoking cues: An ERP study among smokers, ex-smokers and never-smokers. *Journal of Psychopharmacology, 21*(8), 873–882. https://doi.org/10.1177/0269881107078494

Littel, M., & Franken, I. H. A. (2012). Electrophysiological correlates of associative learning in smokers: A higher-order conditioning experiment. *BMC Neuroscience, 13*, 1-13. https://doi.org/10.1186/1471-2202-13-8

Littel, M., Van den Berg, I., Luijten, M., van 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. https://doi.org/10.1111/j.1369-1600.2012.00467.x

Liu, X., Zhou, H., Jiang, C., Xue, Y., Zhou, Z., & Wang, J. (2020). Cognitive control deficits in alcohol dependence are a trait-and state-dependent biomarker: An ERP study. *Frontiers in Psychiatry, 11*, Article 606891. https://doi.org/10.3389/fpsyt.2020.606891

Liu, Y., Masina, F., Ridderinkhof, K. R., & Pezzetta, R. (2023). Addiction as a brain disease? A meta-regression comparison of error-related brain potentials between addiction and neurological diseases. *Neuroscience & Biobehavioral Reviews, 148*, Article 105127. https://doi.org/10.1016/j.neubiorev.2023.105127

Logan, G. D., Schachar, R. J., & Tannock, R. (1997). Impulsivity and inhibitory control. *Psychological Science, 8*(1), 60–64. https://doi.org/10.1111/j.1467-9280.1997.tb00545.x

Lole, L., Gonsalvez, C. J., & Barry, R. J. (2015). Reward and punishment hyposensitivity in problem gamblers: A study of event-related potentials using a principal components analysis. *Clinical Neurophysiology, 126*(7), 1295–1309. https://doi.org/10.1016/j.clinph.2014.10.011

Lopez, R., Dauvilliers, Y., Jaussent, I., Billieux, J., & Bayard, S. (2015). A multidimensional approach of impulsivity in adult attention deficit hyperactivity disorder. *Psychiatry Research, 227*(2-3), 290–295. https://doi.org/10.1016/j.psychres.2015.03.023

Lubman, D. I., Allen, N. B., Peters, L. A., & Deakin, J. W. (2007). Electrophysiological evidence of the motivational salience of drug cues in opiate addiction. *Psychological Medicine, 37*(8), 1203–1209. https://doi.org/10.1017/S0033291707009932

Lubman, D. I., Allen, N. B., Peters, L. A., & Deakin, J. W. (2008). Electrophysiological evidence that drug cues have greater salience than other affective stimuli in opiate addiction. *Journal of Psychopharmacology, 22*(8), 836–842. https://doi.org/10.1177/0269881107083846

Lubman, D. I., Yücel, M., Kettle, J. W. L., Scaffidi, A., MacKenzie, T., Simmons, J. G., & Allen, N. B. (2009). Responsiveness to drug cues and natural rewards in opiate addiction: Associations with later heroin use. *Archives of General Psychiatry, 66*(2), 205–212. https://doi.org/10.1001/archgenpsychiatry.2008.522

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology, 54*(1), 146–157. https://doi.org/10.1111/psyp.12639

Luijten, M., Machielsen, M. W. J., Veltman, D. J., Hester, R., De Haan, L., & Franken, I. H. A. (2014). Systematic review of ERP and fMRI studies investigating inhibitory control and error processing in people with substance dependence and behavioural addictions. *Journal of Psychiatry and Neuroscience, 39*(3), 149–169. https://doi.org/10.1503/jpn.130052

Luijten, M., Van Meel, C. S., & Franken, I. H. A. (2011). Diminished error processing in smokers during smoking cue exposure. *Pharmacology Biochemistry and Behavior, 97*(3), 514–520. https://doi.org/10.1016/j.pbb.2010.10.012

Lumpkin, G. T., Cogliser, C. C., & Schneider, D. R. (2009). Understanding and measuring autonomy: An entrepreneurial orientation perspective. *Entrepreneurship Theory and Practice, 33*(1), 47–69. https://doi.org/10.1111/j.1540-6520.2008.00280.x

Lumpkin, G. T., & Dess, G. G. (1996). Clarifying the entrepreneurial orientation construct and linking it to performance. *Academy of Management Review, 21*(1), 135–172. https://doi.org/10.5465/amr.1996.9602161568

Lutz, M. C., Kok, R., Verveer, I., Malbec, M., Koot, S., Van Lier, P.,A.C., & Franken, I. H. A. (2021). Diminished error-related negativity and error positivity in children and adults with externalizing problems and disorders: A meta-analysis on error processing. *Journal of Psychiatry and Neuroscience, 46*(6), E615–E627. https://doi.org/10.1503/jpn.200031

Maia, T. V., & McClelland, J. L. (2004). A re-examination of the evidence for the somatic marker hypothesis: What participants really know in the Iowa Gambling Task. *PNAS, 101*(45), 16075–16080. https://doi.org/10.1073/pnas.0406666101

Maij, D. L. R., Van de Wetering, B.,J.M., & Franken, I. H. A. (2017). Cognitive control in young adults with cannabis use disorder: An event-related brain potential study.

Journal of Psychopharmacology, 31(8), 1015–1026. https://doi.org/10.1177/0269881117719262

Mapelli, D., Di Rosa, E., Cavalletti, M., Schiff, S., & Tamburin, S. (2014). Decision and dopaminergic system: An ERPs study of Iowa Gambling Task in Parkinson's disease. *Frontiers in Psychology, 5*, Article 684. https://doi.org/10.3389/fpsyg.2014.00684

Marco-Pallares, J., Cucurell, D., Münte, T. F., Strien, N., & Rodriguez-Fornells, A. (2011). On the number of trials needed for a stable feedback-related negativity. *Psychophysiology, 48*(6), 852–860. https://doi.org/10.1111/j.1469-8986.2010.01152.x

Marhe, R., Van de Wetering, B.,J.M., & Franken, I. H. A. (2013). Error-related brain activity predicts cocaine use after treatment at 3-month follow-up. *Biological Psychiatry, 73*(8), 782–788. https://doi.org/10.1016/j.biopsych.2012.12.016

Markiewicz, Ł, & Kubińska, E. (2015). Information use differences in hot and cold risk processing: When does information about probability count in the Columbia Card Task? *Frontiers in Psychology, 6*, Article 1727. https://doi.org/10.3389/fpsyg.2015.01727

Marks-Anglin, A., & Chen, Y. (2020). A historical review of publication bias. *Research Synthesis Methods, 11*(6), 725–742. https://doi.org/10.1002/jrsm.1452

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*, 314–324. https://doi.org/10.3758/s13428-011-0168-7

Maurage, P., Campanella, S., Philippot, P., De Timary, P., Constant, E., Gauthier, S., Miccichè, M., Kornreich, C., Hanak, C., & Noel, X. (2008). Alcoholism leads to early perceptive alterations, independently of comorbid depressed state: An ERP study. *Clinical Neurophysiology, 38*(2), 83–97. https://doi.org/10.1016/j.neucli.2008.02.001

Maurage, P., Philippot, P., Joassin, F., Pauwels, L., Pham, T., Prieto, E. A., Palmero-Soler, E., Zanow, F., & Campanella, S. (2008). The auditory-visual integration of anger is impaired in alcoholism: An event-related potentials study. *Journal of Psychiatry and Neuroscience, 33*(2), 111–122.

Maurage, P., Philippot, P., Verbanck, P., Noel, X., Kornreich, C., Hanak, C., & Campanella, S. (2007). Is the P300 deficit in alcoholism associated with early visual impairments (P100, N170)? An oddball paradigm. *Clinical Neurophysiology, 118*(3), 633–644. https://doi.org/10.1016/j.clinph.2006.11.007

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes.

Perspectives on Psychological Science, 11(5), 730–749. https://doi.org/10.1177/1745691616662243

Meda, S. A., Stevens, M. C., Potenza, M. N., Pittman, B., Gueorguieva, R., Andrews, M. M., Thomas, A. D., Muska, C., Hylton, J. L., & Pearlson, G. D. (2009). Investigating the behavioral and self-report constructs of impulsivity domains using principal component analysis. *Behavioural Pharmacology, 20*(5-6), 390–399. https://doi.org/10.1097/FBP.0b013e32833113a3

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. https://doi.org/10.1111/psyp.12664

Miedl, S. F., Fehr, T., Herrmann, M., & Meyer, G. (2014). Risk assessment and reward processing in problem gambling investigated by event-related potentials and fMRI-constrained source analysis. *BMC Psychiatry, 14*, 1–11. https://doi.org/10.1186/s12888-014-0229-4

Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (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. https://doi.org/10.1162/jocn.1997.9.6.788

Minnix, J. A., Versace, F., Robinson, J. D., Lam, C. Y., Engelmann, J. M., Cui, Y., Brown, V. L., & Cinciripini, P. M. (2013). The late positive potential (LPP) in response to varying types of emotional and cigarette stimuli in smokers: A content comparison. *International Journal of Psychophysiology, 89*(1), 18–25. https://doi.org/10.1016/j.ijpsycho.2013.04.019

Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., & Swann, A. C. (2001). Psychiatric aspects of impulsivity. *American Journal of Psychiatry, 158*(11), 1783–1793. https://doi.org/10.1176/appi.ajp.158.11.1783

Moeller, F. G., Barratt, E. S., Fischer, C. J., Dougherty, D. M., Reilly, E. L., Mathias, C. W., & Swann, A. C. (2004). P300 event-related potential amplitude and impulsivity in cocaine-dependent subjects. *Neuropsychobiology, 50*(2), 167–173. https://doi.org/10.1159/000079110

Morie, K. P., De Sanctis, P., Garavan, H., & Foxe, J. J. (2014). Executive dysfunction and reward dysregulation: A high-density electrical mapping study in cocaine abusers. *Neuropharmacology, 85*, 397–407. https://doi.org/10.1016/j.neuropharm.2014.05.016

Morie, K. P., De Sanctis, P., Garavan, H., & Foxe, J. J. (2016). Regulating task-monitoring systems in response to variable reward contingencies and outcomes in cocaine addicts. *Psychopharmacology, 233*, 1105–1118. https://doi.org/10.1007/s00213-015-4191-8

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. https://doi.org/10.1080/15374416.2013.865191

Motlagh, F., Ibrahim, F., Rashid, R., Seghatoleslam, T., & Habil, H. (2017). Investigation of brain electrophysiological properties among heroin addicts: Quantitative EEG and event-related potentials. *Journal of Neuroscience Research, 95*(8), 1633–1646. https://doi.org/10.1002/jnr.23988

