UNIVERSITA' VITA-SALUTE SAN RAFFAELE

CORSO DI DOTTORATO DI RICERCA
IN NEUROSCIENZE COGNITIVE

NEUROPSYCHOLOGICAL TASKS AS PSYCHOMETRIC MEASURES: ISSUES OF RELIABILITY PARADIGMS AND COGNITIVE PSYCHOMETRICS. THE STRANGE CASE OF THE STOP-IT TASK AND IOWA GAMBLING TASK.

DoS: Chiar.mo Prof. Andrea Fossati
Tesi di DOTTORATO di RICERCA di Giulia Gialdi
matr. 015591

Ciclo di dottorato XXXV°
SSD: M/PSI-03

Anno Accademico 2021/2022

CONSULTAZIONE TESI DI DOTTORATO DI RICERCA

Il/la sottoscritto/*I* GIULIA GIALDI

Matricola / *registration number* 015591

nata a/ *born at* CERNUSCO SUL NAVIGLIO

il/*on* 16/03/1994

autore della tesi di Dottorato di ricerca dal titolo / *author of the PhD Thesis titled*

X AUTORIZZA la Consultazione della tesi / AUTHORIZES *the public release of the thesis*

□ NON AUTORIZZA la Consultazione della tesi per ...... mesi /DOES NOT AUTHORIZE *the public release of the thesis for ...... months*

a partire dalla data di conseguimento del titolo e precisamente */ from the PhD thesis date, specifically*

Dal / *from* ....../....../...... Al / *to* ....../....../......

Poiché /*because*:

□ l'intera ricerca o parti di essa sono potenzialmente soggette a brevettabilità/ *The whole project or part of it might be subject to patentability;*

□ ci sono parti di tesi che sono già state sottoposte a un editore o sono in attesa di pubblicazione/ *Parts of the thesis have been or are being submitted to a publisher or are in press;*

□ la tesi è finanziata da enti esterni che vantano dei diritti su di esse e sulla loro pubblicazione/ *the thesis project is financed by external bodies that have rights over it and on its publication.*

E' fatto divieto di riprodurre, in tutto o in parte, quanto in essa contenuto / *Copyright the contents of the thesis in whole or in part is forbidden*

Data /Date 06/11/2022 Firma /Signature ......Giulia Gialdi......

DECLARATION

This thesis has been:

- composed by myself and has not been used in any previous application for a degree. Throughout the text I use both ‘I’ and ‘We’ interchangeably.
- has been written according to the editing guidelines approved by the University.

Permission to use images and other material covered by copyright has been sought and obtained.

All the results presented here were obtained by myself.

All sources of information are acknowledged by means of reference.

DEDICATION

Ça va sans dire,

👻

👨‍👩‍👧

🐧🦦

ABSTRACT

The aim of the present project was to evaluate the basic psychometric proprieties of two widely used tasks in neuropsychology, namely the Stop-Signal Task and the Iowa Gambling Task.

This project consisted of five independent studies (N = 207; N = 114; N = 174; N = 134; N = 158), composed of Italian community dwelling adult participants, who volunteered to take part in the studies. Specifically, these studies aimed at addressing the problems presented in literature on the reliability and validity of the Stop-Signal Task and the Iowa Gambling Task, also considering more advanced method for computing relevant indices (at least in relation to the Stop-Signal Task). In particular, the studies aimed at evaluating the convergent validity between the Stop-Signal Task and the Iowa Gambling Task and self-report measures of disinhibition, and other computerized behavioral tasks. Moreover, the studies aimed at assessing the temporal stability of the Stop-Signal Task and the Iowa Gambling Task with a three-months test-retest paradigm.

The results of this project may advance our knowledge on the reliability and convergent validity of the Stop-Signal Task and the Iowa Gambling Task.

TABLE OF CONTENTS

Acronyms and Abbreviations

ADHD	Attentional Deficits Hyperactivity Disorder
AMPD	Alternative Model of Personality Disorder
APA	American Psychological Association
APD	Antisocial Personality Disorder
BART	Balloon Analogue Risk Task
BDD	Body Dysmorphic Disorder
BEESTS	Bayesian Ex-Gaussian Estimation of Stop-Signal (RT distributions)
BIS-11	Barratt Impulsiveness Scale-11
BPA	Bayesan Parametric Approach
BPD	Borderline Personality Disorder
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5th edition
EF	Executive Function
EV	Expectancy Valence
FIX	Fixation point
HYP	Hypochondriasis
ICC	Intraclass Correlation Coefficient
IGT	Iowa Gambling Task
ImpSS	Impulsive Sensation Seeking Scale
LATER	Linear Approach Threshold with Ergotic Rate
MAP	Maximum a posteriori
MCMC	Markov chain Monte Carlo
MIC	Mean Inter-item Correlation
ms	Milliseconds
NU	Negative Urgency
OCD	Obsessive Compulsive Disorder
OSF	Open Science Framework
PD	Personality Disorder
PEBL	Psychology Experiment Building
Pers	Perseveration
PG	Pathological Gambling

PID-5	Personality Inventory for *DSM-5*
PID-5-SF	Personality Inventory for *DSM-5*-Short Form
Prem	Premeditation
PU	Positive Urgency
PVL	Prospect Valence Learning
RI	Response Inhibition
RL	Reinforcement-Learning
RT	Reaction Time
SIB	Self-injurious Behaviors
SP	Skin Picking
SS	Sensation Seeking
SSD	Stop-Signal Delay
SSP	Stop-Signal Paradigm
SSRT	Stop-Signal Response Time
SST	Stop-Signal Task
UPPS-P	UPPS-P Impulsive Behavior Scale
WTF	With Trigger Failure

List of Figures

SECTION I: STOP-SIGNAL TASK

SECTION II: IOWA GAMBLING TASK

List of Tables

SECTION I: STOP-SIGNAL TASK

SECTION II: IOWA GAMBLING TASK

Rationale and Objectives

Before entering into the descriptions of the studies presented in this research project, I want to introduce the logic behind them.

In the following pages, I will focus on the basic psychometric characteristics of well-known and widely used tasks developed to assess response inhibition and impulsive decision-making. These tasks are frequently incorporated as primary measures in large scale studies (e.g., Bissett et al, 2021). However, it happened that sometimes their psychometric properties are not sufficiently investigated or taken for granted. For instance, in 2020, Elliott and colleagues carried out an extensive meta-analysis and showed that common task-fMRI measures ware not suitable for brain biomarker discovery or for individual-differences research because of their low reliability indices. Problems with reliability of measure lie at the heart of reproducibility and replicability of findings (Nosek et al, 2022), and research that embraces the psychometric rigor necessary to generate clinically actionable knowledge could represent a first step to move forward (e.g., Elliott et al, 2020; Elliott et al, 2021). For instance, it may be useful to remember that different measures of reliability measure different types of measurement consistency and are not necessarily highly correlated (e.g., Chmielewski & Watson, 2009). Although some tasks were explicitly designed with cognitive models in mind (e.g., the Stop-Signal Task), or the interpretation of their results hinges on assumptions about the performance of healthy participants (e.g., the Iowa Gambling Task), the reliability of the scores resulting from these models or the empirical sustainability of the assumptions is rarely examined (e.g., Bissett et al, 2021; Steingroever et al, 2013).

The studies presented here tried to address these issues by assessing the psychometric properties of the Stop-Signal Task and the Iowa Gambling Task.

SECTION I: STOP-SIGNAL TASK

1. Introduction

The Stop-Signal Task (SST) has been used at assessing the cognitive-control mechanisms involved in inhibitory behavior (Matzke et al, 2018). The parameter describing the stopping process is the Stop-Signal Response Time (SSRT), which has been much criticized since it can be strongly distorted. After presenting an overview of the present literature, the overall objective of this section is to present the psychometric proprieties and measurement models of the improved version of the SST.

In this first Section, several works about the Stop-Signal Task will be presented. Specifically, Chapter 2 presents bibliographic research where response inhibition and stopping process are described (Chapter 2.1). Then (Chapter 2.2), inhibition paradigms used in this project are presented and described in detail (i.e., Go/No-Go Paradigm [Bezdjian et al, 2009] and Stop-Signal Task [Verbruggen et al, 2019]). The Chapter ends with a review of the different Methods of Response Inhibition, that explain the inhibition process (i.e., The Independent Horse Race Model [Logan, 1981], The Interactive Horse Race Model [Boucher et al, 2007], and The Hanes-Carpenter Model [Hanes & Carpenter, 1999]). Chapter 3, focuses on the most used Frequentist Estimations of SSRT, including the Mean Method (Logan & Cowan, 1984), the Integration Method (Logan, 1981), and the Colonius Method (Colonius, 1990). A different SSRT estimation will be presented in Chapter 4, which covers Bayesian estimation methods (i.e., Individual Bayesian Parametric Approach and Hierarchical Bayesian Parametric Approach). Finally, two studies on the psychometric proprieties of the Stop-Signal Task will be presented: the first study aims at assessing the convergent validity between the SST and self-report measures of disinhibition and at testing if one of the different models proposed for estimating the SSRT (e.g., the mean method, the Bayesian parametric method, etc.) would show larger convergent validity with self-reports of disinhibition (i.e., the UPPS-P Impulsive Behavior Scale [UPPS-P; Cyders & Smith, 2007], Impulsive Sensation Seeking Scale [ImpSS; Zuckerman et al, 1991], Barratt Impulsiveness Scale-11 [BIS-11; Patton et al, 1995] and Personality Inventory for *DSM-5* [PID-5; Krueger et al, 2012]); moreover, this study aims at assessing also the convergent validity between the SST and the Go/No-Go Task. Also in this study, SSRT scores will be computed using different measurement models to show their convergence with lab tasks. The last study aims at assessing the test-retest stability of the SST with a three-months test-retest reliability paradigm. In this study I relied on three different approaches to assess the reliability estimations: test-retest reliability (i.e., Spearman *r* coefficient), Intraclass Correlation Coefficient (ICC), and internal consistency estimations.

This section will be concluded with future direction (Chapter 7) and a general conclusion about the Stop-Signal Task (Chapter 8).

2. Literature Review

2.1 Inhibition

2.1.1 Concept of inhibition

The concept of inhibition was described for the first time at the end of the 19^{th} century, by William James, the father of modern psychology, that defined inhibition as "*an essential and unremitting element of our cerebral life*" (James, 1890, p. 583). However, it's only in the second half of this century that the application of the terminology of inhibition became common in neuroscience (Smith, 1992).

Before 19^{th} century, according to Macmillan (1992), inhibition was considered a form of excitement or its opposite (Bari & Robbins, 2013; see also Melzter, 1899). Instead, with the start of the 19^{th} century, and its appearance in psychology (Gall, 1835), inhibition was used to describe numerous and dissimilar experiences (Bari & Robbins, 2013).

Indeed, there are two principal applications of the concept of inhibition: one related to the nervous system (Miller & Cohen, 2001), and a second concepts related to thought and behavior.

Thus far, Bari and Robbins (2013) tried to classify the inhibitory processes, since the relationship between inhibition of mental processes and of physical responses is not entirely clear. As shown in Figure 2.1, the authors divided inhibition or inhibitory control in cognitive inhibition (i.e., the stopping of mental processes as memories, thoughts, perceptions, emotions, etc.) and behavioral inhibition (i.e., the stopping of a manifest behavior). Behavioral inhibition in turn is divided in response inhibition (impulsive action), deferred gratification (impulsive choice) and reversal learning (inflexibility, compulsivity).

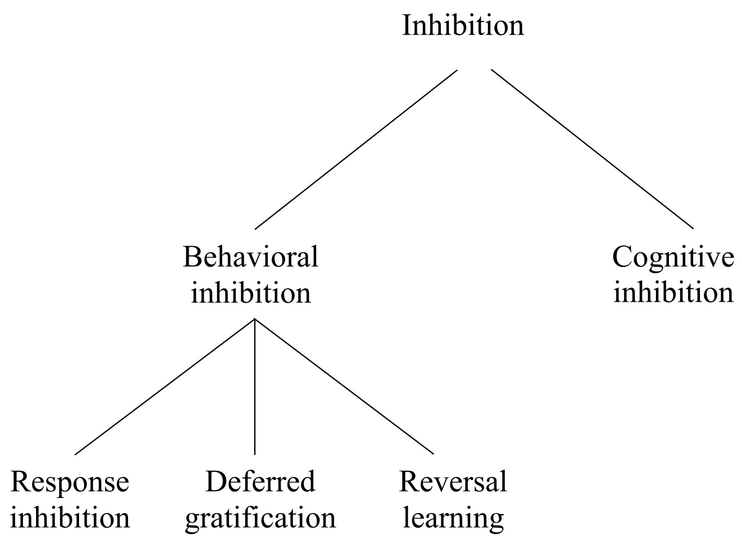


Figure 2.1. Classification of inhibitory control
Adapted from Bari & Robbins, 2013

A recent study of Hendry and colleagues (2022), suggested to describe inhibition by describing its effects, that can be produced by different contexts or tasks. Thus, the authors identified two types of inhibition: directed global inhibition and competitive inhibition (see also, Munakata et al, 2011). Directed global inhibition is a response, for example, to an external prohibition (e.g., "Do not walk on the grass"), whereas during competitive inhibition the subject execute an alternative response to an inhibited action (e.g., when the traffic light turns yellow, and the driver must press on the brake rather than on the accelerator).

However, thus far, there is little consensus about what inhibition is, even in the field of psychology. Werner and colleagues (2022) concluded that the difficulty to define inhibition is because it is included in several psychological areas (e.g., cognitive, social, personality, developmental, clinical psychology, and neuroscience). More recently, indeed, these authors (Werner et al, 2022) considered three interrelated issues about inhibition: (a) there is an increase of different operationalizations about inhibition, because different subfields disagree on how to define it; (b) many areas have lowered the threshold for what counts as inhibition, considering "inhibition" as a construct; and (c) the term “inhibition” has been overextended to concepts that can be more parsimoniously explained by other constructs (Werner et al, 2022).

Another issue about the definition of inhibition concerns if inhibition should be considered as a component of the executive functions or not (see for example, Bari & Robbins, 2013). For example, people need to pay attention to environmental cues that suddenly change to inhibit the action. Indeed, if inhibition is truly needed to implement cognitive control there is no direct evidence (Cohen, 2017).

Lastly, in the cognitive neuroscience literature (e.g., Bari & Robbins, 2013), there are two forms of inhibition: a slower (namely also cool or motor impulsivity; Castellano et al, 2006; Eagle et al, 2009) and a fast form (namely also hot or choice impulsivity; Castellano et al, 2006; Eagle et al, 2009). The first one involves deliberation and the fulfillment of desires with the consideration of negative consequences; whereas the second form do not consider enough the possible negative outcomes (for a review of authors, see Bari & Robbins, 2013).

2.1.2 Response Inhibition

As widely described in the previous paragraph, inhibition is commonly used to describe a wide variety of functions (Kok, 1999); however, in this manuscript, it is considered to the deliberate, controlled suppression of responses. The ability to control impulses and suppress responses when are no longer necessary is the foundation for the possibility to adapt ourselves within an ever-changing environment. This ability is known as Response Inhibition (RI). RI is an important component of the Executive Function (EF; Miyake et al, 2000), and specifies the ability to suppress a dominant, automatic or prepotent responses when they are contextually inappropriate and no longer necessary (Skippen et al, 2019). According to Skaggs (1929) since response inhibition is most of the time voluntary or involving some degree of consciousness, it involves motor-related brain areas.

Response Inhibition, an essential cognitive control process, has been theoretically associated with impulse control (Bari & Robbins, 2013). Therefore, response inhibition becomes necessary when individuals have to adjust their behavior to changing conditions, changing goals, and changing world (Logan, 1994; Matzke et al, 2018; Schachar et al, 2007). For example, RI is necessary when someone is driving, and they have to suddenly slow down because an animal is coming toward them. In addition, RI becomes indispensable when we have to stay focused on a task, ignoring the distractors in the environment (Johnstone, et al, 2007).

Existing research has shown that response inhibition changes throughout a lifetime; for example, developmental studies have shown that RI has an inverted U-shape throughout its lifespan: it is higher during childhood, and it slows down again in older age (e.g., Bedard et al, 2002; van de Laar et al, 2010).

Response inhibition may have important implications for typical and atypical developmental trajectories. Indeed, behavioral consequences in healthy and pathological brain can be measured with observable indices of inhibitory processes. For example, inhibition lack can have relevant implications for the outcome of treatment of people with different psychopathological disorders and problematic behaviors (e.g., Attentional Deficits Hyperactivity Disorder [ADHD], obsessive-compulsive disorder, pathological gambling, substance use disorder, schizophrenia, etc.; Gut-Fayand et al, 1999; Nederkoorn, et al, 2007). In ADHD children impulsivity is often manifested as the inability to wait in a variety of situations and as the tendency to interrupt others' conversations, or to respond before the end of the question (DSM-5, p. 68; APA, 2013). For example, there are numerous study that revealed that individuals with ADHD perform worse on inhibition task compared to individuals without ADHD (e.g., King et al, 2007; Lijffijt et al, 2004; Lijffijt et al, 2005). According to Bari and Robbins (2013), impulsivity is often described also as the main behavioral characteristic of drug abusers, schizophrenic patients, and obsessive compulsive disorder [OCD; see also Bari & Robbins, 2013; Chamberlain et al, 2005; Dumais et al, 2011; Gut-Fayand et al, 2001).

Lack in inhibitory control may also negatively affect the lives of healthy adult individuals (Bari & Robbins, 2013; Fuster, 2008). However, certain impulsive behaviors are not necessarily disadvantageous, but be adaptive (Block, 2002; Dickman, 1990; Harnishfeger & Bjorklund, 1994).

2.2 Inhibition Tasks

Inhibition behavior can be measured in a several ways in individuals, for example by self-report questionnaire, such as the Barratt Impulsiveness Scale, by observing behavior in natural setting, or with behavioral measures of impulsivity (Bari & Robbins, 2013). Moreover, inhibitory control was suggested to be a heterogeneous construct, that allow researchers to include a wide range of tests and tasks to measure it (López-Caneda et al, 2014).

Some of the computerized tasks mostly used to measure Inhibition are the Stroop Task (Stroop, 1935), the Antisaccade Task (Hallett, 1978), and the Stop-Signal Task (Logan, 1994). Briefly, during the Stroop Task (Stroop, 1935) participants watch a sequence of color words, in incongruously colored ink (e.g., the word "yellow" printed in green ink). During this task, individuals are asked to name the ink, suppressing the word meaning. In the Antisaccade Task (Hallett, 1978), once the fixation point has disappeared, participants have to direct their gaze in the direction opposite the cue that appears (e.g., the cue points to the right, participants have to look left). Finally, during the Stop-Signal Task (Logan, 1994), participants are asked to press a key corresponding to stimulus, and to inhibit their response when a different stimulus or an acoustic signal appears. All these tasks, require deliberate stop for a response that is moderately automatic.

However, the Stop-Signal Task has been poorly investigated in neuropsychological contexts; but a similar task (i.e., the Go/No-Go Task) has been well studied in neuropsychological contexts (e.g., Casey et al, 1997; Kiefer et al, 1998). For this reason, in order to evaluate response inhibition in laboratory setting, in this project two different paradigms were used: the Go/No Go Paradigm (Bezdjian et al, 2009) and the Stop-Signal Paradigm (Logan & Cowan, 1984). These two paradigms are the typical tasks used to measure the ability to inhibit a response. Both tasks are based on the execution of a motor response to visual stimuli, while on some trials the stop signal (e.g., visual or acoustic) instructs participants to inhibit the response. The difference between these two tasks is the temporal presentation of the stop signal (Eagle et al, 2008; Schachar et al, 2007): during the Go/No-Go Task, the stop signal is presented instead of the go stimulus (Simmonds et al, 2008), while on the Stop-Signal Task is presented after the go stimulus (Verrbruggen & Logan, 2008). This difference has led several researchers to wonder whether these two tasks measured the same concept. In the next paragraphs, the platform used for the administration of the tasks and the tasks themselves used for this section of the project will be discussed.

2.2.1 Psychology Experiment Building Language platform

Some of the tasks used were administered with the Psychology Experiment Building Language platform (PEBL; Muller, 2013; Mueller & Piper, 2014). Traditionally, Executive Function (EF) is thought to include three main subcomponents: (a) updating (i.e., constant monitoring and rapid addition/deletion of working memory contents), (b) shifting (i.e., switching flexibly between tasks or mental sets), and the main point for this project, (c) inhibition (i.e., deliberate overriding of dominant or prepotent responses). Usually, EF has been assessed using *paper-and-pencil* methods of administration, but computerized administration offers potential advantages ensuring ease of administration, standardization of presentation across test sessions, automatic scoring, randomization of stimuli and tasks, the precision of timing, and the opportunity to transport a large number of potential tests in one laptop (e.g., Collerton et al, 2007; Nicholl et al, 1995).

Despite the fact that computerized cognitive assessments have been shown to be feasible and useful also in the oldest-old age group (e.g., Collerton et al, 2007) and they are widely used in some settings (e.g., pharmacological studies; Simpson et al, 1991), they are less used in clinical practice (e.g., Morris et al, 2000). Based on these problematics, Mueller (2010; 2012) developed the Psychology Experiment Building Language (PEBL).

Researchers and clinicians can create, run, and share behavioral tests using PEBL's platform, that is a free, open-source software system (Mueller & Piper, 2014). Since PEBL is an open system, users are to install the software and share their experiments with others without worrying about licenses (Mueller & Piper, 2014). A major advantage of PEBL is that it is an open-source program, and this led to some advantages: for example, researchers and clinicians can inspect, modify and redistribute the source code, so that experiments can be verified and modified by other researchers (Mueller & Piper, 2014). Generally, PEBL experiments are run through the software launcher, letting users select aspects of the test, such as the way it is conducted, the ability to use “experiment chains” to not interrupt the administrations when a task end. Indeed, in this project a chain with two tasks used in this project (i.e., Balloon Analogue Risk Task and Iowa Gambling Task; see paragraph 5.2 of the second Section) has been created. For the studies assessed in this Section it was not necessary the creation of a chains since the only task administered with PEBL was the Go/No-Go Task. All the tasks were administered in Italian language, randomly and with full screen.

A screenshot of the PEBL launcher is shown in Figure 2.2.

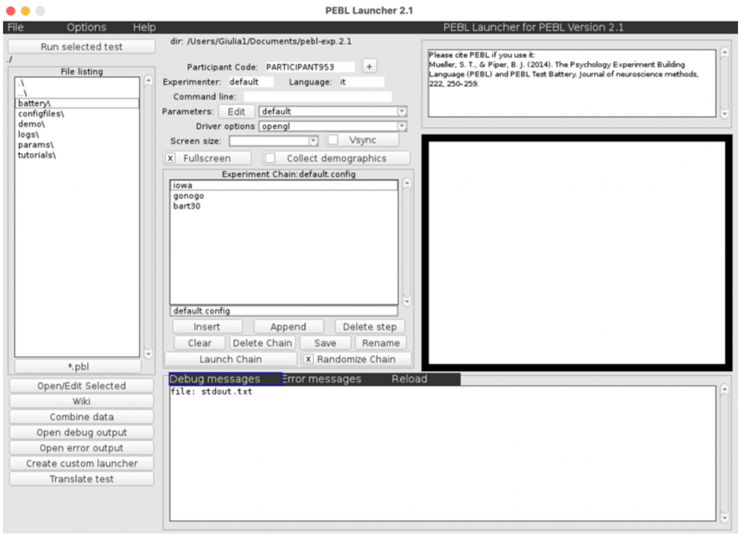


Figure 2.2. Screenshot of the PEBL Launcher.
This platform which allows users to navigate the test battery, run specific tests and execute a “chain” of tests appropriate for a particular study.

The entire test battery presented in PEBL platform focuses on computerized cognitive tests, including paradigms that involve memory, attention, and executive control, for a total of approximately 70 tests and test variants (Mueller, 2010; Mueller, 2012). Notwithstanding the number of tasks included, Piper (2012) had provided evidence for the reliability and validity of the tasks included in the PEBL battery. For this project, all PEBL test battery tasks were completed on an IBM-compatible laptop personal computer. Table 2.1 shows the tasks used in this project available on the PEBL software.

Table 2.1. Description of the three tasks used in this project available on the PEBL Test Battery.

Test name	Directory name	Description	References
Balloon Analog Risk Task	BART	Inflate a balloon for rewards	Lejuez et al, 2002
Bechara's "Iowa" Gambling Task	iowa	Choose decks with different rewards and penalties	Bechara et al, 1994
Go/No-Go task	gonogo	Respond to one stimulus; ignore second stimulus	Bezdjian et al, 2009

Note. Adapted from Muller & Piper, 2014

This platform has been used and can be used in different discipline studies, and Mueller and Piper (2014), in their article, have illustrated part of these disciplines, including for example Artificial Intelligence, cognitive psychology, neurology, clinical psychology cognitive neuroscience, etc. (to a review see Mueller & Piper; 2014)

2.2.2 Go/No-Go Paradigm

The Go/No-Go Paradigm allows researchers to describe symptoms of both impulsivity (i.e., difficulties in inhibiting the response), and inattention (i.e., difficulties in sustained attention). The Go/No-Go Task (Bezdjian et al, 2009) allows evaluating these two aspects. For example, to explore the disinhibitory nature of ADHD, numerous studies have utilized laboratory measures such as the Go/No-Go task (Nigg, 2001). The Go/No-Go Task has a long story (Donders, 1868; Donders, 1969), and its application is increasing in several fields, as bilingualism (Dijkstra et al, 2000), neuropsychology (Goldberg et al, 2001), speech production (Hino & Lupker, 2000), recognition memory (Boldini et al, 2004), etc. Gordon and Camarazza (1982) first applied this procedure to the lexical decision task. The use of the Go/No-Go procedure minimizes response confusion and errors in response selection, by making selection simpler than a two-choice procedure (Gordon & Camarazza, 1982).

In detail, during the Go/No-Go Task, participants are asked to respond when a target stimulus (i.e., letters P or R) appears. During this task, participants watch a sequential presentation of the two letters and have to respond to one of them by pressing the right shift key on the keyboard. The presentation begins with a 2x2 array with four blue stars (one in each square of the array; as shown in Figure 2.3), and the letters appears for a duration of 500 milliseconds with an inter-stimulus interval (ISI) of 1,500 ms.

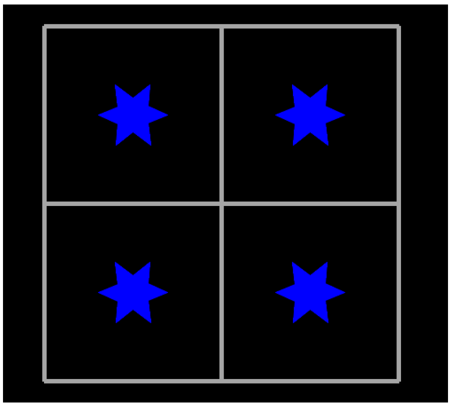


Figure 2.3. Screen presentation of the Go/No Go Task

The Go/No-Go Task consists of two opposite conditions, with 160 trials each. In the first condition, participants are asked to press the right shift key in response to the letter P ("Go" trials) and withhold their response to the letter R ("No-Go" trials); on the contrary, in the second condition participants are asked to make a response to the letter R ("Go" trials) and withhold their response to letter P ("No-Go" trials). Figure 2.4 shows the two different conditions.

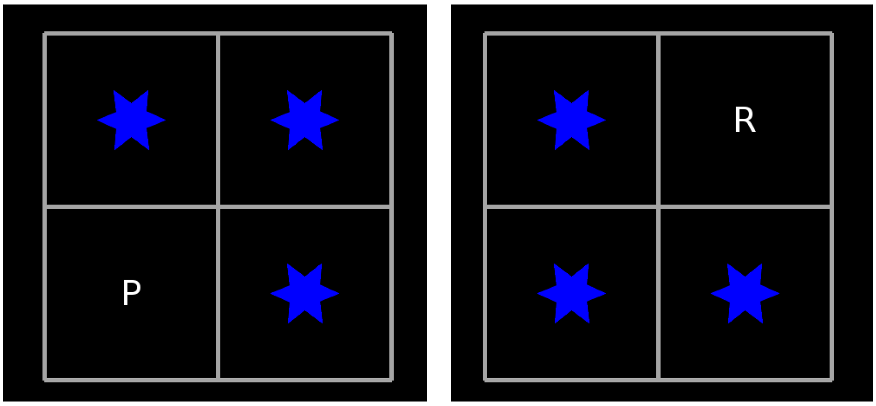


Figure 2.4. Examples of P-condition (left panel) and R-condition (right panel)

Prior to both conditions, participants are administered a few practice trials to ensure the task was fully comprehended. The ratio of Go trials and No-Go trials is equal in both conditions, and it is 80:20. Figure 2.5 represents the sequence of events in the Go/No-Go Task. In this example, participants respond to the letter P (by pressing the right shift key on the keyboard) in the Go trial and inhibit their response when the letter R appears in the Stop trial.

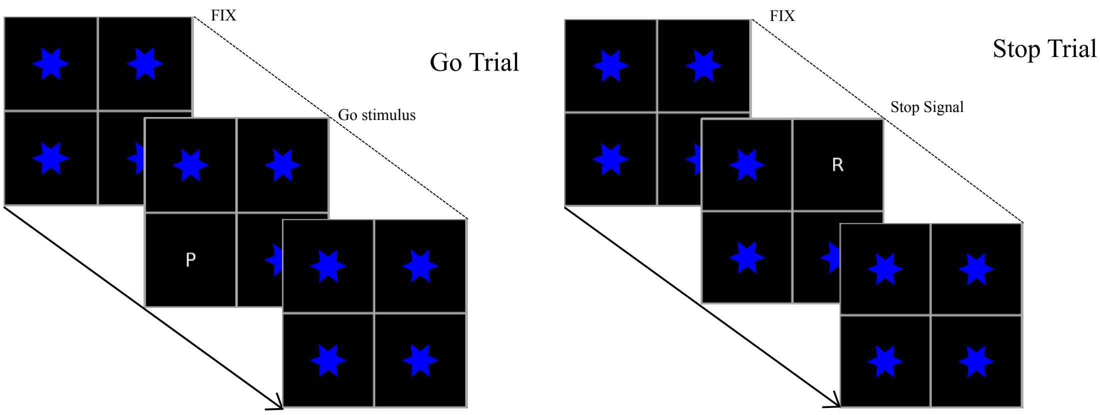


Figure 2.5. Representation of the sequence of events during the Go/No-Go task. FIX = fixation point.

The behavioral performance of the task was assessed by calculating four values both in the whole performance and in each condition:

1) correct responses to the target letter (i.e., participants correctly pressed to the "Go" trials);
2) correct reactions to the No-Go letter (i.e., participants correctly inhibited responses to the "No-Go" trials);
3) errors of omission: responses that occur when no response is required, which are typically considered an indicator of inattention (Barkley, 1991; Halperin et al, 1991).
4) errors of commission: responding incorrectly to the No-Go letter, which is considered an indicator of impulsivity (Barkley, 1991; Halperin et al, 1991).

In addition, reaction time (RT) variability to the "Go" trials was assessed for each participant. For the purpose of this research project, the Go/No-Go Task in the PEBL library has been used in its Italian translation (Fossati et al, 2018).

2.2.3 Stop-Signal Paradigm

The Stop-Signal Paradigm (SSP) is frequently used to assess response inhibition and the most used is the Logan and Cowan paradigm (1984). A wide range of psychological disciplines have used the SSP, over the past decades, including clinical psychology, developmental psychology, experimental psychology, psychopathology, neuropsychology and studies of individual differences studies (e.g., Matzke et al, 2018). Moreover, it is applied to study inhibition deficits in clinical conditions, such as schizophrenia (Matzke et al, 2017), ADHD (Matzke et al, 2019), Obsessive Compulsive Disorder (OCD; Verbruggen & Logan, 2008), and in drug and alcohol-users' studies (e.g., Monterosso et al, 2005; Goudriaan et al, 2006).

To study the underlying process of the Stop-Signal Paradigm, the Stop-Signal Task (SST) is often used. The Figure below show the wide fields where the stop-signal is flourishing (right panel) and the number of articles citing "stop-signal task" per years (left panel).

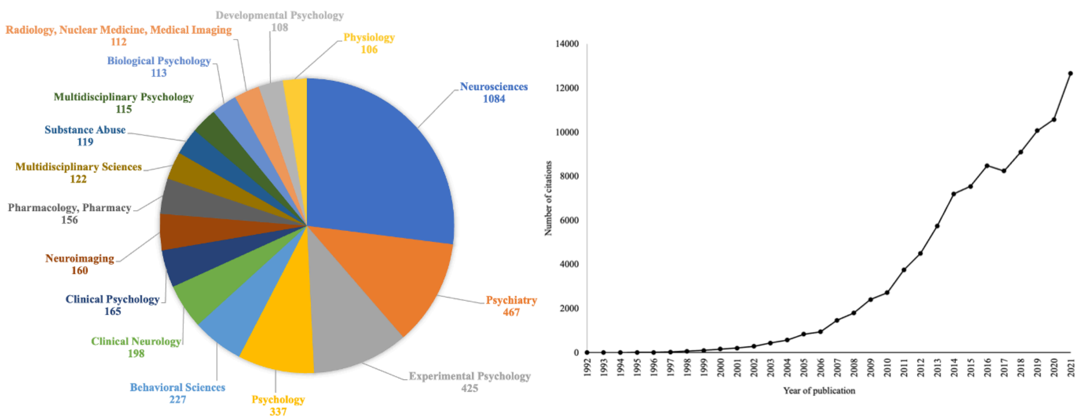


Figure 2.6. Number of stop-signal publications per area (right panel) and number of citations of "stop-signal task" per year (left panel).
Source: Web of Science, 14/07/2022. Search term: "topic = stop-signal task".

The participants in the SST perform a two-choice visual response task. Depending on the paradigm, they have to respond to the color, shape, or direction of the stimulus. Occasionally, this primary stimulus is interrupted by a stop-signal, that can be a sound or any other stimulus (e.g., visual stimulus), that instructs participants to inhibit their response and not to respond (Logan & Cowan, 1984). The time between the presentation of the primary stimulus and the presentation of the stop-signal is called Stop-Signal Delay (SSD). The latency of the stop process is called Stop-Signal Reaction Time (SSRT), and it is the main interest for researchers.