Nermend, K., & Łatuszyńska, M. (2016). Selected issues in experimental economics. *Proceedings of the 2015 Computational Methods in Experimental Economics (CMEE) Conference*. https://doi.org/10.1007/978-3-319-28419-4

Nieß, C., & Biemann, T. (2014). The role of risk propensity in predicting self-employment. *Journal of Applied Psychology, 99*(5), 1000–1009. https://doi.org/10.1037/a0035992

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin, 131*(4), 510–532. https://doi.org/10.1037/0033-2909.131.4.510

Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. H. (2004). Reinforcement-related brain potentials from medial frontal cortex: Origins and functional significance. *Neuroscience & Biobehavioral Reviews, 28*(4), 441–448. https://doi.org/10.1016/j.neubiorev.2004.05.003

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P. H., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology, 38*(5), 752–760. https://doi.org/10.1111/1469-8986.3850752

Nieuwenhuis, S., Yeung, N., Van Den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a Go/No-Go task: Effects of response conflict and trial type frequency. *Cognitive, Affective, & Behavioral Neuroscience, 3*(1), 17–26. https://doi.org/10.3758/CABN.3.1.17

Nikolin, S., Chand, N., Martin, D., Rushby, J., Loo, C. K., & Boonstra, T. W. (2022). Little evidence for a reduced late positive potential to unpleasant stimuli in major depressive disorder. *Neuroimage: Reports, 2*(1), Article 100077. https://doi.org/10.1016/j.ynirp.2022.100077

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231

Norberg, M. M., Kavanagh, D. J., Olivier, J., & Lyras, S. (2016). Craving cannabis: A meta-analysis of self-report and psychophysiological cue-reactivity studies. *Addiction, 111*(11), 1923–1934. https://doi.org/10.1111/add.13472

O'Connell, R. G., Dockree, P. M., Bellgrove, M. A., Kelly, S. P., Hester, R., Garavan, H., Robertson, I. H., & Foxe, J. J. (2007). The role of cingulate cortex in the detection of errors with and without awareness: A high-density electrical mapping study. *European Journal of Neuroscience, 25*(8), 2571–2579. https://doi.org/10.1111/j.1460-9568.2007.05477.x

Oberg, S. A. K., Christie, G. J., & Tata, M. S. (2011). Problem gamblers exhibit reward hypersensitivity in medial frontal cortex during gambling. *Neuropsychologia, 49*(13), 3768–3775. https://doi.org/10.1016/j.neuropsychologia.2011.09.037

Olbrich, H. M., Maes, H., Gann, H., Hagenbuch, F., & Feige, B. (2000). Auditory and visual event-related potentials in alcoholics: Abnormalities of components and brain electrical field. *European Archives of Psychiatry and Clinical Neuroscience, 250*, 215–220. https://doi.org/10.1007/s004060070010

Oliver, J. A., & Foulds, J. (2021). Association between cigarette smoking frequency and tobacco use disorder in US adults. *American Journal of Preventive Medicine, 60*(5), 726–728. https://doi.org/10.1016/j.amepre.2020.10.019

Olvet, D. M., & Hajcak, G. (2008). The error-related negativity (ERN) and psychopathology: Toward an endophenotype. *Clinical Psychology Review, 28*(8), 1343–1354. https://doi.org/10.1016/j.cpr.2008.07.003

Op de Macks, Z. A., Bunge, S. A., Bell, O. N., Kriegsfeld, L. J., Kayser, A. S., & Dahl, R. E. (2017). The effect of social rank feedback on risk taking and associated reward processes in adolescent girls. *Social Cognitive and Affective Neuroscience, 12*(2), 240–250. https://doi.org/10.1093/scan/nsw125

Opitz, A., Zimmermann, J., Cole, D. M., Coray, R. C., Zachäi, A., Baumgartner, M. R., Steuer, A. E., Pilhatsch, M., Quednow, B. B., & Beste, C. (2024). Conflict monitoring and emotional processing in 3, 4-methylenedioxymethamphetamine (MDMA) and methamphetamine users – a comparative neurophysiological study. *NeuroImage: Clinical, 41*, Article 103579. https://doi.org/10.1016/j.nicl.2024.103579

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8*(2), 157–159. https://doi.org/10.3102/10769986008002157

Overbeek, T. J. M., Nieuwenhuis, S., & Ridderinkhof, K. R. (2005). Dissociable components of error processing: On the functional significance of the Pe vis-à-vis the ERN/Ne. *Journal of Psychophysiology, 19*(4), 319–329. https://doi.org/10.1027/0269-8803.19.4.319

Padilla, M. L., Colrain, I. M., Sullivan, E. V., Mayer, B. Z., Turlington, S. R., Hoffman, L. R., Wagstaff, A. E., & Pfefferbaum, A. (2011). Electrophysiological evidence of enhanced performance monitoring in recently abstinent alcoholic men. *Psychopharmacology, 213*, 81–91. https://doi.org/10.1007/s00213-010-2018-1

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., & Brennan, S. E. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71. https://doi.org/10.1136/bmj.n71

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge.

Pandey, A. K., Kamarajan, C., Tang, Y., Chorlian, D. B., Roopesh, B. N., Manz, N., Stimus, A., Rangaswamy, M., & Porjesz, B. (2012). Neurocognitive deficits in male alcoholics: An ERP/sLORETA analysis of the N2 component in an equal probability Go/NoGo task. *Biological Psychology, 89*(1), 170–182. https://doi.org/10.1016/j.biopsycho.2011.10.009

Panno, A., Lauriola, M., & Figner, B. (2013). Emotion regulation and risk taking: Predicting risky choice in deliberative decision making. *Cognition & Emotion, 27*(2), 326–334. https://doi.org/10.1080/02699931.2012.707642

Park, M., Choi, J. S., Park, S. M., Lee, J. Y., Jung, H. Y., Sohn, B. K., Kim, S. N., Kim, D. J., & Kwon, J. S. (2016). Dysfunctional information processing during an auditory event-related potential task in individuals with internet gaming disorder. *Translational Psychiatry, 6*, Article e721. https://doi.org/10.1038/tp.2015.215

Park, M., Jang, J. H., Yoo, S. Y., Choi, A., Kim, H., & Choi, J. (2023). Regional brain activity of resting-state fMRI and auditory oddball ERP with multimodal approach in individuals with internet gaming disorder. *Journal of Behavioral Addictions, 12*(4), 895–906. https://doi.org/10.1556/2006.2023.00063

Park, M., Jung, M. H., Lee, J., Choi, A. R., Chung, S. J., Kim, B., Kim, D. J., & Choi, J. (2020). Neurophysiological and cognitive correlates of error processing deficits in internet gaming disorder. *Cerebral Cortex, 30*(9), 4914–4921. https://doi.org/10.1093/cercor/bhaa083

Park, M., Kim, Y. J., & Choi, J. (2017). Sustained dysfunctional information processing in patients with internet gaming disorder: 6-month follow-up ERP study. *Medicine, 96*(36), 1–6. https://doi.org/10.1097/MD.0000000000007995

Parsons, O. A., Sinha, R., & Williams, H. L. (1990). Relationships between neuropsychological test performance and event-related potentials in alcoholic and nonalcoholic samples. *Alcoholism: Clinical and Experimental Research, 14*(5), 746–755. https://doi.org/10.1111/j.1530-0277.1990.tb01239.x

Parvaz, M. A., Moeller, S. J., Malaker, P., Sinha, R., Alia-Klein, N., & Goldstein, R. Z. (2017). Abstinence reverses EEG-indexed attention bias between drug-related and pleasant stimuli in cocaine-addicted individuals. *Journal of Psychiatry and Neuroscience, 42*(2), 78–86. https://doi.org/10.1503/jpn.150358

Pasion, R., & Barbosa, F. (2019). ERN as a transdiagnostic marker of the internalizing-externalizing spectrum: A dissociable meta-analytic effect. *Neuroscience & Biobehavioral Reviews, 103*, 133–149. https://doi.org/10.1016/j.neubiorev.2019.06.013

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt Impulsiveness Scale. *Journal of Clinical Psychology, 51*(6), 768–774. https://doi.org/10.1002/1097-4679(199511)51:6

Pedroni, A., Rieskamp, J., Pachur, T., Frey, R., Westfall, J. E., & Figner, B. (2018). *A prospect theory account of the "hot" Columbia Card Task*. Open Science Framework. https://osf.io/g7zr3/download

Penfield, T., Baker, M. J., Scoble, R., & Wykes, M. C. (2014). Assessment, evaluations, and definitions of research impact: A review. *Research Evaluation, 23*(1), 21–32. https://doi.org/10.1093/reseval/rvt021

Penolazzi, B., Gremigni, P., & Russo, P. M. (2012). Impulsivity and reward sensitivity differentially influence affective and deliberative risky decision making. *Personality and Individual Differences, 53*(5), 655–659. https://doi.org/10.1016/j.paid.2012.05.018

Pfabigan, D. M., Alexopoulos, J., Bauer, H., & Sailer, U. (2011). Manipulation of feedback expectancy and valence induces negative and positive reward prediction error signals manifest in event-related brain potentials. *Psychophysiology, 48*(5), 656–664. https://doi.org/10.1111/j.1469-8986.2010.01136.x

Pfefferbaum, A., Ford, J. M., Weller, B. J., & Kopell, B. S. (1985). ERPs to response production and inhibition. *Electroencephalography and Clinical Neurophysiology, 60*(5), 423–434. https://doi.org/10.1016/0013-4694(85)91017-X

Platt, M. L., & Huettel, S. A. (2008). Risky business: The neuroeconomics of decision making under uncertainty. *Nature Neuroscience, 11*, 398–403. https://doi.org/10.1038/nn2062

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. https://doi.org/10.1037/a0014245

Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Polich, J., & Ochoa, C. J. (2004). Alcoholism risk, tobacco smoking, and P300 event-related potential. *Clinical Neurophysiology, 115*(6), 1374–1383. https://doi.org/10.1016/j.clinph.2004.01.026

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*(1-2), 130–134. https://doi.org/10.1016/j.neulet.2005.12.003

Potts, G. F., Martin, L. E., Burton, P., & Montague, P. R. (2006b). 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. https://doi.org/10.1162/jocn.2006.18.7.1112

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics, 25*(3), 303–325. https://doi.org/10.1016/0304-4076(84)90004-6

Pripfl, J., Neumann, R., Köhler, U., & Lamm, C. (2013). Effects of transcranial direct current stimulation on risky decision making are mediated by ‘hot’ and ‘cold’ decisions, personality, and hemisphere. *European Journal of Neuroscience, 38*(12), 3778–3785. https://doi.org/10.1111/ejn.12375