This task can be administered in two different ways: with the fixed SSDs procedure or with the tracking procedure (Logan & Cowan, 1984). During the first procedure, the stop-signal is executed with a fixed delay on all stop trials. Despite the simplicity of use of the fixed SDDs procedure, it has few limitations (Band et al, 2003; Congdon et al, 2012; Matzke et al, 2018; Williams et al, 1999), and for this reason the most common setting for the administration of the SSD is the tracking method, where it is adjusted by 50ms after every trial. This procedure is also called the one-up/one down procedure (Logan et al, 1997; Verbruggen & Logan, 2009; Verbruggen et al, 2013): after every successful inhibition, the SSD is increased by 50ms, and, on the contrary, when the inhibition is unsuccessful, the SSD is decreased by 50ms. Using this procedure, the probability that participants respond to the stop-signal is closely to 50% [p(respond|signal) $\approx$ 0.50] (Verbruggen et al, 2013). Unlike the fixed SSDs, with the tracking procedure participants cannot predict when the stop-signal is executed, and thus they cannot wait for its appearance. This is one of the advantages of this procedure and provide more precise estimates of the Stop-Signal Task values (see paragraph e of this Section). In addition, during the presentation of the task with the instruction, participants are asked not to wait for the stop-signal to occur (Verbruggen et al, 2013), also because of the use of the tracking procedure and because it is possible to invalidate the performance. Because of its limitation, that use fixed stop-signal delays require a high number of trials, whereas methods that use the tracking procedure for SSD require a smaller number of observations for still accurate and valid estimation of SSRT (Band et al, 2003; Congdon et al, 2012; Matzke et al, 2018; Williams et al, 1999). However, to obtain realistic performance, researchers are recommended to present participants with approximately 120–150 go-trials and 40–50 stop-signal trials with the tracking procedure (Verbruggen & Logan, 2009).

Following the recommendation of Verbruggen and colleagues (2019), in this project the tracking procedure has been used.

For this project, I relied on an open-source stop-signal paradigm to improve the replicability of the findings (Stop-it Task - The jsPsych version; Verbruggen et al, 2019). In addition, Verbruggen and colleagues (2019) published the guidelines for the administration of the task, which states that if the task is not used adequately, it falls to its own success. Therefore, the authors decided to publish twelve recommendations to improve the quality of future stop-signal research, that should be heterogeneous on how the task is executed, how SSRT is estimated and how to report results (Verbruggen et al, 2019).

Following the established procedure, during this task, participants had to discriminate between a green arrow pointing to the left and a green arrow pointing to the right (Verbruggen et al, 2019). When a go-trial appears (i.e., on the 75% of the trials) participants have to respond quickly and accurately, by pressing the correct key on the keyboard (i.e., the left arrow key for the green arrow pointing to the left and the right arrow key for the green arrow pointing to the right). Conversely, on stop-signal trials (i.e., the 25% of the trials, e.g., Recommendation 3 [Verbruggen et al, 2019]), the apparition of the arrow, is replaced by “XX”, after a variable delay (i.e., Stop-Signal Delay; SSD). This signal instructed participants to not respond to the arrows.

Figure 2.7 represent the sequence of events in the Stop-Signal Paradigm. In this example, participants respond to the arrows by pressing the correct arrow key in the Go Trials and inhibit their response when “XX” appears after the arrow’s presentation. When the stop signal is presented near the presentation of the go stimulus (i.e., short SSD), participants can successfully inhibit their response, but on the contrary, if the stop-stimulus is presented close to the moment of the execution of the response (i.e., long SSD) they will have difficulties in inhibiting their response.

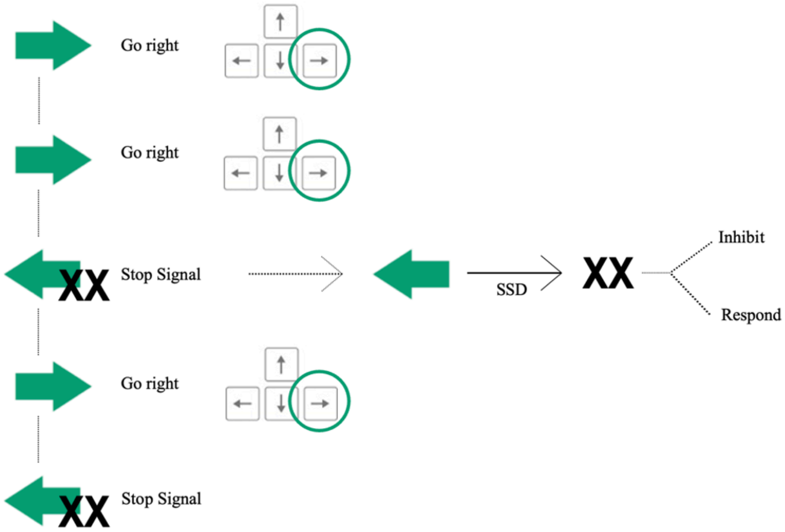


Figure 2.7. Sequence of events in the stop-signal paradigm.
In the go trials, participants respond to the direction of the arrow (the go stimulus, e.g., a "left arrows" requires a left response). On a minority of trials, the go stimulus is followed by two XX after a variable stop-signal delay.

This open-source task is very easy to use and to administer it. Moreover, it can be easily and freely download from the following Open Science Framework (OSF) link: https://osf.io/wuhpv/ (Verbruggen, 2019). This version of the Stop-signal task is based on the jsPsych (De Leeuw, 2015) and can be installed on local computer or on a web server. Thus, for the purpose of the present project, also this task was completed on an IBM-compatible laptop personal computer. The main advantages of this Stop-Signal Task version are (a) the version is platform-independent, (b) it can be used for both offline and online studies, and (c) for basic use, there is no need for additional programming (Verbruggen, 2019).

The administration starts with a welcome message and the informed consent. After having accepted the informed consent, participants are required to enter their age and gender, and the experiment switches to fullscreen mode. At this point, the task instructions are presented in their Italian translation. For this project, experiment consisted of two phases: a practice phase (one block of 32 trials) and an experimental phase, composed by 6 blocks of 64 trials each. To do this, I had to modify the downloaded task of four blocks of 64 trials each, by adding two more blocks, with a result of 384 trials. Using more trials, allows researchers to extend the scope and the applicability of the Stop-Signal Paradigm to the study of response inhibition in the context of difficult choices (Heathcote et al, 2019; Skippen et al, 2019).

In both practice and experimental phases, each trial starts with the presentation of the fixation sign, which is replaced by the go stimulus after 250ms. The stimulus remains on the screen until the response, or until 1,250ms (i.e., the maximal RT) have passed (Verbruggen, 2019). During the practice phase, immediate feedback is presented, to instruct participants to modify their response:

- “incorrect response” when participants execute the incorrect response on go trials (i.e., choice error); for example, participant press the right arrow key, when have to respond by pressing the left arrow key;
- “too slow” when participants do not respond on go trials (i.e., go omission);
- “too fast” when participants press a response key before the go stimulus is presented (i.e., premature response)
- “remember: try to stop” when participants execute a response on a stop trial (i.e., unsuccessful stop or commission errors)

After the practice phase, the immediate feedback is not presented anymore, but between each block, subjects have to wait for 15 seconds before they can start the next block, during which participants will receive information about their performance in the previous block, that include the mean RT on go trials, the number of go omissions, and the percentage of correct stops (Verbruggen, 2019). An example of the performance’s feedback is presented in Figure 2.8.

GO TRIALS:

Average response time = 482 milliseconds

Proportion missed go = 0.00 (should be 0)

STOP-SIGNAL TRIALS:

Proportion correct stops = 0.50 (should be close to 0.5)

You can take a short break, the next block starts in 7

Figure 2.8. Example of performance’s feedback after a block of the Stop-Signal Task

The resulting output file is a .csv file, that can be opened in Microsoft Excel or statistical-software packages, as R and SPSS. The data file consists of some of the following information: which go stimulus was presented (i.e., right or left), if a stop signal was presented or not, the stop-signal delay (in milliseconds), which go response was executed on the trial by the participant (i.e., right or left arrow, undefined), reaction time, and if the answer was correct or not. An example of the stop-signal task output is presented in Table 2.2.

Table 2.2. Example of the Stop-Signal Task output

BLOCK	TRIAL	STIM	SIGNAL	SSD	RESPONSE	RT	CORRECT
5	59	right	no	650	rightarrow	803	true
5	60	right	no	650	rightarrow	1000	true
5	61	right	no	650	rightarrow	848	true
5	62	left	no	650	leftarrow	818	true
5	63	left	no	650	leftarrow	1060	true
5	64	right	yes	650	undefined	null	true
6	1	right	yes	700	rightarrow	580	false
6	2	left	no	650	leftarrow	719	true
6	3	left	yes	650	leftarrow	782	false
6	4	left	no	600	leftarrow	793	true
6	5	right	no	600	rightarrow	739	true
6	6	right	no	600	rightarrow	933	true

Note.
BLOCK: number of blocks; TRIAL: number of trials; STIM: which arrow was presented; SIGNAL: if a stop signal was presented (yes) or not (no); SSD: stop-signal delay; RESPONSE: the key button pressed by the participant; RT: reaction time; CORRECT: correct answer (true) or not (false).

Finally, the program allows researchers also to do the analysis using the software R. The software uses Shiny, an open-source R package that provides a web framework for building web applications using R or Rstudio (https://posit.co/products/open-source/shinyserver/). In the Figure below is presented a screenshot of the Shiny app.

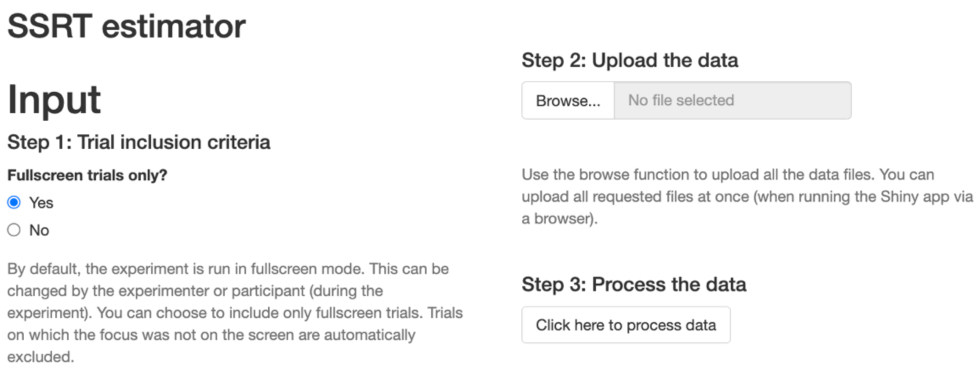
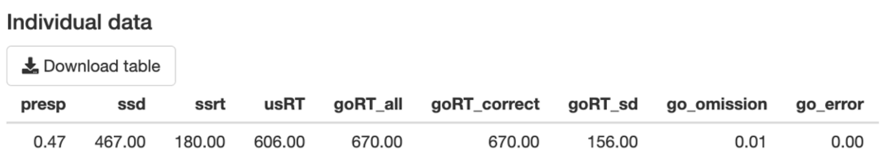


Figure 2.9. Screenshot of the Shiny app used for this project

As soon as the data are uploaded and processed, users will immediately get the outputs with a "Summary" panel showing the main dependent variables and the "Individual data" of the specific participant. In this Shiny App, users can be upload more than one file at a time, to obtain the results of all study participants quicky. In both panels, there are some of the following variables: the probability of responding on a stop trial [i.e., p(respond|signal), the stop-signal reaction time (estimated with the integration method with replacement of go omission; see paragraph 3.2 of this Section), the reaction time on unsuccessful stop trials and go trials, and omission and commission errors, as shown in Figure 2.10.


Individual data

Download table

presp	ssd	ssrt	usRT	goRT_all	goRT_correct	goRT_sd	go_omission	go_error
0.47	467.00	180.00	606.00	670.00	670.00	156.00	0.01	0.00

Figure 2.10. Output of the Shiny app.
Note. presp: p(respond|signal); ssd: stop-signal delay; ssrt: stop-signal reaction time; usRT: reaction time on unsuccessful stop trials (in ms); goRT_all: reaction time on go trials with a response (in ms), including choice errors; goRT_correct: reaction time on go trials with a correct response (in ms), excluding choice errors; goRT_sd: intra-subject variability in response latencies (including all go trials with a response); go_omission: proportion of go trials without a response; go_error: proportion of incorrect responses on go trials with a response (e.g., the go stimulus required a left response but a right response was executed).

Thanks to this paradigm, researchers are able to estimate the latency of the stop processes (i.e., the Stop-Signal Reaction Time; SSRT), based on the underlying horse-race model, which is what gives the stop-signal paradigm its popularity. Since it is an unobservable process, it cannot be directly calculated, but it can be estimated with several methods, that will be presented in the next chapter (see paragraph 3 and 4 of the present Section). The latency of the stop process has been used to study inhibitory deficits in different patient groups and in different research fields, as cognitive neuroscience, and studies investigating life span development, psychopathology, and individual differences (e.g., Monterosso et al, 2005; Goudriaan et al, 2006; Matzke et al, 2017: Matzke et al, 2018; Matzke et al, 2019; Verbruggen & Logan, 1990). For example, several studies have demonstrated that SSRT is elevated in younger children (Williams et al, 1999), older adults (Kramer et al, 1994), impulsive people (Logan et al, 1997), and children with attention-deficit/hyperactivity disorder (ADHD; e.g., Jennings et al, 1997; Schachar & Logan, 1990). In other words, these studies showed that clinical populations and experimental conditions benefit from the use of SSRTs for diagnosing deficient response inhibition.

To conclude this paragraph, it is important to report the advantages and disadvantages of the Stop-Signal Paradigm. The main advantage of this task is the possibility to investigate the processes of response inhibition in laboratory setting and evaluating the latency of the stop process (SSRT). However, the Stop-Signal Task is very long and requires a minimum of 20 minutes for administration (e.g., at least 20 minutes in my project for 384 trials; practice phase excluded), which can negatively affect the performance of the target population (for example, children with ADHD).

2.3 Methods of Response Inhibition

It is possible to estimate Stop-Signal Reaction Time because the performance in the Stop-Signal Task can be formalized as an independent horse race between the go process (i.e., the presentation of the stimulus), and the stop process (i.e., the presentation of the stop-signal) (Logan & Cowan, 1984; Logan et al, 2014). In this chapter, different independent horse-race models are briefly presented (it is possible to find an extended description of the model at the following link: https://osf.io/tudmw/; Gialdi, 2022).

2.3.1 The Independent Horse Race Model

For the assessment of Stop-Signal Reaction Time, Logan (1981) and Logan and Cowan (1984) proposed the Independent Horse Race Model, formalizing the response inhibition as a horse race between two independent processes (see, https://osf.io/tudmw/, Gialdi, 2022; Logan, 1981; Logan & Cowan, 1984; Matzke et al, 2018) : a go process, and a stop process, that start after a specific stop-signal delay (SSD).

Originally proposed by Vince (1948), the idea of the horse race between two processes precedes Logan and Cowan's research (1984). Vince (1948) observed that subjects had difficulties in inhibiting their response when the stop signal appeared after 50ms. The formalization of this process, however, came few years late, with Ollman's study (1973), where a subjective deadline for the go response was set: if the stop signal was prior to the subjective deadline, participants would be able to inhibit their response; contrariwise, if the stop signal appeared after the deadline, the response was an error of commission (i.e., response was erroneously emitted).

Logan and Cowan's model (1984) is consistent whit Ollman's idea (1973). However, based on Logan and Cowan's independent horse race model (1984), there is no understanding of the underlying process that produces the behavior in the Stop-Signal Paradigm, which is why does not explain the differences between subjects in inhibition performance (Matzke et al, 2018). Response inhibition and also aspects related to decision making can be modelled using the present model, since it only describes performance

(Matzke et al, 2018). Moreover, concepts about the nature and the speed of the stop processes can be tested using the model as well.

To facilitate the description of the model, I will describe the independent horse race model considering SSRT as constant during the entire administration of the task. Figure 2.11 (adapted from Matzke et al, 2018) shows a graphical representation of the model with constant SSRT. The gray area on the left shows the response rate (i.e., the probability of responding erroneously to the go stimulus), given a certain SSD; on the other hand, the white area on the right under the curve represents the probability of inhibiting the response (i.e., correct answer on go-trials).