Rakow, T., & Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making, 23*(1), 1–14. https://doi.org/10.1002/bdm.681

Randall, W. M., & Smith, J. L. (2011). Conflict and inhibition in the cued-Go/NoGo task. *Clinical Neurophysiology, 122*(12), 2400–2407. https://doi.org/10.1016/j.clinph.2011.05.012

Rauch, A., & Frese, M. (2007). Let's put the person back into entrepreneurship research: A meta-analysis on the relationship between business owners' personality traits, business creation, and success. *European Journal of Work and Organizational Psychology, 16*(4), 353–385. https://doi.org/10.1080/13594320701595438

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. https://doi.org/10.1111/psyp.12809

Reynolds, B., Ortengren, A., Richards, J. B., & De Wit, H. (2006). Dimensions of impulsive behavior: Personality and behavioral measures. *Personality and Individual Differences, 40*(2), 305–315. https://doi.org/10.1016/j.paid.2005.03.024

Rietdijk, W. J. R., Franken, I. H. A., & Thurik, A. R. (2014). Internal consistency of event-related potentials associated with cognitive control: N2/P3 and ERN/Pe. *PloS One, 9*(7), Article e102672. https://doi.org/10.1371/journal.pone.0102672

Rizvi, S. J., Pizzagalli, D. A., Sproule, B. A., & Kennedy, S. H. (2016). Assessing anhedonia in depression: Potentials and pitfalls. *Neuroscience & Biobehavioral Reviews, 65*, 21–35. https://doi.org/10.1016/j.neubiorev.2016.03.004

Robinson, J. D., Versace, F., Engelmann, J. M., Cui, Y., Slapin, A., Oum, R., & Cinciripini, P. M. (2015). The motivational salience of cigarette-related stimuli among former, never, and current smokers. *Experimental and Clinical Psychopharmacology, 23*(1), 37–48. https://doi.org/10.1037/a0038467

Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain Research Reviews, 18*(3), 247–291. https://doi.org/10.1016/0165-0173(93)90013-P

Rodd, J. M. (2024). Moving experimental psychology online: How to obtain high quality data when we can't see our participants. *Journal of Memory and Language, 134*, Article 104472. https://doi.org/10.1016/j.jml.2023.104472

Rogers, R. D., Owen, A. M., Middleton, H. C., Williams, E. J., Pickard, J. D., Sahakian, B. J., & Robbins, T. W. (1999). Choosing between small, likely rewards and large, unlikely rewards activates inferior and orbital prefrontal cortex. *The Journal of Neuroscience, 19*(20), 9029–9038. https://doi.org/10.1523/JNEUROSCI.19-20-09029.1999

Romer, D. (2010). Adolescent risk taking, impulsivity, and brain development: Implications for prevention. *Developmental Psychobiology, 52*(3), 263–276. https://doi.org/10.1002/dev.20442

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rossion, B., Curran, T., & Gauthier, I. (2002). A defense of the subordinate-level expertise account for the N170 component. *Cognition, 85*(2), 189–196. https://doi.org/10.1016/S0010-0277(02)00101-4

Rossion, B., Joyce, C. A., Cottrell, G. W., & Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *NeuroImage, 20*(3), 1609–1624. https://doi.org/10.1016/j.neuroimage.2003.07.010

Ruchsow, M., Groen, G., Kiefer, M., Hermle, L., Spitzer, M., & Falkenstein, M. (2008). Impulsiveness and ERP components in a Go/NoGo task. *Journal of Neural Transmission, 115*, 909–915. https://doi.org/10.1007/s00702-008-0042-7

Sabatinelli, D., Lang, P. J., Keil, A., & Bradley, M. M. (2007). Emotional perception: Correlation of functional MRI and event-related potentials. *Cerebral Cortex, 17*(5), 1085–1091. https://doi.org/10.1093/cercor/bhl017

Sadeh, B., Podlipsky, I., Zhdanov, A., & Yovel, G. (2010). Event-related potential and functional MRI measures of face-selectivity are highly correlated: A simultaneous ERP-fMRI investigation. *Human Brain Mapping, 31*(10), 1490–1501. https://doi.org/10.1002/hbm.20952

Sahoo, S., & Panda, R. K. (2019). Exploring entrepreneurial orientation and intentions among technical university students: Role of contextual antecedents. *Education + Training, 61*(6), 718–736. https://doi.org/10.1108/ET-11-2018-0247

Saleme, D. M., Kluwe-Schiavon, B., Soliman, A., Misiak, B. ł, Frydecka, D., & Moustafa, A. A. (2018). Factors underlying risk taking in heroin-dependent individuals: Feedback processing and environmental contingencies. *Behavioural Brain Research, 350*, 23–30. https://doi.org/10.1016/j.bbr.2018.04.052

Salim, M. A. M., Van der Veen, F. M., Van Dongen, J.,D.M., & Franken, I. H. A. (2015). Brain activity elicited by reward and reward omission in individuals with psychopathic traits: An ERP study. *Biological Psychology, 110*, 50–58. https://doi.org/10.1016/j.biopsycho.2015.07.001

Sambrook, T. D., & Goslin, J. (2015). A neural reward prediction error revealed by a meta-analysis of ERPs using great grand averages. *Psychological Bulletin, 141*(1), 213–235. https://doi.org/10.1037/bul0000006

San Martín, R. (2012). Event-related potential studies of outcome processing and feedback-guided learning. *Frontiers in Human Neuroscience, 6*, Article 304. https://doi.org/10.3389/fnhum.2012.00304

Sattler, J. M., & Ryan, J. J. (2009). *Assessment with the WAIS-IV*. Jerome M. Sattler Publisher.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Scheel, A. M., Schijen, M. R. M. J., & Lakens, D. (2021). An excess of positive results: Comparing the standard psychology literature with registered reports. *Advances in Methods and Practices in Psychological Science, 4*(2), Article 25152459211007467. https://doi.org/10.1177/25152459211007467

Schindler, S., & Bublatzky, F. (2020). Attention and emotion: An integrative review of emotional face processing as a function of attention. *Cortex, 130*, 362–386. https://doi.org/10.1016/j.cortex.2020.06.010

Schmidt, B., Kessler, L., Holroyd, C. B., & Miltner, W. H. R. (2019). Wearing a bike helmet leads to less cognitive control, revealed by lower frontal midline theta power and risk indifference. *Psychophysiology, 56*(12), Article e13458. https://doi.org/10.1111/psyp.13458

Schmidt, B., Mussel, P., & Hewig, J. (2013). I'm too calm – Let's take a risk! On the impact of state and trait arousal on risk taking. *Psychophysiology, 50*(5), 498–503. https://doi.org/10.1111/psyp.12032

Schonberg, T., Fox, C. R., & Poldrack, R. A. (2011). Mind the gap: Bridging economic and naturalistic risk-taking with cognitive neuroscience. *Trends in Cognitive Sciences, 15*(1), 11–19. https://doi.org/10.1016/j.tics.2010.10.002

Schuermann, B., Endrass, T., & Kathmann, N. (2012). Neural correlates of feedback processing in decision-making under risk. *Frontiers in Human Neuroscience, 6*, Article 204. https://doi.org/10.3389/fnhum.2012.00204

Schultz, W., Preuschoff, K., Camerer, C., Hsu, M., Fiorillo, C. D., Tobler, P. N., & Bossaerts, P. (2008). Explicit neural signals reflecting reward uncertainty. *Philosophical Transactions of the Royal Society, 363*(1511), 3801–3811. https://doi.org/10.1098/rstb.2008.0152

Schumpe, B. M., Brizi, A., Giacomantonio, M., Panno, A., Kopetz, C., Kosta, M., & Mannetti, L. (2017). Need for cognitive closure decreases risk taking and motivates discounting of delayed rewards. *Personality and Individual Differences, 107*, 66–71. https://doi.org/10.1016/j.paid.2016.11.039

Schupp, H. T., Flaisch, T., Stockburger, J., & Junghöfer, M. (2006). Emotion and attention: Event-related brain potential studies. *Progress in Brain Research, 156*, 31–51. https://doi.org/10.1016/S0079-6123(06)56002-9

Schwartz, S. J., Lilienfeld, S. O., Meca, A., & Sauvigné, K. C. (2016). The role of neuroscience within psychology: A call for inclusiveness over exclusiveness. *American Psychologist, 71*(1), 52–70. https://doi.org/10.1037/a0039678

Sehrig, S., Weiss, A., Miller, G. A., & Rockstroh, B. (2019). Decision-and feedback-related brain potentials reveal risk processing mechanisms in patients with alcohol use disorder. *Psychophysiology, 56*(12), Article e13450. https://doi.org/10.1111/psyp.13450

Sellke, T., Bayarri, M. J., & Berger, J. O. (2001). Calibration of ρ values for testing precise null hypotheses. *The American Statistician, 55*(1), 62–71. https://doi.org/10.1198/000313001300339950

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. https://doi.org/10.1037/a0034418

Shen, I., Lee, D., & Chen, C. (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. https://doi.org/10.1016/j.ijpsycho.2013.11.004

Shuttleworth, M. (2008). *Operationalization.* https://explorable.com/operationalization

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. https://doi.org/10.1177/0956797611417632

Simmons, J. P., & Simonsohn, U. (2017). Power posing: P-curving the evidence. *Psychological Science, 28*(5), 687–693. https://doi.org/10.1177/0956797616658563

Simons, R. F., Graham, F. K., Miles, M. A., & Chen, X. (2001). On the relationship of P3a and the novelty-P3. *Biological Psychology, 56*(3), 207–218. https://doi.org/10.1016/S0301-0511(01)00078-3

Simonsohn, U. (2012). It does not follow: Evaluating the one-off publication bias critiques by Francis (2012a, 2012b, 2012c, 2012d, 2012e, in press). *Perspectives on Psychological Science, 7*(6), 597–599. https://doi.org/10.1177/1745691612463399

Simonsohn, U. (2024). *R code behind the p-curve app* (version 4.10) [source code] https://p-curve.com/app4/pcurve_app4.10.r

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143*(2), 534–547. https://doi.org/10.1037/a0033242

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). P-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9*(6), 666–681. https://doi.org/10.1177/1745691614553988

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better P-curves: Making P-curve analysis more robust to errors, fraud, and ambitious P-hacking, a reply to Ulrich and Miller (2015). *Journal of Experimental Psychology: General, 144*(6), 1146–1152. https://doi.org/10.1037/xge0000104

Simonsohn, U., Simmons, J., & Nelson, L. D. (2022). Above averaging in literature reviews. *Nature Reviews Psychology, 1*, 551–552. https://doi.org/10.1038/s44159-022-00101-8