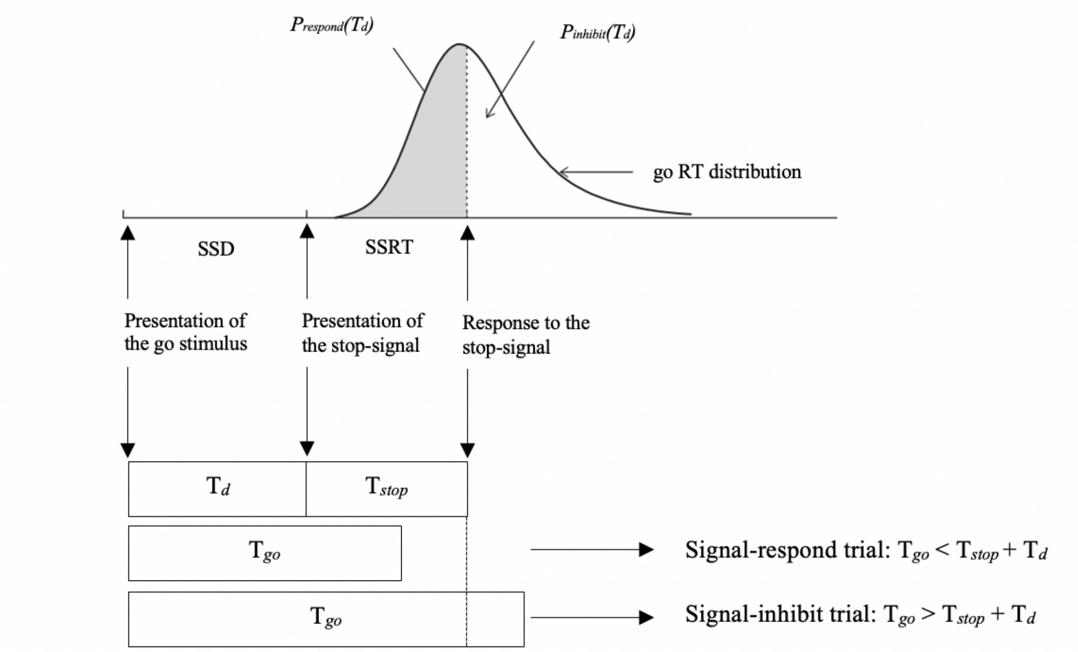


Figure 2.11. Representation of the independent horse-race model with constant stop-signal reaction time.
The Figure shows that response rate [i.e., $P_{respond}(T_d)$] and the probability of inhibition [i.e., $P_{inhibit}(T_d)$] are determined by SSD, the SSRT, and the go RT distribution. Go response is incorrectly emitted if $T_{go} < T_{stop} + T_d$ (i.e., signal-respond trial) and go response is successfully inhibited if $T_{go} > T_{stop} + T_d$ (i.e., signal-inhibit trial).
Adapted from Matzke et al, 2013.

The model also predicts that mean reaction time (RT) on commission errors increases when stop-signal delay increases and approaches to the mean RT on go trials (Matzke et al, 2018). The mean RT of commission error is necessarily faster than the mean RT of the go. When this does not happen, we have the first violation of the Independent Horse Race Model: participants with longer RT on commission errors than RT on go trials, must be excluded from the sample (see also paragraph 5.3 of the present Section; Skippen et al, 2019). Therefore, since the commission errors depend on the presentation of the stop-signal, and thus on the stop-signal delay, the vertical line moves to the right, when the stop-signal delay increase in timing, based on the previous Figure (i.e., Figure 2.11), resulting in an increase in mean commission error RTs, as shown in the modified Figure below (Logan & Cowan, 1984).

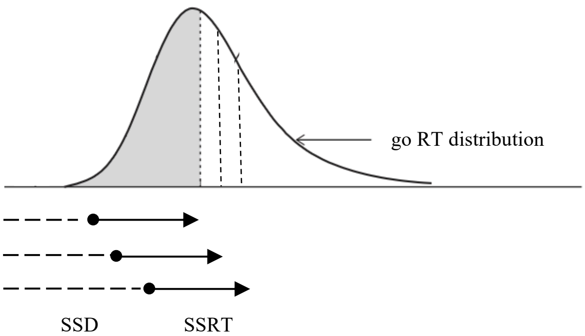


Figure 2.12. Representation of the increasing of the Stop-Signal Delay, with constant SSRT

According to these considerations on the independent horse race model, the differences in inhibition performance can be explained by the interaction between SSD, RT on go trials, and SSRT (Matzke et al, 2018). The subject's ability to control the responses can be explained by the interactions between these variables. These interactions are usually called inhibition functions and can be used to compare the performance of different experimental groups (Matzke et al, 2018). The inhibition function can be modified to represent different performance of participants. As previously mentioned and as shown in Figure 2.12, the finishing time of the stop process is influenced by SSD: subjects will be unable to inhibit their response if SSD occurs later, since the stop process is activated later; on the other hand, if SSD occurs early, subjects may be able to inhibit their response to the stop-signal. In contrast, if the go reaction time increases, the distribution of go RTs moves to the right, increasing the number of correct responses (Matzke et al, 2018). Therefore, the correct response rate is kept constant by participants by slowing down the response time (Lappin & Eriksen, 1966). Finally, if SSD is kept constant during the entire administration of the task, an increase of the response rate will be produced by an increase in SSRT (Matzke et al, 2018).

Starting from this simplified model, Logan and Cowan (1984) have derived the complete independent horse-race model, that assumes that both go RT and SSRT are independent random variables. Figure 2.13 shows that SSRT can now have a different value on each stop-signal trial.

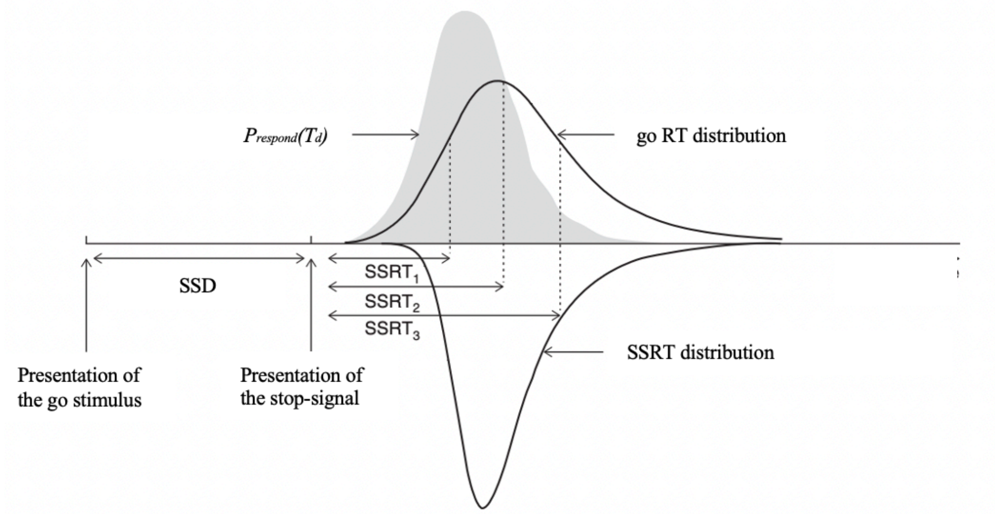


Figure 2.13. The complete independent horse race model (Logan & Cowan, 1984)
This image shows that response rate and the probability of inhibition are determined by the stop-signal delay (SSD), the stop-signal reaction time (SSRT), and the go RT distribution. Adapted from Matzke et al, 2013.

The complete independent horse race model assumes that increasing the mean of RT on go trials, the probability that the go process wins the race decreases (Matzke et al, 2018). On the other hand, if mean SSRT increases, the probability that the stop process wins the race decreases (Logan & Cowan, 1984). According to this model, the mean of the inhibition function is given by the difference between the mean of go RT and the mean SSRT (Logan & Cowan, 1984; Matzke et al, 2018; Verbruggen et al, 2019, etc.), as described in the following Equation:

$$T_d = T_{go} - T_{stop} \quad (1)$$

Equation (1) suggests one of the simplest methods to estimate mean SSRT (describe in the next paragraph; see paragraph 3.1 of the present Section).

To summarize, this model connects reaction times of responses on incorrect trials (i.e., commission errors or responses erroneously emitted), reaction times on go trials, and the probability of responding on stop-signal trials [p(respond|signal)], as a function of the stop-signal delay (Matzke et al, 2018; Verbruggen et al, 2019, etc.).

As I mentioned before, the go process, and the stop process are supposed to be independent (Logan & Cowan, 1984). Specifically, there are two types of independence: (a) stochastic independence, that assumes that the finishing times of the go and the stop process are independent (Logan & Cowan, 1984):

$$\begin{aligned} &P(T_{go} < t_{go} \cap T_{go} < t_{go}) \\ &\qquad = P(T_{go} < t_{go}) \times P(T_{stop} < t_{stop}) \quad (2) \end{aligned}$$

And, (b) context independence, that, assumes that the go process has the same distribution in both go trials and stop trials (Logan & Cowan, 1984):

$$P(T_{go} < t_{go}) = P(T_{go} < t_{go}|t_d) \quad (3)$$

As mentioned before, the independence assumption can be verified when mean RT in signal-respond trials (i.e., commission errors) is shorter than mean RT on go trials. When the independence assumption is violated, such as the mean RT on signal-respond trials is longer than mean RT on go trials, researchers should abstain from estimating SSRT (e.g., Recommendation 11; Verbruggen et al, 2019). It is always important to remember that the independent horse-race model is purely descriptive; it provides information about the latency of the unobservable stop response (i.e., the stop-signal reaction time) but it does not provide information about the processes that determine how long it take to a participant to inhibit the response.

In the next two paragraphs, I outline two others different models of response inhibition: the interactive race model (Boucher et al, 2007; paragraph 2.3.2 of this Section), and the Hanes-Carpenter model (Hanes & Carpenter, 1999; paragraph 2.3.3 of this Section).

2.3.2 The Interactive Horse Race Model

Logan and Cowan's model (1984) assumes the independence of the finishing times of the go and stop process. However, as stated from other researchers, complete independence between these two processes is implausible (Verbruggen & Logan, 2008). Moreover, there are many results in cognitive neuroscience literature suggesting that there is a strong interaction between the go process and the stop process (see for example, Schall et al, 2002). So, it is quite strange that a model like the independence horse race model fit the data so well.

In order to address this paradox, Boucher and colleagues (2007) developed the interactive horse race model. It is important to specify that this model can also be easily adapted with behavioral studies of saccadic inhibition (Boucher et al, 2007).

As the Logan and Cowan's independent horse race model (1984), also the interactive model assumes a race between the go and the stop process. However, in Boucher and colleagues' model (2007) the two processes are independent at the beginning of the delay of the stop process and interact almost near the end, when the stop-unit is activated. More specifically, in this model, the two processes (i.e., the go process and the stop process) are initiated by the presentation of the go stimulus and by the stop stimulus, respectively. However, the go unit and the stop unit are activated after an afferent delay. It is only after the activation of the stop unit, that the go unit can be strongly and quickly inhibited (Boucher et al, 2007). Whereas a go response is executed when the stop unit reaches the go unit too late to stop the activation (Boucher et al, 2007). Figure 2.14 shows how go activation a go and a stop process interact.

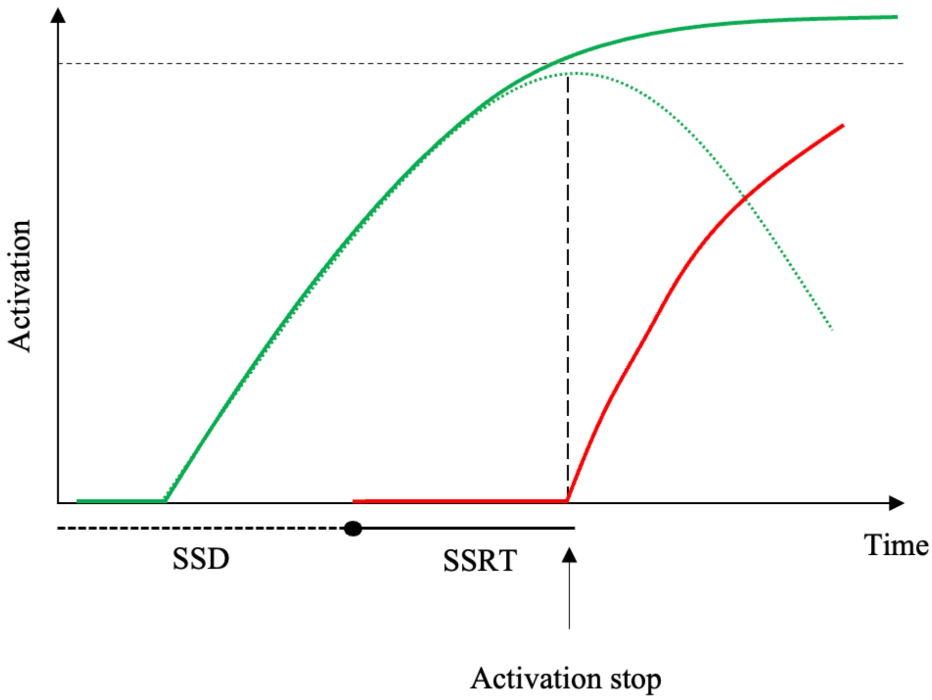


Figure 2.14. Representation of the assumptions of the interactive race model. This model indicates how go process is inhibited when the stop process is activated. Note. Green solid line indicates a Go process; green dotted line indicates a Go process with signal inhibit; red line indicates a Stop process; black dotted line indicates the threshold.

Thus, according to this model, the Stop-Signal Reaction Time represent the period before the stop unit is activated, where stop and go process are independent (Boucher et al, 2007). In other words, in this model SSRT changes as a function of the stop signal delay. Moreover, the behavioral predictions of the interactive horse-race model are similar to the behavioral predictions evaluated with the independent horse-race model (Colonius et al, 2001). Boucher and colleagues (2007) concluded the description of their model by stating that response inhibition consists in a first stage, where the go and stop process are independent, and a stage where the stop process inhibits the go process. As the interruption is brief, SSRT estimates from the independent horse-race model can be considered a valid measure of the latency of stop process, since it reflects the stage of response inhibition during which participants are encoding the information (Matzke et al, 2018).

In sum, according to this model, there is a strong interaction between stop and go processes despite their independence. When comparing the interactive horse race model and the independent horse race model, it should be observed that, (a) the SSRT estimations of both model are a good predictor of the other, when behavioral data are considered (Boucher et al, 2007; Matzke et al, 2018; Verbruggen et al, 2019); (b) second, the interactive horse race model can be fitted for both behavioral data and neurophysiological data (Worthy et al, 2013), while the independent horse race model can be fitted only for behavioral data; lastly, (c) the interactive horse race model can be applied to a constricted range of actions (i.e., inhibition of eye movements; Boucher et al, 2007), while the independent horse race model can be used for both discrete actions, such as arm movements (McGarry et al, 2003) and continuous actions such as speech (Slevc & Ferreira, 2006).

2.3.3 The Hanes-Carpenter Model

A third model was developed by Hanes and colleagues (Hans et al, 1999; Hans et al, 1998; Hans et al, 1995) in order to describe the inhibition of eyes saccades in monkeys and humans. Thus, also the Hanes-Carpenter model (Hans et al, 1999; Hans et al, 1998; Hans et al, 1995) applies exclusively to simple RT go tasks saccade inhibition. Matzke and colleagues (2018) stated that this model can be considered as special case of the Logan and Cowan (1984). For an extended description of similar task used to assess this model see https://osf.io/tudmw/ (Gialdi, 2022). Performance in this task is similar to the performance in other stop-signal tasks with different responses, but in this case reaction time on go and SSRTs are shorter (Matzke et al, 2018).

The Hanes-Carpenter model has two main characteristics: it is based on the Linear Approach to Threshold with Ergodic Rate (LATER; Carpenter, 1981; Carpenter et al, 1995), and it has some specific distributional assumptions about the go process and stop process (Matzke et al, 2013a).

To conclude, comparing the Hanes-Carpenter model and the independent horse race model there are two observations to make. First, on the contrary of the independent horse race model, the Hanes-Carpenter model gives some awareness on the mechanisms of the go process and the stop process (Matzke et al, 2013a). Secondly, if we want to compare the two models, it is important to note that the Hanes-Carpenter model has only been used for saccade inhibition experiments in monkeys and humans, and with a very small number of participants (Colonuis et al, 2001). Lastly, the interactive race model proposes a mechanism by which responses are stopped inhibiting the activation of the go process (Matzke et al, 2018).

3. Frequentist Estimations of SSRT

In this paragraph, two frequentist estimation methods of Stop-Signal Reaction Time are discussed. Several methods are available in literature to estimate SSRT (e.g., Verbruggen & Logan, 2009). The two most popular methods to estimate SSRT focus on obtaining summary measures of the latency of stop process, described in this paragraph, are: the mean method and the integration method (Logan, 1994). These methods differ in whether they treat SSRT as a constant or as a random variable. The integration method assumes SSRT is constant, while the mean method assume SSRT is a random variable.

Based on the design of the study, researcher can decide which estimation method is better for the data. This choice also depends on how stop-signal delay is set during the design of the task (Verbruggen et al, 2019). As previously stated (see paragraph 2.2.3 of the present Section), there are two procedures for setting stop-signal delay: the fixed-SSDs procedure, using a specific number of fixed stop-signal delays (i.e., the stop-signal will be executed with a fixed delay on all stop trials), or the tracking procedure, that is a result of the adjustment of the stop-signal delays dynamically by 50ms after every trial (i.e., after every successful inhibition, the SSD is increased by 50ms, and when the inhibition is unsuccessful, the SSD is decreased by 50ms; Verbruggen et al, 2019). The most common tracking procedure is based on the correction of stop-signal delay after every trial (Verbruggen & Logan, 2009; Verbruggen et al, 2013). As mentioned before, for this research project this tracking procedure has been used (see paragraph 2.2.3 of thi Section for details).

3.1 Mean Method

The mean method, proposed by Logan and Cowan (1984), is the most popular method for estimating SSRT. This model considers SSRT as random variable and can be applied for both fixed stop-signal delay and tracking stop-signal delay studies (Logan & Cowan, 1984; Matzke et al, 2018). However, it is mostly used when the tracking procedure is used to set stop-signal delays (Logan & Cowan, 1984; Logan et al, 1997). As discussed before, in this project I relied on the tracking method, by setting the changing of the Stop-Signal Delay after every successful or unsuccessful trials. For an extended description of this method see https://osf.io/tudmw/ (Gialdi, 2022).