Slovic, P. (1966). Risk-taking in children: Age and sex differences. *Child Development, 37*(1), 169–176. https://doi.org/10.2307/1126437

Snyder, E., & Hillyard, S. A. (1976). Long-latency evoked potentials to irrelevant, deviant stimuli. *Behavioral Biology, 16*(3), 319–331. https://doi.org/10.1016/S0091-6773(76)91447-4

Sokhadze, E., Stewart, C., Hollifield, M., & Tasman, A. (2008). Event-related potential study of executive dysfunctions in a speeded reaction task in cocaine addiction. *Journal of Neurotherapy, 12*(4), 185–204. https://doi.org/10.1080/10874200802502144

Song, F., Parekh-Bhurke, S., Hooper, L., Loke, Y. K., Ryder, J. J., Sutton, A. J., Hing, C. B., & Harvey, I. (2009). Extent of publication bias in different categories of research cohorts: A meta-analysis of empirical studies. *BMC Medical Research Methodology, 9*, Article 79. https://doi.org/10.1186/1471-2288-9-79

Soong, A. C. K., 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. https://doi.org/10.1016/0013-4694(93)90018-Q

Šoškić, A., Styles, S. J., Kappenman, E. S., & Ković, V. (2024). Garden of forking paths in ERP research – Effects of varying pre-processing and analysis steps in an N400 experiment. *Psychophysiology, 61*(10), Article e14628. https://doi.org/10.1111/psyp.14628

Spaniol, J., Di Muro, F., & Ciaramelli, E. (2018). Differential impact of ventromedial prefrontal cortex damage on “hot” and “cold” decisions under risk. *Cognitive,*

Affective, & Behavioral Neuroscience, 19, 477–489. https://doi.org/10.3758/s13415-018-00680-1

Squires, K. C., Wickens, C., Squires, N. K., & Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science, 193*(4258), 1142–1146. https://doi.org/10.1126/science.959831

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology, 38*(4), 387–401. https://doi.org/10.1016/0013-4694(75)90263-1

Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychological Bulletin, 144*(12), 1325–1346. https://doi.org/10.1037/bul0000169

Starcke, K., Tuschen-Caffier, B., Markowitsch, H. J., & Brand, M. (2009). Skin conductance responses during decisions in ambiguous and risky situations in obsessive-compulsive disorder. *Cognitive Neuropsychiatry, 14*(3), 199–216. https://doi.org/10.1080/13546800902996831

Starcke, K., Tuschen-Caffier, B., Markowitsch, H. J., & Brand, M. (2010). Dissociation of decisions in ambiguous and risky situations in obsessive-compulsive disorder. *Psychiatry Research, 175*(1-2), 114–120. https://doi.org/10.1016/j.psychres.2008.10.022

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review, 28*(1), 78–106. https://doi.org/10.1016/j.dr.2007.08.002

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. https://doi.org/10.1037/a0012955

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa. *Journal of the American Statistical Association, 54*(285), 30–34. https://doi.org/10.1080/01621459.1959.10501497

Sterne, J. A. C., Becker, B. J., & Egger, M. (2005). The funnel plot. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication bias in meta-analysis* (pp. 75–98). Wiley. https://doi.org/10.1002/0470870168

Suomala, J. (2018a). Benefits of neuromarketing in the product/service innovation process and creative marketing campaign. In L. Moutinho & M. Sokele (Eds.), *Innovative*

research methodologies in management (pp. 159–177). Springer Nature. https://doi.org/10.1007/978-3-319-64400-4

Suomala, J. (2018b). The neuroscience research methods in management. In L. Moutinho & M. Sokele (Eds.), *Innovative research methodologies in management* (pp. 135–158). Springer Nature. https://doi.org/10.1007/978-3-319-64400-4

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*, 857–868. https://doi.org/10.3758/s13415-018-0609-1

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science, 150*(3700), 1187–1188. https://doi.org/10.1126/science.150.3700.1187

Takács, Á, Kóbor, A., Janacsek, K., Honbolygó, F., Csépe, V., & Németh, D. (2015). High trait anxiety is associated with attenuated feedback-related negativity in risky decision making. *Neuroscience Letters, 600*, 188–192. https://doi.org/10.1016/j.neulet.2015.06.022

Tamburin, S., Maier, A., Schiff, S., Lauriola, M. F., Di Rosa, E., Zanette, G., & Mapelli, D. (2014). Cognition and emotional decision-making in chronic low back pain: An ERPs study during Iowa Gambling Task. *Frontiers in Psychology, 5*, Article 1350. https://doi.org/10.3389/fpsyg.2014.01350

Taylor, J. B., Visser, T. A. W., Fueggle, S. N., 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. https://doi.org/10.1016/j.dcn.2018.01.003

Thalemann, R., Wölfling, K., & Grüsser, S. M. (2007). Specific cue reactivity on computer game-related cues in excessive gamers. *Behavioral Neuroscience, 121*(3), 614–618. https://doi.org/10.1037/0735-7044.121.3.614

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica, 26*(1), 24–36. https://doi.org/10.2307/1907382

Trotzke, P., Starcke, K., Pedersen, A., Müller, A., & Brand, M. (2015). Impaired decision making under ambiguity but not under risk in individuals with pathological buying – behavioral and psychophysiological evidence. *Psychiatry Research, 229*(1-2), 551–558. https://doi.org/10.1016/j.psychres.2015.05.043

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297–323. https://doi.org/10.1007/BF00122574

Van Assen, M. A. L. M., Van Aert, R., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods, 20*(3), 293–309. https://doi.org/10.1037/met0000025

Van Boxtel, G. J. M., Van der Molen, M. W., & Jennings, J. R. (2005). Differential involvement of the anterior cingulate cortex in performance monitoring during a stop-signal task. *Journal of Psychophysiology, 19*(1), 1–10. https://doi.org/10.1027/0269-8803.19.1.1

Van Boxtel, G. J. M., Van der Molen, M. W., Jennings, J. R., & Brunia, C. H. M. (2001). A psychophysiological analysis of inhibitory motor control in the stop-signal paradigm. *Biological Psychology, 58*(3), 229–262. https://doi.org/10.1016/S0301-0511(01)00117-X

Van den Berg, I., Franken, I. H. A., & Muris, P. (2010). A new scale for measuring reward responsiveness. *Frontiers in Psychology, 1*, Article 239. https://doi.org/10.3389/fpsyg.2010.00239

Van den Berg, I., Franken, I. H. A., & Muris, P. (2011). Individual differences in sensitivity to reward. *Journal of Psychophysiology, 25*(2), 81–86. https://doi.org/10.1027/0269-8803/a000032

Van Duijvenvoorde, A. C., Huizenga, H. M., Somerville, L. H., Delgado, M. R., Powers, A., Weeda, W. D., Casey, B. J., Weber, E. U., & Figner, B. (2015). Neural correlates of expected risks and returns in risky choice across development. *The Journal of Neuroscience, 35*(4), 1549–1560. https://doi.org/10.1523/JNEUROSCI.1924-14.2015

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. https://doi.org/10.1016/j.psychres.2006.05.011

Van Praag, C. M. (1999). Some classic views on entrepreneurship. *De Economist, 147*, 311–335. https://doi.org/10.1023/A:1003749128457

Van Ravenzwaaij, D., Dutilh, G., & Wagenmakers, E. (2011). Cognitive model decomposition of the BART: Assessment and application. *Journal of Mathematical Psychology, 55*(1), 94–105. https://doi.org/10.1016/j.jmp.2010.08.010

Van Veen, V., & Carter, C. S. (2002). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of Cognitive Neuroscience, 14*(4), 593–602. https://doi.org/10.1162/08989290260045837

Verleger, R. (2020). Effects of relevance and response frequency on P3b amplitudes: Review of findings and comparison of hypotheses about the process reflected by P3b. *Psychophysiology, 57*(7), Article e13542. https://doi.org/10.1111/psyp.13542

Volkow, N. D., & Boyle, M. (2018). Neuroscience of addiction: Relevance to prevention and treatment. *American Journal of Psychiatry, 175*(8), 729–740. https://doi.org/10.1176/appi.ajp.2018.17101174

Volpe, U., Mucci, A., Bucci, P., Merlotti, E., Galderisi, S., & Maj, M. (2007). The cortical generators of P3a and P3b: A LORETA study. *Brain Research Bulletin, 73*(4-6), 220–230. https://doi.org/10.1016/j.brainresbull.2007.03.003

Volz, K. G., & Gigerenzer, G. (2012). Cognitive processes in decisions under risk are not the same as in decisions under uncertainty. *Frontiers in Neuroscience, 6*, Article 105. https://doi.org/10.3389/fnins.2012.00105

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior*. Princeton University Press.

Wallsten, T. S., Pleskac, T. J., & Lejuez, C. W. (2005). Modeling behavior in a clinically diagnostic sequential risk-taking task. *Psychological Review, 112*(4), 862–880. https://doi.org/10.1037/0033-295X.112.4.862

Walsh, M. M., & Anderson, J. R. (2012). Learning from experience: Event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neuroscience & Biobehavioral Reviews, 36*(8), 1870–1884. https://doi.org/10.1016/j.neubiorev.2012.05.008

Wang, C., Yeo, J. E., & Shou, Y. (2025). Association between risk propensity and substance use: A multilevel meta-analysis. *Drug and Alcohol Dependence, 271*, Article 112640. https://doi.org/10.1016/j.drugalcdep.2025.112640

Wang, J., & Li, H. (2023). Emotional processing deficits in individuals with problematic pornography use: Unpleasant bias and pleasant blunting. *Journal of Behavioral Addictions, 12*(4), 1046–1060. https://doi.org/10.1556/2006.2023.00058

Wang, W., & Wang, Y. (2001). Sensation seeking correlates of passive auditory P3 to a single stimulus. *Neuropsychologia, 39*(11), 1188–1193. https://doi.org/10.1016/S0028-3932(01)00051-3

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*, Article 43. https://doi.org/10.3389/fnins.2012.00043

Weafer, J., Baggott, M. J., & De Wit, H. (2013). Test-retest reliability of behavioral measures of impulsive choice, impulsive action, and inattention. *Experimental and Clinical Psychopharmacology, 21*(6), 475–481. https://doi.org/10.1037/a0033659

Webber, H. E., de Dios, C., Kessler, D. A., Schmitz, J. M., Lane, S. D., & Suchting, R. (2022). Late positive potential as a candidate biomarker of motivational relevance in substance use: Evidence from a meta-analysis. *Neuroscience & Biobehavioral Reviews, 141*, Article 104835. https://doi.org/10.1016/j.neubiorev.2022.104835