As shown in Equation (1) (see paragraph 2.3.1 of the present Section), mean SSRT can be computed by subtracting the mean of stop-signal delay from mean go RT (Matkze et al, 2018):

$$T_d = T_{go} - T_{stop} \qquad (1)$$

Consequently, Equation (1) becomes Equation (4):

$$T_{stop} = T_{go} - T_d \qquad (4)$$

Due to its easiness of computation, the mean method is the most used method for estimating SSRTs (Matzke et al, 2018) and has been implemented in the Verbruggen and colleagues' STOP-IT software (Verbruggen et al, 2008b). When we consider the mean method, there are other advantages: first, it is computationally easy; second, the mean method does not require unrealistic assumptions on SSRT. And lastly, the mean method's estimates for SSRT given p(respond|signal) = .50, are the most reliable (Band et al, 2003), making the dominant method for estimating SSRT. On the other hand, its disadvantages are that it is impossible to provide an estimate of SSRT for a given single delay; and it lacks estimates for other parameters of the SSRT distribution. Finally, Verbruggen and colleagues (2019) stated that the mean method is strongly influenced by skew of the go RT distribution, as well as by go omissions (i.e., go trials where no response is executed, without the stop-signal).

3.2 Integration Method

A second frequentist approach is the integration method. For the purpose of this project, I relied on the integration method with replacement of go omission, because it is the least biased and most reliable (Verbruggen et al, 2019). Thus, in order to estimate SSRT with the integration method, go omissions are assigned the maximum RT because of the missing response (Verbruggen et al, 2019). Indeed, Verbruggen and colleagues (2019) in their consensus guide to the Stop-Signal Task, reported in one of the recommendations (i.e., Recommendation 8) the importance to estimate SSRT using the integration method with replacement of go omission, if using a non-parametric approach. For an extended description of this method see https://osf.io/tudmw/ (Gialdi, 2022).

Until 2003, Logan's integration method (Logan, 1981) was the most commonly used method in the assessment of SSRT. Logan's integration method assumes that SSRT is constant, it is mostly used for fixed stop signal delay experiments and allows for the estimation of SSRT for each stop-signal delay separately (Logan & Cowan, 1984; Matzke et al, 2018).

Equation (3) represent the adjustment of p(respond|signal) to compensate for go omissions (Tannock et al, 1989; Verbruggen et al, 2019):

$$p(respond|signal)_{adjusted} = 1 - \frac{P(inhibit|signal) - (omission|go)}{1 - p(omission|go)} \quad (3)$$

However, as other estimation methods, it may be difficult to use the integration method if the assumptions of the independent horse-race model are violated, since it is more susceptible (Verbruggen et al, 2019).

To conclude, Verbruggen and colleagues (2013) showed that the mean method produce overestimation of the SSRTs when participants wait for the stop-signal presentation, slowing down their responses (Matzke et al, 2018); on the contrary, the integration method is less sensitive, but it can underestimate SSRTs when participants slow down their response (Matzke et al, 2018; Verbruggen et al, 2013). This bias might disappear if researchers apply the integration method to smaller blocks of trials (i.e., for this project the integration method has been evaluated for the entire experiment and in the six different blocks) as opposed to the entire experiment (Matzke et al, 2018). Verbruggen and colleagues (2013) therefore recommended in their guide to the Stop-Signal Task, that whether researchers want to estimate SSRT, might better use the integration method in combination with the tracking procedure (Matzke et al, 2018).

Lastly, several works (for a review see Band et al, 2003) stated that both non-parametric methods previously described, when used with fixed stop-signal delays, produce reliable estimations of SSRT (Matzke et al, 2018). However, as Matzke and colleagues sustained, this procedure with fixed stop-signal delays require a large number of observations. For example, in the work of Band and colleagues (2003), the authors stated that to obtain reliable estimates using the integration method, researchers have to present participants a stop signal task with at least more than 900 trials and five different stop-signal delays.

3.3 Conclusion

It is well known that relying on measures of central tendency, such as the mean, produce results without important characteristic of the data (e.g., Heathcote et al, 1991; Matzke et al, 2018; Matzke & Wagenmakers, 2009). Similarly with the Stop-Signal Task, using only summary measures of the estimation of the SSRT may cover important aspects of stop-signal data and lead to erroneous conclusions (Matzke et al, 2018; Verbruggen et al, 2019). For example, if we consider only the mean estimation of SSRT, two group may have the same mean, but their distributions may be completely different (Matzke et al, 2018). However, frequentists estimations are still commonly used, due to their simplicity. Simulation study of Band and colleagues (2003) showed that SSRT estimates derived from the central part of the distribution (i.e., mean method) are most reliable. These central estimations are less influenced by variability in go RT and SSRT and rarely violate the independence assumptions (Band et al, 2003; Verbruggen & Logan, 2009). Thus, based on these assumptions, the mean method or the median method typicallyare more reliable than integration when estimates SSRT (Verbruggen & Logan, 2009).

However, Verbruggen and colleagues (2019) showed the overestimation and the underestimation of SSRT with four different frequentist methods: (1) the mean method, (2) the integration method with replacement of go omissions, (3) integration method with exclusion of go omission, and (4) integration method with adjustment of *p*(respond|signal). These authors calculated the difference between the estimated SSRT and the actual SSRT (Verbruggen et al, 2019). They also differentiated between included and excluded participants (i.e., participants who violated the independent horse race model, with RT on signal respond trials are larger than RT on go trials).

The authors found that although these two non-parametric methods represent the two most widely used approaches to SSRT estimations, the mean method was more biased in simulation studies, as shown in Table 3.1. Moreover, estimates were generally more biased for excluded participants than for included participants.

Table 3.1. Differences between estimated and true SSRT for included participants and exclude participants (i.e., mean RT on unsuccessful stop trial greater than mean RT on go trials).

Estimation method	Included	Excluded
Mean	-16.0	-46.34
Integration with replacement of go omissions	-6.4	-35.8
Integration with exclusion of go omissions	-19.4	-48.5
Integration with adjusted p(respond|signal)	12.5	-17.4

Note. Positive values: SSRT overestimated; Negative values: SSRT underestimated.
Source: Verbruggen et al, 2019

To conclude, as described in previous paragraphs and as deeply described by Verbruggen and colleagues (2019), the integration method with replacement of go omissions is the least biased and most reliable; nevertheless, the mean method is still the method most used despite the issues described below.

4. Bayesian Estimation Methods of SSRT

As I mentioned before, several researchers have shown that focusing only on measures of central tendency to estimate SSRT gives insufficient information regarding the nature of the data (see for example, Matzke et al, 2018; Soltanifar et al, 2001). As seen in an ADHD group compared to controls (Leth-Steensen et al, 2000), or in a schizophrenia group versus controls (Belin & Rubin, 1995), it is possible that the mean of SSRT in the two studied group is similar, but the shape of its distribution is completely different. In fact, the existing methods to estimate SSRT (i.e., mean and integration method) are unable to estimate the shape of entire SSRT distributions. Crucial features of the data may be loss if researchers focused only on the mean SSRT, ignoring the shape of SSRT distributions. Most importantly, the results obtained by these methods may be result erroneous (Matzke et al, 2018).

Others SSRT estimations that can be used, are the parametric methods, that are less biased than non-parametric methods (Verbruggen et al, 2019). Although the non-parametric approach Colonius's method (Colonius, 1990) theoretically allows researchers to estimate the non-parametric distribution of SSRT, it is difficult to be implemented in practice because of the elevated number of trials that require to obtain valid SSRT estimations. Moreover, the estimation with this method underestimates the mean SSRT and overestimates its variance (Band et al, 2003).

For the purpose of this project, I considered Bayesian Parametric Approach (BPA), that are used to obtain posterior distributions for the model parameters. From this point of view, successful response inhibition requires fast stop, and the stop process must also be successfully triggered before the beginning of the race against the go process. Participants will have difficulties to inhibit their response, if they fail to interpret the stop-signal.

The Bayesian Parametric Approach allow the possibility to estimate the entire distribution of SSRT, and this can be applicable to real data with a low number of trials (Matzke et al, 2013a; Matzke et al, 2017; Dupuis et al, 2018), and, as previously mentioned, enables researchers to evaluate the presence of differences in the shape of SSRT distributions between two different groups.

Matzke and colleagues (2013a) developed a descriptive Bayesian Parametric Approach that enables researchers to estimate the entire distribution of SSRT. This approach can provide more accurate estimates of SSRT even with a less number observations (Matzke et al, 2013a). The BPA consider the same assumption of the independent horse race model: the reaction time in go trials and SSRT have independent distributions (Matzke et al, 2013a). This method estimates the entire distribution of SSRT under the assumption that the go RTs and SSRTs are ex-Gaussian distributed and relies on Markov chain Monte Carlo (MCMC) sampling to obtain posterior distributions for the model parameters (Heathcote et al, 1991; Matzke, 2014; Matzke et al, 2013a; Skippen et al, 2019; Wagenmaker et al, 2008).

Matzke (2014) stated that the ex-Gaussian is frequently used as a distributional model that usually produces excellent fit to empirical RT distributions (Matzke et al, 2013a). Figure 4.1 represents some examples of ex-Gaussian distributions.

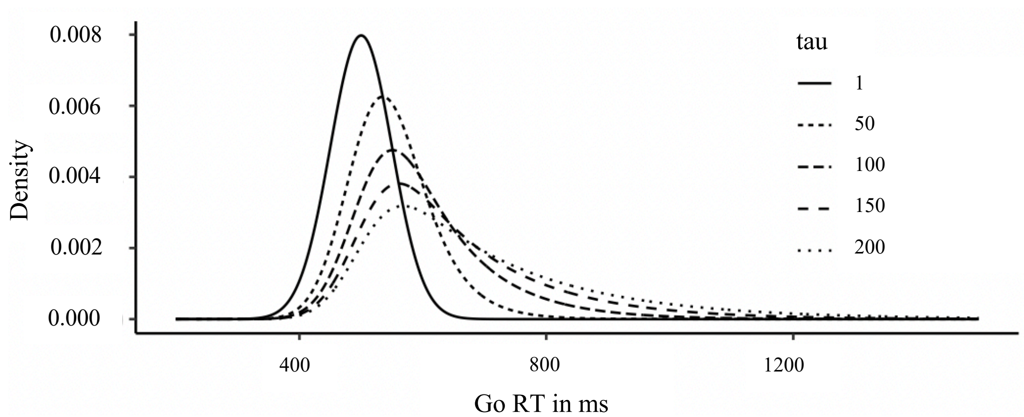


Figure 4.1. Examples of ex-Gaussian RT distributions.

The ex-Gaussian is a three-parameter distribution that is given by the convolution of a Gaussian and an exponential distribution (Matzke, 2014; Matzke & Wagenmakers, 2009). For an extended description of the parameters of the ex-Gaussian distribution see https://osf.io/tudmw/ (Gialdi, 2022).

The Bayesian Parametric Approach can be applied to individual and to hierarchial data (Matzke, 2014). In the first estimation, SSRT is estimated separately for each participant, whereas the hierarchical approach allows to estimate the individual parameters considering the information from the entire group analyzed (Matzke, 2014).

In the next paragraph a user-friendly method for the estimation of SSRT distribution with the Bayesian Parametric Approach developed by Matzke and colleagues (2013b) will be presented.

4.1. Bayesian Ex-Gaussian Estimation of Stop-Signal RT distributions

To use the Bayesian Parametric Approach, Matzke and colleagues, developed a fast and user-friendly software to estimate the entire SSRT distributions, namely Bayesian Ex-Gaussian Estimation of Stop-Signal RT distributions, BEESTS (Matzke et al, 2013b). Following the Independent Horse Race Model, BEESTS assumes that response inhibition is determined by the relative finishing time of the stop process and the go process, as two independent processes. The new approach can be applied to individual as well as hierarchical data (Matzke, 2014). BEESTS is open-source software for the estimation of SSRT distribution and can be downloaded from Matzke profile on OSF (https://osf.io/kk287/). Figure 4.2 shows the graphical user interface.

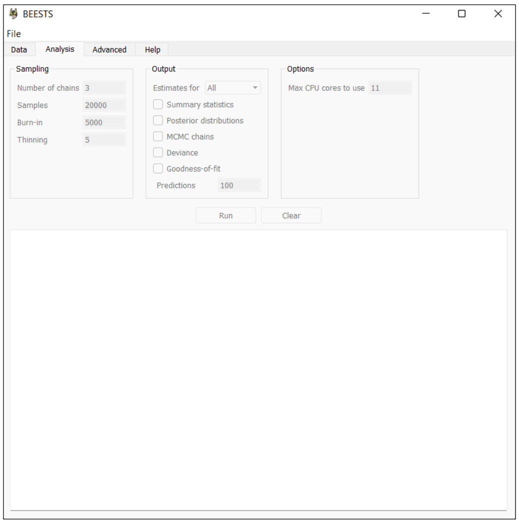
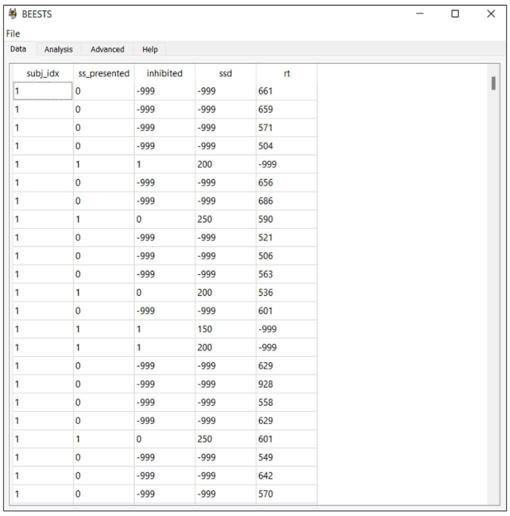


Figure 4.2. BEESTS user friendly interface.

After loading data in .csv (i.e., comma-separated values) file, as shown in Figure 4.3, users can specify the details of the MCMC sampling, the required output.



subj_idx	ss_presented	inhibited	ssd	rt
1	0	-999	-999	661
1	0	-999	-999	659
1	0	-999	-999	571
1	0	-999	-999	504
1	1	1	200	-999
1	0	-999	-999	656
1	0	-999	-999	686
1	1	0	250	590
1	0	-999	-999	521
1	0	-999	-999	506
1	0	-999	-999	563
1	1	0	200	536
1	0	-999	-999	601
1	1	1	150	-999
1	1	1	200	-999
1	0	-999	-999	629
1	0	-999	-999	928
1	0	-999	-999	558
1	0	-999	-999	629
1	1	0	250	601
1	0	-999	-999	549
1	0	-999	-999	642
1	0	-999	-999	570

Figure 4.4. The upload data for the hierarchical analysis.
Note "subj_idx" contains the participant code; "ss_presented" contains the trial type, where go trials are coded with 0 and stop-signal trials are coded with 1; "inhibited" contains the inhibition data, where signal-respond trials are coded with 0 (i.e., unsuccessful inhibition), signal-inhibit trials are coded with 1 (i.e., successful inhibition), and go trials are coded with −999; "ssd" contains the stop-signal delay in ms., where go trials are coded with −999; "rt" column contains the go RT for go trials and the signal-respond RT for signal-respond trials in ms., where signal-inhibit trials are coded with −999.

Whereas Figure 4.4 shows the details that users can modify (i.e., sampling or CPU cores) or select (i.e., estimations) with BEESTS.

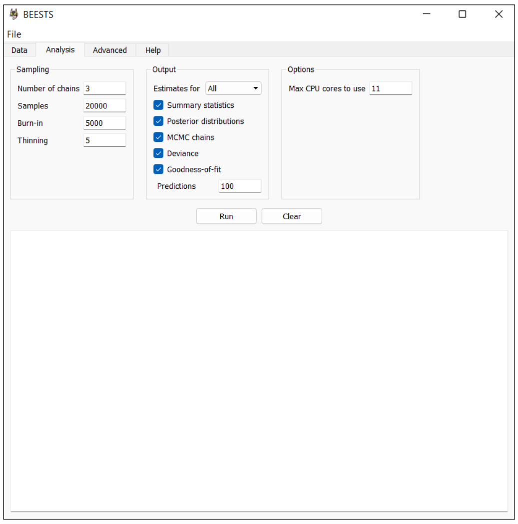


Figure 4.4. *Details that users can define for the analysis.*

As shown in Figure 4.5, for this project I relied on the default options for the sampling, I also selected all the options in the output section, to estimate summary statistics, posterior distribution, MCMC chains, deviance and goodness of fit.

Moreover, BEESTS allows researchers to indicate whether they want to do the analysis and the estimation of SSRT with or without including Trigger Failure in the model. The default setting is without trigger failures, but users can change this setting by checking the option in advanced option. Figure 4.6 shows the screen of the selected option of trigger failures.

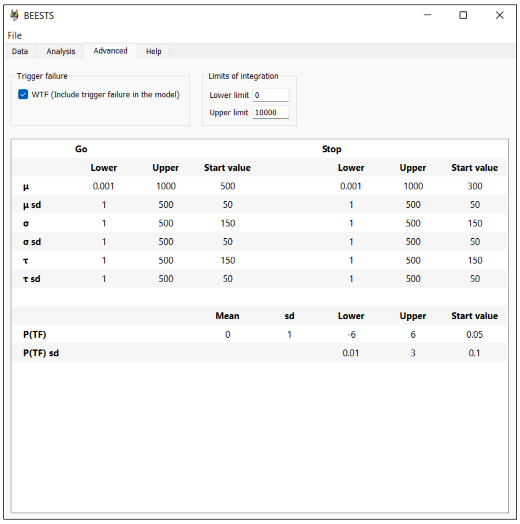


Figure 4.5. Option "WTF (include trigger failure in the model)" selected.

Trigger failures pose well-known theoretical and methodological challenges to the interpretation of stop-signal data (Logan, 1994; Matzke et al, 2017). As Matzke and colleagues (2017) stated, differences in inhibition performance across groups might just reflect differences in the probability of triggering the stop process. For example, poor response inhibition in clinical population may reflect a difficult to evidence a stop process not triggered (Matzke et al, 2017; Schachar & Logan, 1990). Band and colleagues (2003) have shown that trigger failures can overestimate SSRTs. Similarly, Matzke and colleagues (2017) demonstrated that trigger failures can bias the estimation of entire SSRT distributions. Despite its importance, it is very difficult to quantifying the contribution of trigger failures during a performance on Stop-Signal Task. For this reason, Matzke and colleagues (2017), identified a Bayesian method that allows researchers to measure the probability of trigger failures and the whole distribution of SSRTs (see Figure 4.5).

In Figure 4.6, the representation of the beginning (part left) and the ending (part right) of the analysis of BEESTS with the progress of the sampling and the saving process of the output.

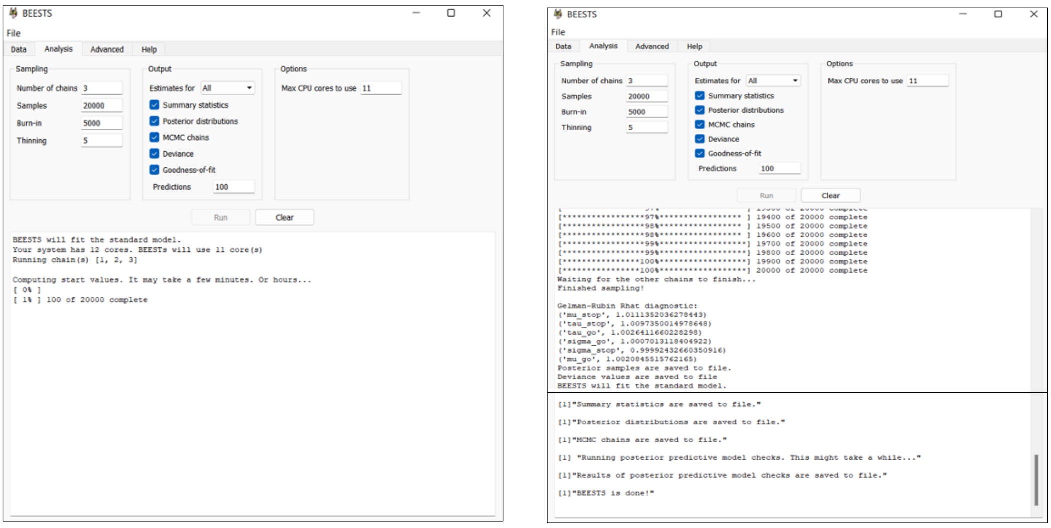


Figure 4.6. Running analysis on BEEST.

Part left shows the start of the analysis; for individual analysis BEESTS took almost five minutes; for hierarchical analysis BEESTS took almost 24 hours. Part right shows the end and the saving process of the analysis.

Depending on whether researchers want to estimate SSRT with individual or hierarchical data, the BPA estimates parameters of the SSRT distribution separately for each participant (i.e., individual estimation) or estimates the parameters of SSRT for each participant with information from the entire population, as well as parameters of SSRT for the entire population (i.e., hierarchical estimations; Gelman & Hill, 2007; Matzke et al, 2017; Rouder et al, 2005). The hierarchical approach can provide more accurate and less variable estimates than individual estimation, especially if there are only few numbers of trials per participant (e.g., Matzke et al, 2017; Rouder et al, 2003). Specifically, the individual approach provides accurate and precise parameter estimates with 250 trials, and on the contrary, the hierarchical approach requires a sample size of approximately 25 participants, each with at least 100 stop-signal trials (Matzke et al, 2013b; Matzke et al, 2017). As previously described (see paragraph 2.2.3), for this project I relied on a Stop-Signal Task with more than 380 trials, and more than 100 participants (see paragraph 5 and 6 of this Section), that allow me to estimate hierarchical analysis for the Stop-Signal Task.

4.2 Individual BPA

The individual BPA assumes that participant are independent and uses regular Bayes theorem for both reaction time on go trials and SSRT distributions.

The goal is to estimate the μ_{go}, σ_{go}, τ_{go}, μ_{stop}, σ_{stop}, and τ_{stop} parameters for each participant separately. As I mentioned before, I used BEEST software (Matzke et al, 2013b), with the default sampling: number of chains = 3, samples = 2000, burn-in = 5000, and thinning = 5. The analysis for individual estimations took about 5 minutes with BEESTS. Figure 4.8 shows part of the output of an individual analysis (the complete output is available in Appendix 1). This figure represents the posterior and prior distributions (left part of the Figure) and the MCMC chains for the six model parameters of one of the participants of this project (right part of the Figure). Specifically, the right part of Figure 4.8 shows MCMC chains sampled from the posterior distribution of the parameters of the SSRT distribution (Matzke et al, 2017).

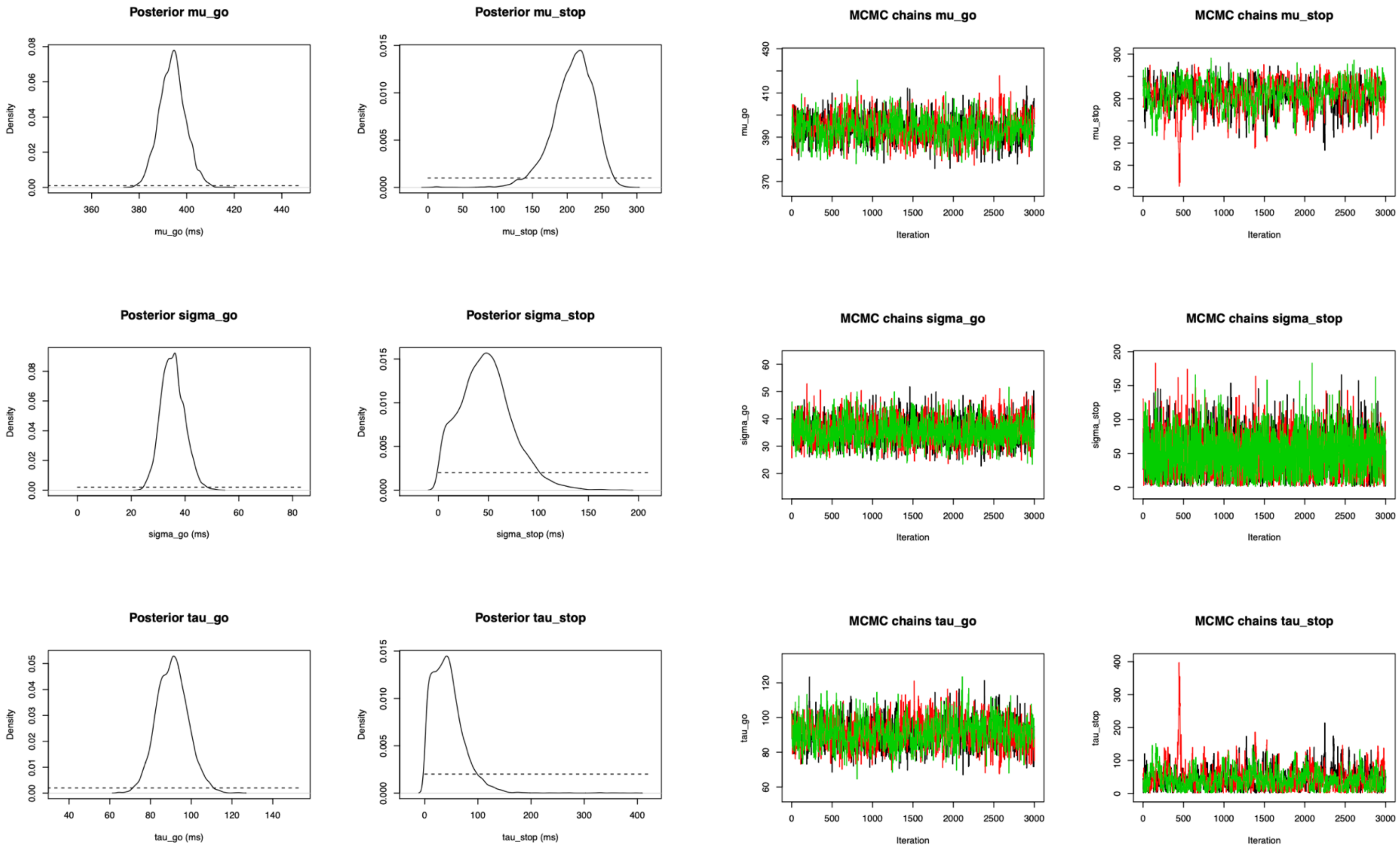


Figure 4.8. Output of individual analysis
Left panel: density plot for the posterior (solid line) and prior distributions (dashed line) of the individual model parameters.
Right panel: trace plots for the individual model parameters.

As Dora Matzke (2014) stated in her study "*if the model adequately describes the data, the predictions based on the model parameters should closely approximate the observed data*". Moreover, these analyses allow researchers to compute posterior predictive *p* values (e.g., Matzke et al, 2014).

The individual analysis also produces Summary statistic (see Appendix 1) and the posterior goodness of fit. As Matzke and colleagues (2017) stated, for this project, I assessed model fit only on SSDs with a at least one observed signal-respond RTs (i.e., unsuccessful inhibition; see validity checks in paragraph 5.3 of this Section). As shown in Figure 4.9, and based on Matzke description of the analysis (2014), the one-side *p* values on this participants' SSDs (i.e., 100, 150, 200, 250, 300, 350) are far from 0 or 1, and the two-side *p* values are all above 0.05. On the contrary, if the posterior predictive *p* values would have been very close or are equal to 0 or 1, the BPA had failed to account for the median of the observed signal-respond RTs (Matzke, 2014).

Posterior predictive p values for median SRRT

	SSD=100	SSD=150	SSD=200	SSD=250	SSD=300	SSD=350
Number of observed SRRT	2	4	13	13	9	4
Observed median SRRT	376.5	384.5	413	406	458	478
Average predicted SRRT	407.87	411.87	420.44	429.49	443.91	456.42
One−sided p value	0.75	0.828	0.68	0.91	0.27	0.25
Two−sided p value	0.5	0.343	0.64	0.18	0.54	0.5

Figure 4.9. Output for the Goodness of fit of the posterior analysis
SRRT = Signal-Respond Reaction Time

According to Matzke and colleagues (2013a) the individual Bayesian Parametric Approach for individual estimations analysis provided good posterior distributions for most participants, shoeing that the mean of SSRTs computed with the BPA approximated the mean SSRTs obtained with non-parametric estimations.

4.3 Hierarchical BPA

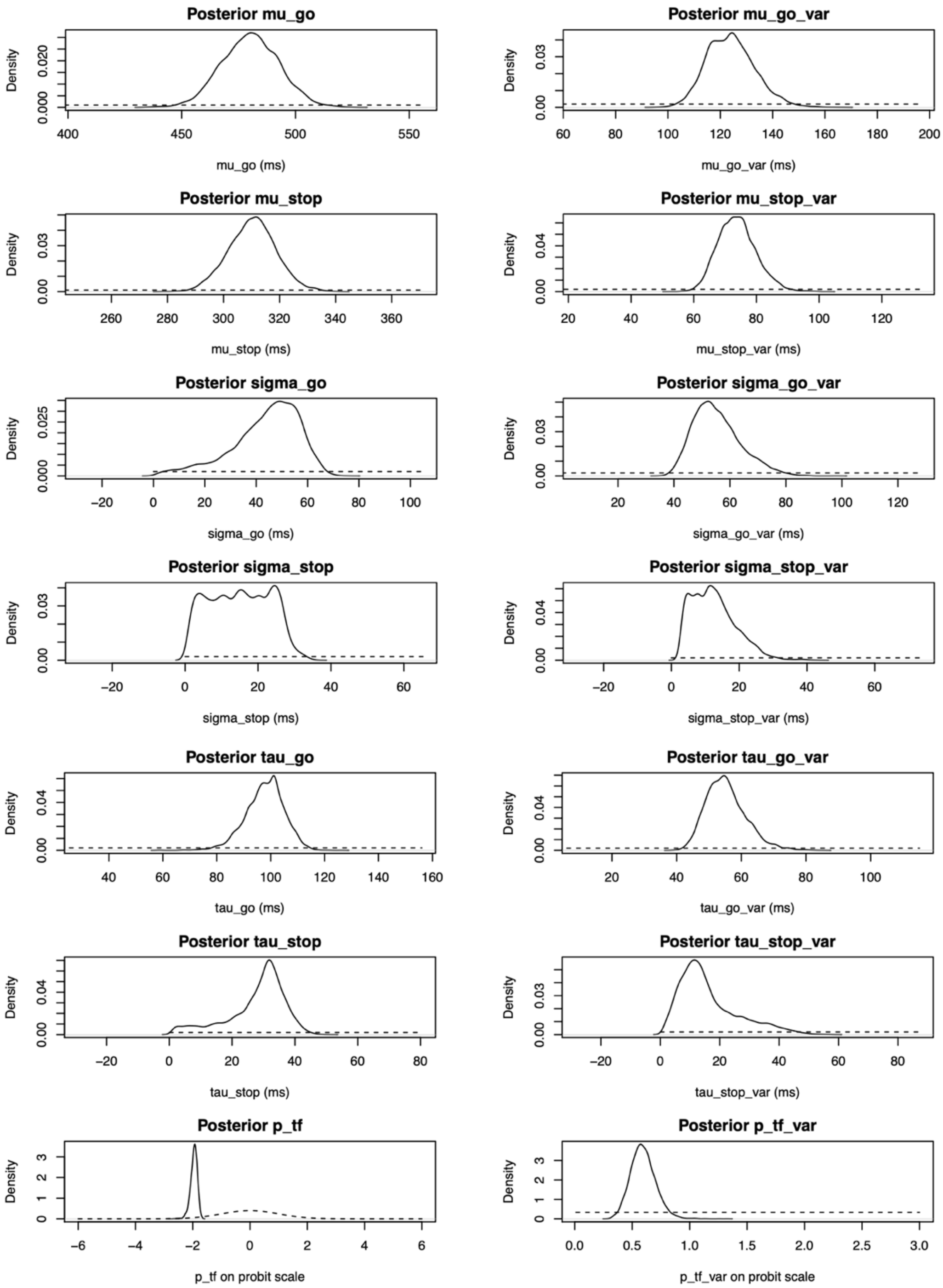
Matzke (2014) described the Bayesian Parametric hierarchical approach as the one that allows researchers to model individual differences and similarity between participants (Farrell & Ludwig, 2008; Lee, 2011; Matzke, 2014; Nilsson et al, 2011). Indeed, hierarchical approach consider participants as both independent identical (Matzke, 2014). Indeed, the hierarchical BPA aims at estimating both individual participant parameters, by considering the mean distribution of the population, as well as population parameters. Moreover, posterior median of each participants' parameters can be used to estimate the SSRT distribution of each participant (Matzke et al, 2013a).

Hierarchical methods can provide accurate and less variable parameter estimates (Matzke, 2014), with a small number of observations per participant (for example, 100 stop-signal trials per participants; Gelman & Hill, 2007; Matzke, 2014; Matzke et al, 2013b). Indeed, the hierarchical modeling take the information from the entire group to improve individual estimation (Matzke, 2014). For additional details of this method see https://osf.io/tudmw/ (Gialdi, 2022).

Real experiment does not provide normally distributed data for individual, particularly with clinical participants.

Figure 4.10 shows part of the output of a hierarchical analysis. This figure represents the posterior and prior distributions (part A) and the MCMC chains for the six model parameters of the entire groups of participants (part B). A complete output for the hierarchical estimations can be found at the following link: www.osf.io/w5fpt (Gialdi, 2022).