Webber, H. E., de Dios, C., Kessler, D. A., Schmitz, J. M., Lane, S. D., & Suchting, R. (2024). A meta-analysis of electrophysiological biomarkers of reward and error monitoring in substance misuse. *Psychophysiology, 61*(5), Article e14515. https://doi.org/10.1111/psyp.14515

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. https://doi.org/10.1016/j.jrp.2012.07.001

Wei, S., Zheng, Y., Li, Q., Dai, W., Sun, J., Wu, H., & Liu, X. (2018). Enhanced neural responses to monetary rewards in methamphetamine use disordered individuals compared to healthy controls. *Physiology & Behavior, 195*, 118–127. https://doi.org/10.1016/j.physbeh.2018.08.003

Weller, J. A., King, M. L., Figner, B., & Denburg, N. L. (2019). Information use in risky decision making: Do age differences depend on affective context? *Psychology and Aging, 34*(7), 1005–1020. https://doi.org/10.1037/pag0000397

Wetzel, N., Widmann, A., Berti, S., & Schröger, E. (2006). The development of involuntary and voluntary attention from childhood to adulthood: A combined behavioral and event-related potential study. *Clinical Neurophysiology, 117*(10), 2191–2203. https://doi.org/10.1016/j.clinph.2006.06.717

Whipple, S. C., Berman, S. M., & Noble, E. P. (1991). Event-related potentials in alcoholic fathers and their sons. *Alcohol, 8*(4), 321–327. https://doi.org/10.1016/0741-8329(91)90497-K

White, T. L., Lejuez, C. W., & De Wit, H. (2008). Test-retest characteristics of the Balloon Analogue Risk Task (BART). *Experimental and Clinical Psychopharmacology, 16*(6), 565. https://doi.org/10.1037/a0014083

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. https://doi.org/10.1016/S0191-8869(00)00064-7

Wieczorek, O., Unger, S., Riebling, J., Erhard, L., Koß, C., & Heiberger, R. (2021). Mapping the field of psychology: Trends in research topics 1995–2015. *Scientometrics, 126*, 9699–9731. https://doi.org/10.1007/s11192-021-04069-9

Wiersema, J. R., Van der Meere, J. J., & Roeyers, H. (2005). ERP correlates of impaired error monitoring in children with ADHD. *Journal of Neural Transmission, 112*, 1417–1430. https://doi.org/10.1007/s00702-005-0276-6

Wiersema, J. R., Van der Meere, J. J., & Roeyers, H. (2009). ERP correlates of error monitoring in adult ADHD. *Journal of Neural Transmission, 116*, 371–379. https://doi.org/10.1007/s00702-008-0165-x

Wölfling, K., Flor, H., & Grüsser, S. M. (2008). Psychophysiological responses to drug-associated stimuli in chronic heavy cannabis use. *European Journal of Neuroscience, 27*(4), 976–983. https://doi.org/10.1111/j.1460-9568.2008.06051.x

Wölfling, K., Mörsen, C. P., Duven, E., Albrecht, U., Grüsser, S. M., & Flor, H. (2011). To gamble or not to gamble: At risk for craving and relapse–learned motivated attention in pathological gambling. *Biological Psychology, 87*(2), 275–281. https://doi.org/10.1016/j.biopsycho.2011.03.010

Wronka, E. A., Kaiser, J., & Coenen, A. M. L. (2012). Neural generators of the auditory evoked potential components P3a and P3b. *Acta Neurobiologiae Experimentalis, 72*(1), 51–64. https://doi.org/10.55782/ane-2012-1880

Wu, Y., & Zhou, X. (2009). The P300 and reward valence, magnitude, and expectancy in outcome evaluation. *Brain Research, 1286*, 114–122. https://doi.org/10.1016/j.brainres.2009.06.032

Xu, S., Pan, Y., Wang, Y., Spaeth, A. M., Qu, Z., & Rao, H. (2016). Real and hypothetical monetary rewards modulate risk taking in the brain. *Scientific Reports, 6*, Article 29520. https://doi.org/10.1038/srep29520

Yang, B., Yang, S., Zhao, L., Yin, L., Liu, X., & An, S. (2009). Event-related potentials in a Go/NoGo task of abnormal response inhibition in heroin addicts. *Science in China Series C: Life Sciences, 52*, 780–788. https://doi.org/10.1007/s11427-009-0106-4

Yang, L., Chen, Z., Qi, L., Yang, H., & Zhang, Y. (2024). Processing abnormalities in monetary outcome evaluations among male individuals with opioid use disorder: Evidence from feedback-related negativity. *The American Journal of Drug and Alcohol Abuse, 50*(2), 207–217. https://doi.org/10.1080/00952990.2024.2304036

Yang, L., Zhang, J., & Zhao, X. (2015). Implicit processing of heroin and emotional cues in abstinent heroin users: Early and late event-related potential effects. *The American Journal of Drug and Alcohol Abuse, 41*(3), 237–245. https://doi.org/10.3109/00952990.2015.1020383

Yang, L., Zhang, Y., Zhang, J., Li, N., & Chen, Z. (2022). Altered neural processing of social reward in male heroin abstainers. *International Journal of Psychophysiology, 176*, 142–148. https://doi.org/10.1016/j.ijpsycho.2022.04.005

Yau, Y. H. C., Potenza, M. N., Mayes, L. C., & Crowley, M. J. (2015). Blunted feedback processing during risk-taking in adolescents with features of problematic internet use. *Addictive Behaviors, 45*, 156–163. https://doi.org/10.1016/j.addbeh.2015.01.008

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111*(4), 931–959. https://doi.org/10.1037/0033-295X.111.4.931

Yeung, N., & Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *The Journal of Neuroscience, 24*(28), 6258–6264. https://doi.org/10.1523/JNEUROSCI.4537-03.2004

Young, M. E., & McCoy, A. W. (2019). Variations on the balloon analogue risk task: A censored regression analysis. *Behavior Research Methods, 51*, 2509–2521. https://doi.org/10.3758/s13428-018-1094-8

Zamarian, L., Sinz, H., Bonatti, E., Gamboz, N., & Delazer, M. (2008). Normal aging affects decisions under ambiguity, but not decisions under risk. *Neuropsychology, 22*(5), 645–657. https://doi.org/10.1037/0894-4105.22.5.645

Zhang, L., Dong, Y., Ji, Y., Tao, R., Chen, X., Ye, J., Yu, F., Zhu, C., & Wang, K. (2015). Trait-related decision making impairment in obsessive-compulsive disorder: Evidence from decision making under ambiguity but not decision making under risk. *Scientific Reports, 5*, Article 17312. https://doi.org/10.1038/srep17312

Zhang, L., Dong, Y., Ji, Y., Zhu, C., Yu, F., Ma, H., Chen, X., & Wang, K. (2015). Dissociation of decision making under ambiguity and decision making under risk: A neurocognitive endophenotype candidate for obsessive-compulsive disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 57*, 60–68. https://doi.org/10.1016/j.pnpbp.2014.09.005

Zhang, Y., Ou, H., Yuan, T., & Sun, J. (2021). Electrophysiological indexes for impaired response inhibition and salience attribution in substance (stimulants and depressants) use disorders: A meta-analysis. *International Journal of Psychophysiology, 170*, 133–155. https://doi.org/10.1016/j.ijpsycho.2021.10.008

Zhao, Q., Li, H., Hu, B., Wu, H., & Liu, Q. (2017). Abstinent heroin addicts tend to take risks: ERP and source localization. *Frontiers in Neuroscience, 11*, Article 681. https://doi.org/10.3389/fnins.2017.00681

Zheng, Y., Sheng, W., Xu, J., & Zhang, Y. (2014). Sensation seeking and error processing. *Psychophysiology, 51*(9), 824–833. https://doi.org/10.1111/psyp.12240

Zhong, N., Chen, T., Zhu, Y., Su, H., Ruan, X., Li, X., Tan, H., Jiang, H., Du, J., & Zhao, M. (2020). Smaller feedback-related negativity (FRN) reflects the risky decision-making deficits of methamphetamine dependent individuals. *Frontiers in Psychiatry, 11*, Article 320. https://doi.org/10.3389/fpsyt.2020.00320

Zhou, Z. H., Yuan, G., Yao, J., Li, C., & Cheng, Z. (2010). An event-related potential investigation of deficient inhibitory control in individuals with pathological internet use. *Acta Neuropsychiatrica, 22*(5), 228–236. https://doi.org/10.1111/j.1601-5215.2010.00444.x

Zhou, Z., Yu, R., & Zhou, X. (2010). To do or not to do? Action enlarges the FRN and P300 effects in outcome evaluation. *Neuropsychologia, 48*(12), 3606–3613. https://doi.org/10.1016/j.neuropsychologia.2010.08.010

Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking*. Cambridge university press.

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. https://doi.org/10.1037/0022-3514.65.4.757

Zuckerman, M., & Neeb, M. (1979). Sensation seeking and psychopathology. *Psychiatry Research, 1*(3), 255–264. https://doi.org/10.1016/0165-1781(79)90007-6

Summary in English

Risk-taking propensity (RTP) is a core concept in several fields of research, including psychology, economics, and management. Consequently, an important question is how to best measure RTP. Recent years have seen an increase in the use of behavioural and (neuro)physiological measures for operationalising RTP and related concepts like reward responsiveness and impulsivity. In part, this increase seems to originate from the assumption that behavioural and (neuro)physiological measures provide more accurate operationalisations of these constructs than do self-reports, due to the latter being influenced by social desirability bias and limits on participants' introspective abilities. However, the suitability of behavioural and (neuro)physiological measures for operationalising RTP and related concepts has been examined only sparsely. The present dissertation addresses this gap.

Chapters 2, 3, and 4 evaluate *behavioural measures* of risk-taking propensity. In particular, they look at the Balloon Analogue Risk Task (BART) and the Columbia Card Task (CCT), which are both frequently used in the psychological literature to measure RTP. The chapters demonstrate that the BART suffers from various methodological shortcomings. For example, it cannot distinguish properly between adaptive and maladaptive risk-taking. Also, a subset of observations in the BART is censored, because trials can finish while participants may have wanted to take more risk. Together, the problems discussed in Chapters 2 and 3 show that the BART is an unfit measure for operationalising risk-taking propensity. The CCT, on the other hand, receives positive acclaim. While in some ways the task is similar to the BART – both are sequential risk-taking tasks in which participants can win and lose money – the CCT circumvents the methodological problems that make the BART unsuited for measuring RTP. In addition, the design of the CCT makes it possible for researchers to examine the impact that information on gains, losses, and the probability of losing has on the amount of risk participants take. This makes the CCT relevant not only to researchers from psychology, but to researchers from other fields like economics and management as well. This relevance is illustrated in Chapter 4, which uses the CCT to measure risk-taking propensity in relation to persons' entrepreneurial orientation.

Chapters 5, 6, and 7 subsequently assess the suitability of operationalising RTP with the use of *electroencephalography (EEG)*, which measures changes in the brain's electrical activity. To this end, Chapters 5 and 6 record EEG at relevant moments in the decision-making process, such as when participants lose money in the CCT. Across several large and well-powered samples, strong EEG signals are found. Unfortunately, these show poor

convergent validity as they correlate only weakly with self-reported and behavioural measures of RTP, reward responsiveness, and impulsivity. At the outset, these correlations therefore indicate that EEG may not be a suitable operationalisation for these constructs. However, it is further found that self-reports and behavioural measures gauging these constructs also do not correlate strongly, and that measures of the same type (such as two EEG measures) do correlate well. This could mean that EEG *is* suitable for operationalising RTP and related constructs, but that it measures a different dimension or aspect of it than for example self-reports do. Furthermore, Chapter 7 finds support for the existence of EEG differences between healthy persons and persons with addiction, indicating that EEG is a suitable operationalisation when differentiating between groups of people on the more extreme ends of a spectrum. In all, EEG may be appropriate to operationalise RTP and related constructs, but only under specific conditions.

By examining the suitability of behavioural and EEG measures for operationalising RTP and related constructs, the present dissertation contributes to the fundamental literature on the measurement of RTP as well as to applied research that uses the measurement instruments that are evaluated. The findings emphasise how choosing an operationalisation is a complicated decision that should involve careful consideration of the options, as the conclusions that can be drawn from empirical analyses depend on accurate measurement of the concepts of interest.

Summary in Dutch

Risicobereidheid (de neiging van een individu om risico te nemen) speelt een prominente rol in diverse onderzoeksgebieden, waaronder in de psychologie, economie, en in management. Een belangrijke vraag is dan ook hoe risicobereidheid het best kan worden gemeten. De laatste jaren worden risicobereidheid en gerelateerde constructen zoals beloningsgevoeligheid en impulsiviteit steeds vaker geoperationaliseerd door middel van gedragstaken en (neuro)fysiologische metingen. Deze toename lijkt deels gedreven door het idee dat gedragstaken en (neuro)fysiologische metingen een nauwkeurigere operationalisering van deze constructen bieden dan zelf-rapportages, welke onderhevig zijn aan sociale wenselijkheid *bias* en beperkingen in de introspectieve kennis van proefpersonen. Echter, tot op heden is slechts beperkt onderzocht hoe geschikt gedragstaken en (neuro)fysiologische metingen eigenlijk zijn voor het operationaliseren van risicobereidheid en gerelateerde constructen. Het huidige proefschrift gaat in op deze vraag.

Hoofdstuk 2, 3, en 4 onderzoeken *gedragstaken* die risicobereidheid beogen te meten. De focus ligt hierbij op de Balloon Analogue Risk Task (BART) en de Columbia Card Task (CCT), die beide in de psychologie veelvuldig worden gebruikt voor dit doel. De hoofdstukken laten zien dat dat de BART verschillende methodologische tekortkomingen heeft. Zo kan de taak adaptief en maladaptief risicogedrag niet goed van elkaar onderscheiden. Een ander probleem is dat een deel van de observaties in de BART slechts beperkt informatief is, omdat *trials* kunnen eindigen terwijl proefpersonen mogelijk meer risico hadden willen nemen. Op basis van de problemen die in Hoofdstukken 2 en 3 naar voren komen, kan worden geconcludeerd dat de BART geen geschikt instrument is voor het operationaliseren van risicobereidheid. De CCT wordt daarentegen juist wel goed beoordeeld. Hoewel deze gedragstaak in meerdere opzichten op de BART lijkt – beide zijn sequentiële risicotaken waarbij proefpersonen geld kunnen winnen en verliezen – omzeilt de CCT de methodologische problemen die de BART ongeschikt maken voor het meten van risicobereidheid. Bovendien biedt het ontwerp van de CCT de mogelijkheid om te onderzoeken wat de invloed is van informatie over winst, verlies, en de verlieskans op de hoeveelheid risico die proefpersonen nemen. Dit maakt de CCT niet alleen relevant voor onderzoekers in de psychologie, maar ook voor onderzoekers in andere vakgebieden, zoals economie en management. Dit wordt geïllustreerd in Hoofdstuk 4, waarin de CCT wordt gebruikt om risicobereidheid te meten in relatie tot de ondernemerschapsoriëntatie van mensen.

Hoofdstuk 5, 6, en 7 onderzoeken vervolgens de mogelijkheid om risicobereidheid te operationaliseren met behulp van *elektro-encefalografie (EEG)*. Deze technologie meet veranderingen in de elektrische activiteit van het brein. Hoofdstuk 5 en 6 meten EEG-signalen op relevante momenten tijdens het beslisproces, bijvoorbeeld wanneer proefpersonen geld verliezen in de CCT. In drie grote steekproeven met goede statistische *power* worden sterke EEG-signalen gevonden. Deze vertonen echter slechte convergente validiteit aangezien ze zwak correleren met risicobereidheid en gerelateerde constructen zoals gemeten met zelf-rapportages en gedragsmetingen. Dit wijst er in eerste instantie op dat EEG mogelijk ongeschikt is om deze constructen te operationaliseren. Echter, verder blijkt dat zelfrapportages en gedragsmetingen van deze constructen ook niet sterk correleren, en dat metingen van hetzelfde type (zoals twee EEG-metingen) juist wel goed correleren. Het is daarom mogelijk dat EEG *wel* geschikt is om risicobereidheid en gerelateerde constructen te operationaliseren, maar dat het een andere dimensie of een ander aspect van deze constructen meet dan bijvoorbeeld zelf-rapportages doen. Hoofdstuk 7 laat bovendien zien dat er EEG verschillen bestaan tussen gezonde personen en personen met een verslaving. Dit suggereert dat EEG een geschikte operationalisering is als gedifferentieerd wordt tussen groepen mensen die zich op tegenoverliggende uiteinden van een spectrum bevinden. Alles samengenomen is het mogelijk dat EEG, onder specifieke voorwaarden, risicobereidheid en gerelateerde constructen kan operationaliseren.

Door te onderzoeken hoe geschikt gedragstaken en EEG-metingen zijn voor het operationaliseren van risicobereidheid en gerelateerde constructen, levert het huidige proefschrift een bijdrage aan zowel de fundamentele literatuur naar het meten van risicobereidheid, als aan toegepast onderzoek dat de geëvalueerde meetinstrumenten gebruikt. De uitkomsten benadrukken dat het kiezen van een operationalisering een complexe beslissing is waarbij de opties zorgvuldig tegen elkaar moeten worden afgewogen, aangezien de conclusies die uit empirisch onderzoek getrokken kunnen worden afhankelijk zijn van het accuraat meten van de relevante concepten.

About the Author

Kristel de Groot was born on November 8, 1993 in Gouda, the Netherlands. Before starting her PhD, she obtained a Bachelor of Science in Psychology at the Erasmus University Rotterdam in 2015, specialising in biological and cognitive psychology and with a minor in neuroeconomics; and a Master of Science in Psychology (*cum laude*) at the Erasmus University Rotterdam in 2016, again with a specialisation in biological and cognitive psychology.

In December 2017, Kristel started her PhD candidacy at the Erasmus Research Institute of Management (ERIM), with combined supervision from the Erasmus School of Economics (ESE) and the Erasmus School of Social and Behavioural Sciences (ESSB). Her research is focused on the evaluation and implementation of behavioural tasks and electroencephalography in psychology, economics, and management, and aims to bridge the gap between these fields. In addition, she coordinates a longitudinal research study into university students' mental health. The work she has done during her PhD has been published in *Applied Psychology*, *Frontiers in Psychology*, *Frontiers in Education*, the *Journal of Open Psychology Data*, the *Journal of Trial and Error*, *Biological Psychology*, *Psychophysiology*, and *Psychometrika*. Kristel has a passion for teaching and has, in addition to thesis and workgroup supervision, coordinated the Electrophysiology Course for the Brain & Cognition Master for several years. As of November 2024, Kristel is a post-doctoral researcher at the Strategy & Entrepreneurship department of the Tilburg School of Economics and Management (TiSEM).

Portfolio

Journal Publications

Bernoster, I., De Groot, K., Wieser, M. J., Thurik, R., & Franken, I. H. (2019). Birds of a feather flock together: Evidence of prominent correlations within but not between self-report, behavioural, and electrophysiological measures of impulsivity. *Biological Psychology, 145*, 112-123. https://doi.org/10.1016/j.biopsycho.2019.04.008

De Groot, K. (2020). Burst beliefs – Methodological problems in the Balloon Analogue Risk Task and implications for its use. *Journal of Trial and Error, 1*, 43-51. https://doi.org/10.36850/mr1

De Groot, K. (2020). Non-clinical autistic traits correlate with social and ethical but not with financial and recreational risk-taking. *Frontiers in Psychology, 11*, Article 360. https://doi.org/10.3389/fpsyg.2020.00360

De Groot, K., & Thurik, R. (2018). Disentangling risk and uncertainty: When risk-taking measures are not about risk. *Frontiers in Psychology, 9*, Article 2194. https://doi.org/10.3389/fpsyg.2018.02194

De Groot, K., & Van Strien, J. W. (2019). Event-related potentials in response to feedback following risk-taking in the hot version of the Columbia Card Task. *Psychophysiology, 56*, Article e13390. https://doi.org/10.1111/psyp.13390

De Groot, K., Wieman, S. M., Van Strien, J. W., & Lindemann, O. (2024). To each their own: Sociodemographic disparities in student mental health. *Frontiers in Education, 9*, Article 391067. https://doi.org/10.3389/feduc.2024.1391067

Dijkstra, N. F., De Groot, K., & Rietveld, C. A. (2023). Entrepreneurial orientation and decision-making under risk and uncertainty: Experimental evidence from the Columbia Card Task. *Applied Psychology, 72*, 1577-1592. https://doi.org/10.1111/apps.12436

Ellis, J. L., Sijtsma, K., De Groot, K., & Groenen, P. J. (2024). Reliability theory for measurements with variable test length, illustrated with ERN and Pe collected in the Flanker Task. *Psychometrika, 89*, 1280-1303. https://doi.org/10.1007/s11336-024-09982-5

Terry, J., Ross, R. M., Nagy, T., Salgado, M., Garrido-Vásquez, P., Sarfo, J. O., Cooper, S., Buttner, A. C., Lima, T. J. S., Öztürk, İ., Akay, N., Santos, F. H., Artemenko, C.,

Copping, L. T., Elsherif, M. M., Milovanović, I., Cribbie, R. A., Drushlyak, M. G., Swainston, K., … Field, A. P. (2023). Data from an international multi-centre study of statistics and mathematics anxieties and related variables in university students (the SMARVUS dataset). *Journal for Open Psychology Data, 11*, Article 8. https://doi.org/10.5334/jopd.80

Working Papers

Quantifying evidential value for event-related potentials as biomarkers of Substance Use Disorders: A literature review and p-curve analysis. *Working paper*, with Eric Slob, Oliver Lindemann, & Jan van Strien.