Part A


Posterior mu_go
Density
mu_go (ms)
Posterior mu_go_var
mu_go_var (ms)
Posterior mu_stop
mu_stop (ms)
Posterior mu_stop_var
mu_stop_var (ms)
Posterior sigma_go
sigma_go (ms)
Posterior sigma_go_var
sigma_go_var (ms)
Posterior sigma_stop
sigma_stop (ms)
Posterior sigma_stop_var
sigma_stop_var (ms)
Posterior tau_go
tau_go (ms)
Posterior tau_go_var
tau_go_var (ms)
Posterior tau_stop
tau_stop (ms)
Posterior tau_stop_var
tau_stop_var (ms)
Posterior p_tf
p_tf on probit scale
Posterior p_tf_var
p_tf_var on probit scale

Part B

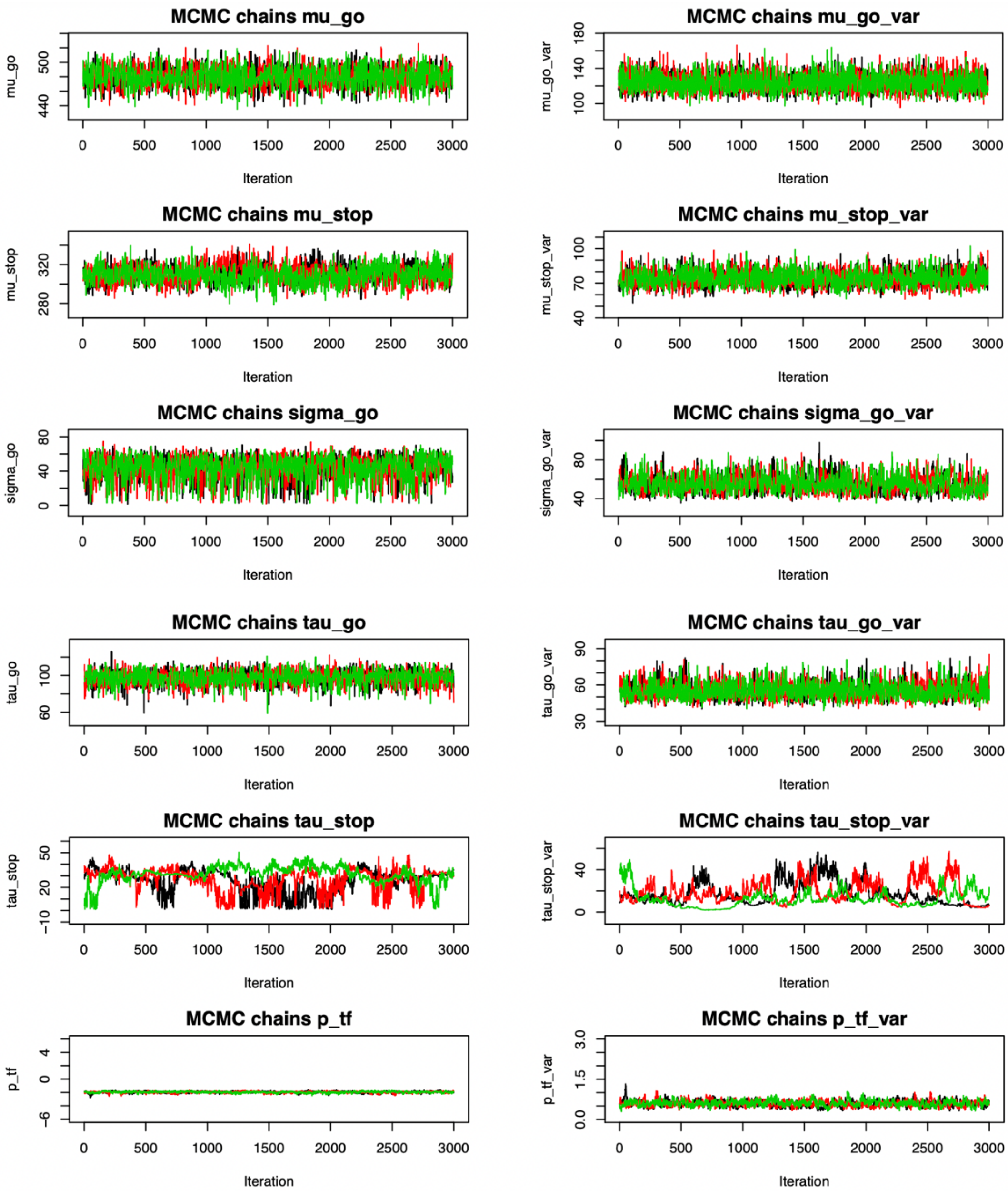


Figure 4.10. Output of hierarchical analysis
Part A: density plot for the posterior (solid line) and prior distributions (dashed line). Part B: trace plots for the individual model parameters.

Although the hierarchical BPA is more complex than individual BPA and more long to compute, it has three main advantages. First, as the hierarchical BPA uses information from the distribution for the entire group, it has more precise posterior parameter estimates for individual participants. Second, the posterior parameter estimates for the distribution for the entire group can be used to compare different clinical or experimental groups. And lastly, it needs a smaller sample size per participant with a small number of trials per participants, to have moderately precise estimates for posterior parameters compared to the individual BPA. This last advantage may be useful for clinical and experimental practice, when comparing subjects with difficulties sustaining prolonged administrations.

4.4 Conclusion

To conclude, based on Bayesian Parametric Approach method, researchers can evaluate differences between clinical populations or experimental groups in the shapes of SSRT distributions (Matzke, 2014). As a result, stop-signal data can be interpreted more easily, and some unknown aspects of response inhibition can be revealed. The authors can choice between one of the two approaches considering different aspects of the study: for instance, Matzke (2014) suggested the quality of the data, the number of participants and the number of trials per participant.

BPA can accurately estimate SSRT in experimental stop-signal data containing a relative low number of trials in both individual and hierarchical data structures. Indeed, the hierarchical BPA provide accurate estimations also with a small sample size (Matzke, 2014; Matzke et al, 2013a), and with relatively few observations per participant, that is more common in studies with clinical population (e.g., in my project I relied on a Stop-Signal Paradigm with 384 trials and, based on participants feedbacks, the task was too long) (Matzke et al, 2013a).

About this approach, there are also has some limitations. One weakness is related to the amount of data required for precise estimates (Matzke, 2014; Matzke et al, 2013a), and consequently, the time spent to organize and analyze the data. Indeed, individual estimations need a very large number of data, that are not always collected in clinical fields. However, the hierarchical approach may provide a solution in such situations (see, Matzke, 2014). On the contrary, the computational analysis for individual estimates is definitely easier and faster, than hierarchical estimations (e.g., for the present project hierarchical estimations took more than 24 hors).

Based on these limitations, for this project I relied on BEEST software using hierarchical estimations.

5. Study 1

The first study of my project aimed at assessing the correlations between self-reports of impulsivity and behavioral tasks. The importance of testing these associations between two types of measurement of the same construct, is due to a substantial number of articles describing modest correlations between them (see for a review, Sharma et al, 2014). In Sharma and colleagues' meta-analysis (2014), low correlations between self-reports and laboratory tasks are due to an inconsistent definition of impulsivity across methods. As described in the Introduction, impulsivity is considered both as a unidimensional and multidimensional construct. Due to its complexity, impulsivity can be conceptualized as a label that include similar behavioral patterns representing distinct traits.

However, both laboratory tasks and self-reports can describe more than one dimensions of impulsive behavior (Cyders & Coskunpinar, 2011, 2012; Lane et al, 2003). For this reason, it is important to use multiple assessment strategies to obtain an accurate description of impulsivity (Sharma et al, 2014; Stahl et al, 2014). Moreover, Sharma and colleagues (2014) found that the use of both self-reports and laboratory tasks is useful in order to predict external criteria.

According to existent literature (e.g., Cyders & Coskunpinar, 2011, 2012; Sharma et al, 2014), studying correlations between different methods that assess impulsivity (i.e., self-reports and behavioral task) seems very important.

The validity of the Stop-Signal Task and the corresponding SSRT, have been demonstrated in several studies (e.g., Sharma et al, 2014; Verbruggen et al, 2019; Wöstmann et al, 2013). In the same way, numerous self-report measures showed adequate psychometric proprieties in non-clinical and clinical sample (Cyders & Smith, 2007; Fossati et al, 2001; Patton et al, 1995; Sharma et al, 2014; Whiteside & Lynam 2001; Zuckerman et al, 1991). Nevertheless, meta-analytic findings suggested that the association between SSRT and impulsivity self-reports is modest ($r \approx .1$) (Sharma et al, 2014). It is also important to consider that improved estimation of the SSRT (i.e., parametric estimations vs. non-parametric estimations) may better identify relationships between computerized tasks and self-reports (Skippen et al, 2019).

Just as it important to assess correlations between behavioral tasks and self-reported measures, it is also important to measure the convergent validity between the Stop-Signal

Task with an alternative measure of response inhibition. Convergent validity allows researchers to verify whether the Stop-Signal Task evaluates response inhibition in the same way as a complementary validated task (i.e., Go/No-Go Task; Bezdjian et al, 2009).

A second purpose of this study was to evaluate the association between two tasks that aim to measure the same construct: response inhibition. As previously described (see paragraph 2.2), these two tasks have a similar goal (i.e., press a key when a visual go-signal appears, and suppress the answer when a visual stop-signal appears). However, during the Go/No Go Task the go-stimulus is replaced with a stop-signal; whereas during the Stop-Signal Task, the go-stimulus is always presented, but is immediately followed by the stop-signal, so that the response is in the process of completion (Verbruggen & Logan, 2008). Researchers have been arguing whether the Go/No Go Task and the Stop-Signal Task measure the same inhibitory mechanism (Raud et al, 2020; Schachar et al, 2007). Existent literature is controversial about the distinction (Raud et al, 2020) or similarity (e.g., Eagle et al, 2008) between these two tasks. Eagle and colleagues (2008), indeed, stated the interchangeability of the tasks, without providing a methodological rationale behind the choice. On the contrary, other studies (Raud et al, 2020; van Gaal et al, 2010, 2009, 2008) demonstrated the distinction between action cancellation in the Go/No Go Task, and automatic response inhibition in the Stop-Signal Task.

In the light of the still open debate about the differences between these two tasks, it is methodologically important to evaluate and compare the performance in Go/No Go Task and Stop-Signal Task. Moreover, it is also important to consider improved estimation of SSRT when comparing two different tasks.

5.1 Aim

Starting from these considerations, the first study aimed at evaluating how different methods for estimating SSRT could produce different correlations with self-reports of impulsivity. For estimating the SSRT, I considered the following methods: the mean method (see paragraph 3.1 of the present Section), the integration method (see paragraph 3.2 of this Section), the Bayesian estimation of ex-Gaussian SSRT (BEESTS method; see paragraph 4.1 of this Section), BEESTS with trigger failure (also see paragraph 4.1 of the present Section).

Based on previous findings (Gialdi et al, 2020; Skippen et al, 2019), I expected that frequentist estimations methods (i.e., mean method) of SSRT were weakly associated with self-reports measures, whereas BEESTS estimates, were more associated with self-reports.

In addition, this study aimed at evaluating convergent validity between the Go/No-Go Task and the Stop-Signal Task, taking into consideration if different methods for estimating SSRT could produce different results. As previously mentioned, for estimating the SSRT I considered the mean method (see paragraph 3.1 of the present Section), the integration method (see paragraph 3.2 of this Section), the Bayesian estimation of ex-Gaussian SSRT (BEESTS method; see paragraph 4.1 of this Section), BEESTS with trigger failure (also see paragraph 4.1 of the present Section).

Based on previous findings (Eagle et al, 2008), I expected that the probability of responding to the stop signal during the Stop-Signal Task showed association with the commission errors in the Go/No-Go Task. In the same way, I expected that the omission errors on both tasks were moderately associated (e.g., > .30). Finally, I supposed that different SSRT estimation methods would be associated with the reaction time of the Go/No-Go Task.

5.2 Material and Methods

5.2.1 Participants

The sample was composed of 207 Italian community dwelling adult participants with a mean age of 26.79 years (*SD* = 6.71 years; age range: 19 years – 61 years). In my sample eighty-seven participants (42.0%) were male, and 119 participants (57.5%) were female; one participant (0.5%) refuse to disclose his/her gender.. Five participants (2.4%) were left-handed. The sample was composed of two hundred three (98.1%) unmarried participants, 2 (1.0%) married participants, and 2 (1.0%) divorced participants. Eight participants (3.9%) had a junior high school degree, 98 (47.3%) had a high school degree, 87 (42.0%) had a university degree, and 13 (6.3%) had a post-lauream degree; one participant (0.5%) refuse to report his/her educational level. One hundred thirty participants (62.8%) were students, 34 (16.4%) were blue collars, 4 (1.9%) were white collars, 22 (10.6%) were managers, 15 (7.2%) were liberal arts practitioners, and 2 (1.0%) were unemployed.

Forty-six participants (22.2%) completed a shorter version of the Stop-Signal Task (i.e., four blocks, instead of six). Participants who completed the shorter version of the Stop-Signal Task did not differ significantly from participants who completed the longer version on age, $t(205) = -2.50$, $p > .86$, $d = -.35$, and gender, $\chi^2(2) = 3.68$, $p > .16$, Cramer $V = .13$. As expected, the two sample were significantly different on civil status, $\chi^2(2) = 7.61$, $p < .05$, Cramer $V = .19$, educational level, $\chi^2(3) = 21.54$, $p < .001$, Cramer $V = .32$, and occupation, $\chi^2(5) = 21.44$, $p < .01$, Cramer $V = .32$. The differences can be explained by the fact that the sample at issue was part of a different study (see, Gialdi et al, 2020) where only university students were included.

To be included in the sample, participants had to document that they were of adult age (i.e., 18 years of age or older), they had no psychiatric or neurological disorders, and had normal or corrected-to-normal vision, and to agree to the written informed consent in which the study was extensively described. To avoid cultural and lexical bias in questionnaire responses, to participate in the present study, participants were required to speak Italian as their first language. All participants were treated in accordance with the Ethical Principles of Psychologists and Code of Conduct.

5.2.2 Measures

- Stop-Signal Task (Verbruggen et al, 2019). As previously described, in this project, I relied on an open-source stop-signal paradigm to improve the replicability of the findings (Stop-it Task - The jsPsych version; Verbruggen et al, 2019). During this task, participants had to discriminate between two arrows: when a Go-trials appears (75% of the trials) participants have to respond as fast and accurate as possible. Conversely, on stop-signal trials (25% of the trials), after the apparition of the arrows, they are replaced by “XX” after a variable delay (Stop-Signal Delay; SSD), and participants have to inhibit their response. For an extensive description see paragraph 2.2.3.

- Go/No-Go Task (Bezdjian et al, 2009). During the Go/No-Go Task, participants have to respond to the presence of a target stimulus (i.e., letters P or R). In the first condition, participants are asked to press a key in response to the letter P (“Go” trials) and withhold their response to the letter R (“No-Go” trials); on the contrary, in the second condition participants are asked to make a response to the letter R (“Go” trials) and withhold their response to letter P (“No-Go” trials). For an extensive description see paragraph 2.2.2.

- Barratt Impulsiveness Scale – 11 (BIS-11; Patton et al, 1995). The BIS-11 is a 30-items self-report questionnaire designed to assess three facets of impulsivity. Items are rated on a 4-point Likert-type scale (*1 = Rarely/Never; 2 = Occasionally; 3 = Often; 4 = Almost Always/Always)*. The score of 4 indicates the most impulsive response, but in this questionnaire some items are reversed to avoid response biases. In the original study (Patton et al, 1995) and also in its Italian translation (Fossati et al, 2001), the authors have found three facets of impulsiveness: motor impulsivity, attention impulsivity, and non-planning impulsivity. These three facets are summed to produce a total score, and the higher the BIS-11 total score, the higher impulsivity level.

- UPPS-P Impulsive Behavior Scale (UPPS-P; Lynam et al, 2006). The UPPS-P is composed of 59 items, assessed on a 4-point Likert-type scale (*1 = Agree Strongly; 2 = Agree Some; 3 = Disagree Some; 4 = Disagree Strongly*). This questionnaire was designed to measure five dimensions of impulsive behavior: the tendency to commit rash actions as a result of intense negative affect (Negative Urgency; 12 items), the tendency to think and reflect on the consequences of an act (Premeditation; 11 items), the ability to remain with a task until completion (Perseverance; 10 items), the tendency to seek excitement (Sensation Seeking; 12 items), the tendency to act rashly in response to a positive mood (Positive Urgency; 14 items). The UPPS-P showed adequate psychometric properties (Cyders & Smith, 2007; Whiteside & Lynam 2001) also in its Italian translation (Fossati et al, 2016; Gialdi et al, 2021).

- Zuckerman-Kuhlman Personality Questionnaire Impulsive Unsocialized Sensation Seeking Scale (ImpSS; Zuckerman et al, 1993). The ImpSS is a self-report questionnaire, composed of 19 true-false items. This scale describes a lack of planning and a tendency to act on impulse without thinking, a need for excitement, change and novelty, and a preference for unpredictable situation. Carlotta and colleagues (2003) previously assessed the reliability and validity of the Italian translation of the ImpSS.

- Personality Inventory for *DSM-5* (PID-5; Krueger et al, 2012). The PID-5 is a 220-items self-report with a 4-point response scale (*0 = Very False or Often False; 1 = Sometimes or Somewhat False; 2 = Sometimes or Somewhat True; 3 = Very True or Often True*). The PID-5 was designed to assess the *DSM-5* traits presented in the Alternative Model of Personality Disorder (AMPD), provided in Section III (APA; 2013). The PID-5 has 25 scales that can be summed to generate five higher order dimensions (Krueger et al, 2012), which represents dysfunctional variants of the Five-Factor Model personality dimensions (APA, 2013). Specifically, the five domains of the PID-5 are: Negative Affectivity (i.e., frequent, and intense experiences of high levels of a wide range of negative emotions), Detachment (i.e., avoidance of socio-emotional experience), Antagonism (i.e., behaviors that put the individual at odds with other people), Disinhibition (i.e., orientation toward immediate gratification and impulsive behavior), and Psychoticism (i.e., a wide range of culturally incongruent odd, eccentric, or unusual behaviors and cognition). The psychometric proprieties of the Italian translation of the PID-5 in nonclinical adults have been previously published (Fossati et al, 2013).