To share or not to share: How researchers' beliefs about benefits, risks, and norms impact their attitude to open access sharing. *Working paper*, with Eric Slob & Sander Wieman.

Presentations

2020	Discussing Failures in Science (Open Science Community Rotterdam)
2021	Do Disabilities Impact Students' Success? (Community for Learning & Innovation)
2023	Sociodemographic Disparities in Student Mental Health (AE PhD Day)
2024	Disabilities and Students' Performance and Wellbeing (study advisors meeting)
2024	To Share or not to Share (Open Science Community Rotterdam)

Teaching Activities

2018-2024	Bachelor and Master thesis supervision, ESE and ESSB
2019-2020	Tutor Affective Neuroscience course, ESSB
2019-2022	Tutor Electrophysiology course, ESSB
2022-2024	Coordinator Electrophysiology course, ESSB
2023-2024	Guest lecturer Neuroeconomics, ESE Honours Class
2023-2024	Guest lecturer Societal Impact, ESSB Graduate School

Academic Reviewing

Acta Psychologica, Advances in Neurodevelopmental Disorders, Applied Economics, Applied Psychology, Autism, Behaviour Research Methods, Biological Psychology, Journal of Autism and Developmental Disorders, Journal of Gambling Studies, Journal of Trial and Error, Psychological Reports, Psychological Science, Psychophysiology

Education

2012-2015	BSc in Psychology, Erasmus University Rotterdam
2015-2016	MSc in Psychology, Erasmus University Rotterdam
2017-2025	PhD, Erasmus University Rotterdam (leave from June 2020 to May 2022)

Formal Certificates

2018	Certificate of Proficiency in English Level C2, Cambridge Assessment English
2023	Basic Examination Qualification, Risbo

PhD Courses ERIM

Mediation, Moderation, and Conditional Process Modelling

Methods Stumblers: Pragmatic Solutions to Challenges in Behavioural Research

My First Bayes: A Gentle introduction to Bayesian Analysis

Necessary Condition Analysis: Theory and Practice

Programming in R

Publishing Strategy

Scientific Integrity

Teaching, Presenting, and Writing in English

Testing and Interpreting Moderation and Mediation with SPSS

PhD Courses EGSH

Brush up your Research Design

Data Analysis with R

English Academic Writing

How to Survive your PhD

Making an Academic Poster

Multilevel Modelling I: An Introduction to Multilevel Modelling

Multilevel Modelling II: Multilevel Structural Equation Modelling

Professionalism and Integrity in Research

Qualitative Coding with ATLAS.ti

Responsible Research Data Management

Self-Presentation: Presenting Yourself and your Research

Grants

2018 €20.000 to study student wellbeing, Community for Learning & Innovation

2024 €2.500 to implement and study open science solutions, DPECS Dragon's Den

Awards

2023 Excellence Award Societal Impact, ESSB Graduate School

The ERIM PhD Series

The ERIM PhD Series contains PhD dissertations in the field of Research in Management defended at Erasmus University Rotterdam and supervised by senior researchers affiliated to the Erasmus Research Institute of Management (ERIM). Dissertations in the ERIM PhD Series are available in full text through: https://pure.eur.nl.

ERIM is the joint research institute of the Rotterdam School of Management (RSM) and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam (EUR).

Dissertations in the last four years

Abdelwahed, A., *Optimizing Sustainable Transit Bus Networks in Smart Cities*, Supervisors: Prof. W. Ketter, Dr. P. van den Berg & Dr. T. Brandt, EPS-2022-549-LIS

Alkema, J., *READY, SET, GO(AL)! New Directions in Goal-Setting Research*, Supervisors: Prof. H.G.H. van Dierendonck & Prof. S.R. Giessner, ESP-2022-555-ORG

Andrei, A.G., *Essays on Behavioral Corporate Governance*, Supervisors: Prof. J. van Oosterhout & Dr. M.H. Benischke, EPS-2025-577-S&E

Badenhausen, K., *IoT – Inducing Organizational Transformation?* Supervisors: Prof. R.A. Zuidwijk & Dr. M. Stevens, EPS-2022-559-LIS

Balocco, F.A.M., *Asymmetric information in programmatic advertising: Three studies on adverse selection, mechanism choices, and fee structures*, Supervisors: Prof. T. Li & Prof. E. van Heck, EPS-2023-565-LIS

Bilgin, B., *Visionary Leadership and the Pursuit of Organizational Visions*, Supervisors: Prof. D.L. van Knippenberg & Dr. I.J. Hoever, EPS-2024-583-ORG

Breet, S., *A Network Perspective on Corporate Entrepreneurship: How workplace relationships influence entrepreneurial behavior*, Supervisors: Prof. J.J.P. Jansen, Prof. J. Dul & Dr. L. Glaser, EPS-2022-545-S&E

Bunders, D., *Gigs of their Own: can platform cooperatives become resilient*? Supervisors: Prof. T. de Moor, Prof. A. Akkerman & Prof. P. Dykstra, EPS-2024-580-ORG

Celiktutan, B., *Self-Other Differences in Human-AI Interactions and Consumption*, Supervisors: Prof. S. Puntoni, Dr. A.K. Klesse & Dr. M.A. Tuk, EPS-2025-598-MKT

Chun, Y., *Navigating Temporal Dimensions Consumer decisions from past to future*, Supervisors: Prof. B. Van Den Bergh & Dr. C.J.P. Lembregts, EPS-2024-592-MKT

Chung, Y.S., *Valorizing Innovation through Imaginativeness in Business Venturing*, Supervisors: Prof. PH.B.F. Franses & Prof. H.P.G. Pennings, EPS-2022-561-MKT

Dijkstra, N., *Statistical Modeling Approaches for Behavioral and Medical Sciences*, Supervisors: Prof. P.J.F. Groenen, Prof. H. Tiemeier & Prof. A.R Thurik, EPS-2022-539-S&E

Duijsters, J., *Change in Inter-Organizational Relationship Portfolios and Social Networks in the Context of Corporate Venturing*, Prof. V.J.A. van de Vrande, Prof. J.J.P. Jansen & Prof. P.P.M.A.R. Heugens, EPS-2024-504-S&E

Fu, G., *Agency Problems in the Mutual Fund Industry*, Supervisors: Prof. M.J.C.M. Verbeek & Dr. E. Genc, EPS-2022-526-F&A

Genc, A., *The Triad of Innovation: Exploring management innovation, business model innovation, and ownership in unlocking firms' innovative potential*, Supervisors: Prof. H.W. Volberda & Prof. J.S. Sidhu, EPS-2024-579-S&E

Giessen, M. van der, *Co-creating Safety and Security: Essays on bridging disparate needs and requirements to foster safety and security*, Supervisors: Prof. G. Jacobs, Prof. J.P. Cornelissen & Prof. P.S. Bayerl, EPS-2022-542-ORG

Giudici, A., *Cooperation, Reliability, and Matching in Inland Freight Transport*, Supervisors: Prof. R.A. Zuidwijk, Prof. C. Thielen & Dr. T. Lu, EPS-2022-533-LIS

Gunadi, M.P., *Essays on Consumers and Numbers*, Supervisors: Prof. S. Puntoni & Dr. B. Van Den Bergh, EPS-2022-558-MKT

Hagen, L. van der., *Machine Learning for Time Slot Management in Grocery Delivery*, Supervisors: Prof. N.A.H. Agatz, Prof. M.B.M. de Koster & Dr. R. Spliet, EPS-2024-595-LIS

Harter, C., *Vulnerability through vertical collaboration in transportation: A complex networks approach*, Supervisors: Prof. R.A. Zuidwijk & Dr. O.R. Koppius, EPS-2023-560-LIS

Hoogendoorn, Y.N., *Vehicle Routing with Varying Levels of Demand Information*, Supervisors: Prof. A.P.M. Wagelmans, Prof. R. Dekker & Dr. R. Spliet, EPS-2024-578-LIS

Hulsen, M., *Wait for others? Social and intertemporal preferences in allocation of healthcare resources*, Supervisors: Prof. K.I.M. Rohde & Prof. N.J.A. van Exel, EPS-2023-556-MKT

Hut, X.M., *Essays on Sustainable Finance*, Supervisors: Prof. M.A. Van Dijk & Prof. dr. A. Andonov, EPS-2025-594-F&A

Jellema, S.F., *Brace for Impact: Good intentions, unintended consequences, and the role of performative micro-processes in organized efforts for societal change*, Supervisors: Prof. J.P. Cornelissen & Prof. T.H. Reus, EPS-2024-587-ORG

Kalvapelle, S.G., *Breaking the Conduit: A Relational Approach to Communication in Management and Entrepreneurship*, Supervisors: Prof. J.P. Cornelissen & Prof. P.P.M.A.R. Heugens, EPS-2023-575-ORG

Kanellopoulos, I., *Navigating Digital Platform Challenges: Three Studies on Blockchain-based Platform Competition, Platform Subsidies for Non-Fungible Token Creation, and Ghosting in the Dating Market*, Supervisors: Prof. T. Li & Prof. H.W.G.M. van Heck, EPS-2024-585-LIS

Karaca, U., *Privacy in Resource Allocation Problems*, Supervisors: Prof. Ş.İ. Birbil & Prof. A.P.M. Wagelmans, EPS-2023-567-LIS

Lam, H.H.W., *Lonely-ship: The emergence and experience of leader loneliness*, Supervisors: Prof. S.R. Giessner, Prof. M. Shemla & Dr. M.D. Werner, EPS-2023-525-ORG

Legierse, W., *The Timing and Pricing of Initial Public Offerings: Evidence from the Low Countries*, Supervisors: Prof. A. de Jong & Prof. P.G.J. Roosenboom, EPS-2022-543-F&A

Leung, Y.K., *The Mind with a Touch of Madness? Mental health and well-being of entrepreneurs*, Supervisors: Prof. A.R. Thurik & Prof. I.H.A. Franken, EPS-2022-546-S&E

Liu, M., *Essays on Financial Disclosure and Innovation*, Supervisors: Prof. J.P.M. Suijs & Dr. P.Y.E. Leung, EPS-2023-582-F&A

Lieshout, R. van, *Integration, Decentralization and Self-Organization Towards Better Public Transport*, Supervisors: Prof. D. Huisman, Dr. P.C. Bouman & Dr. J.M. van den Akker, EPS-2022-547-LIS

Liu, W., *An indigenous perspective on institutions for sustainable business in China*, Supervisors: Prof. P.P.M.A.R. Heugens & Dr. F.H. Wijen, EPS-2023-568-S&E

Mazzola, F., *Externalities in Economics and Finance: Essays on spillover effects and economic decisions*, Supervisors: Prof. W.B. Wagner & Prof. D.G.J. Bongaerts, EPS-2023-573-F&A

Mehrdar, A., *Algorithmic and Behavioral Pricing of Rental Space: Studies on Price Cannibalization, Starting Prices and Overoptimism, and Automated Dynamic Pricing*, Supervisors: Prof. T. Li & Prof. H.W.G.M. van Heck, EPS-2025-584-LIS

Mobini Dehkordi, Z., *Supply Chain Coordination with Separating and Pooling Contracts*, Supervisors: Prof. A.P.M. Wagelmans, Dr. W. van den Heuvel, EPS-2022-554-LIS

Müller, M.J., *Strangers in a Strange Land: Essays on migrant entrepreneurs and managers*, Supervisors: Prof. T.H. Reus & Prof. A.H.L. Slangen, EPS-2024-593-S&E

Musa, SM., *Making a Life on the Margins: An ethnographic account from Kutupalong*, Supervisors: Prof. P.P.M.A.R. Heugens & Prof. L. Berchicci, EPS-2024-586-S&E

Novales Uriarte, A., *Thriving with Digitized Products: How Firms Leverage their Generative Capacity via Experimentation, Learning, and Collaboration*, Supervisors: Prof. H.W.G.M. van Heck & Prof. M. Mocker, EPS-2022-544-LIS

Orlandi, I., *Unringing the stigma bell: Investigating informational and social mechanisms behind boards of directors' appointments*, Supervisors: Prof. P.M.A.R. Heugens & Prof. V. F Misangyi, EPS-2023-564-S&E

Osrouti, F. El, *Norms versus Organizational Practices: Essays in the context of innovation, standardization, and religion*, Supervisors: Prof. H.J. de Vries & Prof. H.W. Volberda, EPS-2024-570-LIS

Pasparakis, A.M., *Leader-Follower Relationships in Technologically Advanced Operations*, Supervisors: Prof. M.B.M. de Koster & Dr. J. de Vries, EPS-2023-571-LIS

Paul, J., *Online Grocery Operations in Omni-channel Retailing: Opportunities and Challenges*, Supervisors: Prof. M.B.M de Koster & Dr. N.A.H. Agatz, EPS-2022-541-LIS

Paundra, J., *The Search for Alternatives to Private Vehicles The multi-perspective investigation of transportation mode choice preference based on policy, ICT development, and psychology*, Supervisors: Prof. W. Ketter, Dr. J. van Dalen & Dr. L. Rook, EPS-2023-496-LIS

Pocchiari, M., *Managing Successful and Resilient Shared-Interest Communities: The Role of Digitization Technologies and Disruptive Events*, Supervisors: Prof. G.H. van Bruggen & Dr. J.M.T. Roos, EPS-2022-552-MKT

Ralcheva, A., *Crowdfunding: A Disruptive Innovation in Entrepreneurial Finance?* Supervisors: Prof. P.G.J. Roosenboom & Dr. T. Lambert, EPS-2023-581-F&A

Ratara, C., *Behavioural and Neural Evidence for Processes Underlying Biases in Decision-Making*, Supervisors: Prof. A. Smidts & Dr. M.A.S. Boksem, EPS-2022-548-MKT

Rijn, L. van, *Improving Access to Healthcare in Underserved Rural Areas*, Supervisors: Prof.dr. A.P.M. Wagelmans & Dr. H. de Vries, EPS-2025-590-LIS

Rossum, B. van, *Optimising Fair Work Allocation: Applications in railway crew planning*, Supervisors: Prof. D. Huisman & Dr. T.A.B. Dollevoet, EPS-2025-600-LIS

Schutter, L.D.M. De., *How Power Differences Shape Interpersonal Trust Processes*, Supervisors: Prof. M.H. van Dijke & Prof. F.M. Bridoux, EPS-2024-588-ORG

Spigt, R.M., *The Role of Central Banks in the Corporate Landscape*, Prof. J.T.J. Smit & Dr. T. Eisert, EPS-2024-591-F&A

Tierean, S.H., *Mind the Gap: The role of psychic distance and supplier's reputation in international buyer-supplier relationships*, Supervisors: Prof. C.B.M. van Riel, Prof. G. van Bruggen & Dr. G.A.J.M. Berens, EPS-2022-551-ORG

Vafa Arani, H., *Creating Shared Value: An operations and supply chain management perspective*, Supervisors: Prof. M.B.M. de Koster & Dr. H. de Vries, EPS-2023-576-LIS

Versteegt, L., *Moral Self-Regulation Across Contexts: From pandemic responses to behaviors in the workplace*, Supervisors: Prof. M.H. van Dijke & Prof. K. van den Bos, EPS-2025-596-ORG

Wagenvoort, M., *Mathematical Models for Planning in Military and Humanitarian Logistics*, Supervisors: Prof. D. Huisman, Dr. P.C. Bouman & Dr. M. van Ee, EPS-2025-601-LIS

Waltré, E., *Leading for Performance in Adversity: Managing Failure, Negative Emotions, and Self-Threats*, Supervisors: Prof. D.L. van Knippenberg & Dr. H.M.S. Dietz, EPS-2022-513-ORG

Welz, M., *Robust Categorical Data Analysis*, Supervisors: Prof. P.J.F. Groenen & Dr. A. Alfons, EPS-2025-597-MKT

Wiessner, Y.T., *The Societal Impact of Multinational Organizations*, Supervisors: Prof. F.H. Wijen, Prof. P.P.M.A.R. Heugens & Dr. B.C. Pinkham, EPS-2024-589-S&E

Willemsen, R.S.H., *Mathematical Programming in Dimension Reduction and Clustering*, Supervisors: Dr. W. van den Heuvel & Dr. M. van de Velden, EPS-2025-599-LIS

Wismans, A., *Behavioural Insights from the COVID-19 Pandemic Studies on Compliance, Vaccination, and Entrepreneurship*, Supervisors: Prof. A.R. Thurik & Prof. I.H.A. Franken, EPS-2023-562-S&E

Yalcin, G., *Consumers in the Age of AI: Understanding Reactions Towards Algorithms and Humans in Marketing Research*, Supervisors: Prof. S. Puntoni & Dr. A. Klesse, EPS-2022-550-MKT

Yang, A., *Corporate Bond Markets: Investor Preferences and Intermediary Frictions*, Supervisors: Prof. P. Verwijmeren & Prof. S. van Bekkum, EPS-2022-553-F&A

Yang, Z., *Human-Centric Digital Transformation: Exploring Human Information Behavior throughThree Field Studies in Digital Policy, Artificial Intelligence, and Emerging Technologies*, Supervisors: Prof. T. Li & Prof. H. W. G. M. van Heck, EPS-2025-602-LIS

Zhang, Q., *Making sense of sensitivity in the workplace: Coping with contextual information in innovation and social networks*, Supervisors: Prof. D. Stam & Dr. S. Tasselli, EPS-2023-574-LIS

Zhou C., *Exploring The Role of Context and Interpretative Dynamics in Large-Scale Cross-Cultural Collaborations*, Supervisors: Prof. T. Simons & Dr. M.D. Werner, EPS-2023-572-ORG

One of the greatest challenges in the social sciences is operationalisation: deciding how to define and measure concepts that cannot be directly observed. This includes risk-taking propensity (RTP), a construct central to various fields of research, such as psychology, economics, and management, due to its link with many impactful behaviours. For instance, individuals high in RTP are more likely to start a business, move away from where they live, or use illegal substances. However, there is no consensus on how individuals' RTP can best be measured. Therefore, this dissertation provides a critical evaluation of several approaches. Part I of the dissertation explores the suitability of measuring RTP with behavioural tasks, particularly two popularised in psychology: the Balloon Analogue Risk Task and the Columbia Card Task. Part II of the dissertation focuses on measuring RTP using electroencephalography (EEG), often recorded while participants perform these same behavioural tasks. In addition, Part II addresses the problems associated with large researcher degrees of freedom and the use of small sample sizes in EEG research. Together, the findings from Parts I and II provide practical guidance to help researchers studying risk-taking propensity make a more informed decision about their choice of measurement.

Kristel de Groot (1993) was born in Gouda, the Netherlands. She started her PhD in December 2017 after obtaining a BSc and MSc in biological and cognitive psychology at the Erasmus University Rotterdam.

ERiM

The Erasmus Research Institute of Management (ERIM) of Erasmus University Rotterdam (EUR) is one of the top management research centres in Europe. ERIM was founded in 1999 by the Rotterdam School of Management (RSM) and Erasmus School of Economics (ESE) to jointly nurture internationally recognised management research.

Research excellence is at the heart of ERIM: It runs EUR's PhD programmes in Business and Management, provides research support for faculty and PhD students, and maintains a solid research infrastructure. Over 450 senior researchers and PhD candidates participate in ERIM's research environment. Coming from myriad areas of expertise, the ERIM Community is constantly striving for excellence at the forefront of the academic world.

This PhD thesis is a result of ERIM's Full-Time PhD Programme in Business and Management. The full-time programme aims to develop international academic talent and produce outstanding research across a wide range of disciplines. Students receive innovative training and coaching from distinguished academic experts – setting students on track to become thought leaders and top researchers at the world's best universities and business schools.

ERiM

Research in Business and Management

ERIM Full-Time PhD

Erasmus School of Economics

Erasmus Research Institute of Management
Erasmus University Rotterdam
Burgemeester Oudlaan 50
Mandeville (T) Building
3062 PA Rotterdam, The Netherlands

P.O. Box 1738
3000 DR Rotterdam, The Netherlands
T: +31 10 408 1182
E: info@erim.eur.nl
W: www.erim.eur.nl
www.eur.nl/ese