For this project, in order to avoid a long and inaccurate compilation from participants, I relied only on the Disinhibition domain (46 items) and its corresponding facets (i.e., Distractibility, Impulsivity, Rigid Perfectionism, Risk Taking, Irresponsibility). The subsample previously described (see, paragraph 5.2.1), completed the shorter version of PID-5 with 100 items (PID-5-SF; Maples et al, 2015). However, previously studies (e.g., Somma et al, 2019) suggested that the reduced number of items of the PID-5-SF, identified by Maples and colleagues (2015), can be used to evaluate the domain scales and the corresponding facets also in its Italian translation.

5.2.3 Procedures

The questionnaires and the behavioral computerized tasks were randomly administered to the sample. In the whole project, all measures and tasks were administered in their Italian translation. In order to match the self-reports scores and tasks results and to maintain anonymity, each participant included in the sample created an alphanumeric ID code.

Participants completed the study online using Online Surveys Jisc, an online survey tool designed for academic research (https://www.onlinesurveys.ac.uk/); participants volunteered to take part in the study receiving no economic incentive or academic credit for their participation. Self-report measures were administered in random order and scored blind to the computerized task results. The computerized tasks were administered using a laptop computer in individual session and each session lasted on average two hours per participant.

Written informed consent was obtained prior to study participation; all participants were of adult age and volunteered to take part in the present study after it was extensively described. Institutional Review Board was obtained for all aspects of the study.

5.3 Data Analysis

Cronbach's alpha coefficient, Omega (ω) coefficient, and mean inter-item correlation (MIC) were used to estimate the internal consistency reliability of the self-report measures of impulsivity (i.e., BIS-11, UPPS-P, ImpSS, PID-5) in the whole sample. Spearman *r* coefficient was used to assess bivariate associations between neuropsychological indices and self-reported measures.

In the present study, I relied on both parametric and non-parametric methods to estimate SSRT (see paragraph 3 and 4 of the present Section). Specifically:

- *Non-parametric estimation methods*. Although the mean method is known to be biased (i.e., because of the skewness of the go RT distribution and by go omissions errors), I computed this method in order to assess if it provides meaningful associations with self-reported measures of impulsivity. Noteworthy, when the tracking procedure is used, it is still a widely used non-parametric estimation method because of its easiness of computation (Verbruggen et al, 2019). According to the mean method, SSRT can be estimated easily by subtracting mean SSD from mean RT on go trials (Verbruggen et al, 2019). Moreover, I computed non-parametric SSRT estimates based on the integration method with replacement of go omissions. Indeed, according to Verbruggen and colleagues' (2019) simulation study, I reported SSRT obtained by subtracting mean SSD from the n^{th} RT. To this aim, all go trials with a response were included (including go trials with a choice error and go trials with a premature response). Importantly, go omissions (i.e., go trials on which the participant did not respond before the response deadline) are assigned the maximum RT in order to compensate for the lacking response. Premature responses on unsuccessful stop trials (i.e., responses executed before the stop signal is presented) were also included when calculating p(respond|signal) and mean SSD (Verbruggen et al, 2019).

- *Parametric estimation methods.* Different from non-parametric methods, parametric methods allow for the estimation of the entire distribution of SSRTs (Matzke et al, 2013a). Bayesian parametric approach relies on Bayesian parameter estimation, and it relies on Markov chain Monte Carlo sampling (MCMC; Gilks et al, 1996; Gamerman & Lopes, 2006) to obtain posterior distributions for the go and stop parameters.

  For the purposes of the present study, the number of MCMC chains (i.e., sequences of values sampled from the posterior distribution of the parameters) was 3; the start values were automatically set to the maximum a posteriori probability estimates of the parameters. The total number of MCMC samples per chain was set at 20,000, the number of burn-in samples to discard at the beginning of each chain was 5,000, and a thinning factor of 5 was selected, meaning that only every 5th MCMC the sample was retained. In the present study, the central tendency of the posterior was used as a point estimate of the parameter. In particular, I relied on the median of the posterior distribution as point estimate for the parameters; the parameters posterior distributions quantify knowledge about the parameters after the data have been observed; the larger the dispersion, the greater the uncertainty in the estimated parameter. The 95% credible interval (i.e., 2.5th and 97.5th percentile of the distribution) were computed to quantify estimation uncertainty (e.g., Matzke et al, 2013a). To ensure that the chains have converged from their starting values to their stationary distributions, I verified that the posterior distributions of the model parameters were unimodal; then I run multiple MCMC chains and ascertain that the chains have mixed well. At convergence, the individual MCMC chains should look like “hairy caterpillars” and should be indistinguishable from one another. Lastly, I relied on the $\hat{R}$ (Gelman & Rubin, 1992) convergence diagnostic measure for each model parameter. $\hat{R}$ compares the between-chain variability to the within-chain variability. As a rule of thumb, $\hat{R}$ should be lower than 1.1 if the chains have properly converged.

  In the presents study I relied on two different BEESTS models (see paragraph 4.1 of this Section), namely, the BEESTS method (Matzke et al, 2013b), and the BEESTS method with trigger failure (Matzke et al, 2017b). The BEEST methods relied on a Bayesian parametric approach that allows for the estimation of the entire distribution of SSRTs. SSRTs are assumed to follow an ex-Gaussian distribution and Markov chain Monte Carlo sampling are used to estimate the parameters of the SSRT distribution (e.g., Matzke et al, 2013a). The BEESTS method with trigger failure enables researchers to simultaneously estimate the probability of trigger failures (i.e., deficiencies in triggering the stop process) and the entire distribution of stopping latencies (Matzke et al, 2017b); the resulting SSRT estimates are corrected for the bias that results from deficiencies in triggering the stop process (Matzke et al, 2017b).

For the purposes of the present study, I relied on a hierarchical estimation (e.g., Matzke et al, 2013a; Matzke et al, 2017b), so that the estimation of each individual's model parameters is informed by data from the entire sample, resulting in more precise and, on average, more accurate estimates of the true parameters (e.g., Farrell & Ludwig, 2008). Moreover, hierarchical modelling provides inference on both the participant and the population level (Farrell & Ludwig, 2008). The hierarchical approach has the potential to provide accurate parameter estimates with relatively few (i.e., 384 trials for participants) observations per participant (Matzke et al, 2013b).

Before estimating non-parametric and parametric SSRT, the data were screened for two basic validity checks: 384 recorded trials, and participant responses detected for at least one go trials. Moreover, in order to compute parametric SSRT, each individual data should be screened for performance exclusion criteria to ensure that participants included in BEESTS estimation analyses were adequately engaged in the task and had behavioral data that were broadly consistent with race model assumptions (e.g., Matzke et al, 2019). This procedure is consistent with previous studies (e.g., Skippen et al, 2019; Skippen et al, 2020; Weigard et al, 2021), and usually results in the exclusion of a significant number of participants (almost 16% of the original samples; see Skippen et al, 2019; Skippen et al, 2020; Weigard et al, 2021). In line with previous research (Skippen et al, 2019; Skippen et al, 2020; Weigard et al, 2022), in the present study, I relied on the following exclusion criteria:

- high overall error rate (i.e., >10%);
- go omission rate >10%;
- no error on go trials;
- excessive rate of omission (>.25) on go trials;
- staircase tracking failure to converge to 0.50 (i.e., these participants may be slowing or speeding of go responses);
- reaction times (RT) on unsuccessful stop trials numerically longer than RT on go trials. In line with Skippen and colleagues (2019), I compared the mean RT on unsuccessful stop trials with the mean RT on go trials; in this comparison I included all trials with a response (including choice errors and premature responses; Skippen et al, 2019);
- high response rate on shortest delay. Indeed, this results in extreme trigger failure rate and, thus, no sufficient information for parameter estimation.

5.4 Results

Detailed comparison between participants who completed a shorter version of the Stop-Signal Task and the PID-5-SF (n = 46) and participants who completed the longest version on the Stop-Signal Task and the PID-5 (n = 161) are reported in Table 5.1.

Table 5.1.

Participants who completed a shorter version of the Stop-Signal Task and the PID-5-SF (n = 46) versus participants who completed the longer version on the Stop-Signal Task and the PID-5 (n = 161): Detailed Comparisons on Demographic Variables.

	Longer Version of PID-5 (n=161)		Shorter Version of PID-5 (n=46)				
Demographic Variables	*n / M*	*% / SD*	*n / M*	*% / SD*	χ^2 / *t*	*df*	*V / d*
Civil Status:							
Unmarried	159	98.8	44	95.70			
Married	2	1.20	--	--			
Divorced	--	--	2	4.30	7.61	2	.19
Education Level:							
Junior High School	8	5.0	--	--			
High School	63	39.10	35	76.10			
University Degree	78	48.40	9	19.60			
Post-lauream Degree	12	7.50	1	2.20	21.54	3	.32
Profession:							
Student	91	56.50	39	84.80			
Blue Collar	34	21.10	--	--			
White Collar	4	2.50	--	--			
Manager	15	9.30	7	15.20			
Liberal Art Practitioner	15	9.30	--	--			
Retired	2	1.20	--	--	21.44	5	.32
Age (years)	27.40	5.97	24.63	8.58	-2.50	205	-.35

Note. V: Cramer's V coefficient; d: Cohen's d coefficient. --: statistic not present

The Cronbach's α values, Omega (ω) values, and mean inter-item correlation (MIC) for the BIS-11 subscale and total score in this sample are reported in Table 5.2.

Table 5.2

The BIS-11 Scales: Descriptive Statistics, Internal Consistency Reliability (i.e., Cronbach's α and ω Coefficient) Estimates, and Scale Intercorrelation ($N = 207$).

BIS-11 subscale	*M*	*SD*	α	ω	MIC
Motor impulsivity	20.17	4.63	.76	.88	.25
Attention impulsivity	22.60	3.94	.52	.71	.09
Non-planning impulsivity	16.83	3.98	.40	.77	.15
BIS-11 Total Score	59.59	9.05	.69	.84	.11

Note. MIC = Mean Inter-Item Correlation

The Cronbach's α values, Omega (ω) values, and mean inter-item correlation (MIC) for the UPPS-P subscale in this sample are reported in Table 5.3.

Table 5.3

The UPPS-P Scales: Descriptive Statistics, Internal Consistency Reliability (i.e., Cronbach's α and ω Coefficient) Estimates, and Scale Intercorrelation ($N = 207$).

UPPS-P subscale	*M*	*SD*	α	ω	MIC
Negative Urgency	26.82	7.04	.86	.91	.35
Premeditation	20.24	5.18	.86	.92	.38
Perseverance	18.06	4.98	.84	.92	.37
Sensation Seeking	28.19	8.49	.89	.93	.40
Positive Urgency	24.50	8.30	.92	.95	.47

Note. MIC = Mean Inter-Item Correlation

The Cronbach's α values, Omega (ω) values, and mean inter-item correlation (MIC) for the ImpSS scale total score in this sample are reported in Table 5.4.

Table 5.4

The ImpSS Scale: Descriptive Statistics, Internal Consistency Reliability (i.e., Cronbach's α and ω Coefficient) Estimates, and Scale Intercorrelation ($N = 207$).

ImpSS Scale	*M*	*SD*	α	ω	MIC
ImpSS Total Score	6.36	4.56	.86	.90	2.42

Note. MIC = Mean Inter-Item Correlation

The Cronbach's α values, Omega (ω) values, and mean inter-item correlation (MIC) for the PID-5 Disinhibition subscales in part of the sample are reported in Table 5.5.

Table 5.5

The PID-5 Facets: Descriptive Statistics, Internal Consistency Reliability (i.e., Cronbach's α and ω Coefficient) Estimates, and Scale Intercorrelation (N = 161).

PID-5 facet	*M*	*SD*	α	ω	MIC
Distractibility	.87	.68	.92	.96	.56
Impulsivity	.75	.64	.88	.96	.55
Rigid Perfectionism	1.53	.71	.91	.95	.50
Risk Taking	1.18	.62	.92	.94	.45
Irresponsibility	.49	.43	.71	.87	.28
Disinhibition	.95	.40	.89	.95	.16

Note. MIC = Mean Inter-Item Correlation

The Cronbach's α values, Omega (ω) values, and mean inter-item correlation (MIC) for the PID-5-SF Disinhibition subscales in part of the sample are reported in Table 5.6.

Table 5.6

The PID-5-SF Facets: Descriptive Statistics, Internal Consistency Reliability (i.e., Cronbach's α and ω Coefficient) Estimates, and Scale Intercorrelation (N = 46).

PID-5 facet	*M*	*SD*	α	ω	MIC
Distractibility	.90	.74	.91	.98	.70
Impulsivity	.81	.71	.90	.99	.70
Rigid Perfectionism	1.04	.82	.84	.98	.58
Risk Taking	.58	.56	.79	.98	.52
Irresponsibility	.46	.49	.61	.95	.29
Disinhibition	.94	.43	.85	.93	.24

Note. MIC = Mean Inter-Item Correlation

Convergent validity coefficients (i.e., Spearman's r coefficient) between the Go/No-Go Task and the Stop-Signal Task are reported in Table 5.7. It is important to note that SSRT has been evaluated with and without trigger failure, however these two estimations were highly correlated, at least in this sample, $r = .96$, $p < .01$.

Based on the criteria of Matzke and colleagues (2013b) previously described (see paragraph 5.3 in this Section), the sample included for the analysis of advanced estimations for SSRT was reduced to 168 participants (mean age = 26.68 years, *SD* = 5.87

years). This sample was composed of 78 male (46.4%) and 89 female (53.0%); one participant (0.6%) refused to disclose his/her gender. However, the sample to which could not be possible estimate the BEEST methods, was significantly older than the sample included for the analysis of BEEST, $t(205) = .48$, $p < .05$, $d = .07$. Though, Cohen's d value suggested that the difference between the mean scores was trivial, based on Cohen's conventional cut off values (Cohen, 1988), although it was significant. Moreover, the sample not included for the analysis was composed of more female participants, $\chi^2(2) = 7.50$, $p < .05$, Cramer $V = .19$.

Lastly, the sample was further reduced when BEEST with trigger failure analysis were computed. This third sample was composed of 123 participants (mean age = 27.15, SD = 5.95. The sample included 57 male (46.3%) and 65 female (52.8%); one participant (0.8%) refused to disclose his/her gender. Nevertheless, this sample did not show significant differences from the original sample on age, $t(205) = -.95$, $p > .40$, $d = -.13$, and on gender, $\chi^2(2) = 3.16$, $p > .20$, Cramer $V = .12$.

Table 5.7

The Stop-Signal Task: Correlations (i.e., Spearman's r coefficient) with the Go/No-Go Task ($N = 207$).

	Stop-Signal Task						
	p(r|s)	RT	Omission	SSRT Mean method	SSRT Integration method	SSRT BEEST[1]	SSRT BEEST WTF[2]
Go/No-Go Task	*r*	*r*	*r*	*r*	*r*	*r*	*r*
Omission errors	-.09	**.22**	**.31**	**.21**	.16	.14	.01
Commission errors	**.24**	-.13	-.01	-.08	-.00	-.00	-.01
RT	**-.27**	**.58**	**.35**	**.66**	**.60**	**.51**	**.48**

Note. p(r|s): probability of responding to the stop-signal; RT: reaction time; SSRT: stop-signal reaction time; WTF: with trigger failure. Bold indicate $p < .01$. 1: $n = 168$; 2: $n = 123$.

Table 5.8 represents the correlations coefficient (i.e., Spearman's *r* coefficient) between the Stop-Signal Task estimations and the disinhibition measures used in this sample.

Table 5.8

The Stop-Signal Task: Correlations (i.e., Spearman's r coefficient) with the disinhibition self-report measures (N = 207).

	Mean Method	Integration Method	BEEST[1]	BEEST WTF[2]
Self-report	*r*	*r*	*r*	*r*
BIS-11 Motor	-.02	-.03	.10	-.08
BIS-11 Attention	.01	-.01	.04	.04
BIS-11 Non-Planning	.13	.13	.08	.21
BIS-11 Total Score	.05	.04	.11	.07
UPPS-P NU	.08	.06	.12	.09
UPPS-P Prem	.06	.03	.13	-.04
UPPS-P Pers	-.02	-.05	.00	-.07
UPPS-P SS	.00	.01	.10	.03
UPPS-P PU	-.03	-.04	.06	.00
ImpSS Total Score	.04	.02	.16	.10
PID-5 Distractibility	-.19	-.26	-.04	-.28
PID-5 Impulsivity	-.21	-.24	.07	-.36
PID-5 Rigid Perfectionism	-.08	-.07	-.08	-.03
PID-5 Risk Taking	-.29	-.21	.14	-.42
PID-5 Irresponsibility	-.08	-.22	.06	-.02
PID-5 Disinhibition	-.16	-.28	.10	-.24

Note. Underlined correlations: p < .05. NU: Negative Urgency, Prem: Premeditation; Pers: Perseveration; SS: Sensation Seeking; PU: Positive Urgency. 1: n = 168; 2: n = 123.

5.5 Discussion

As a whole, the results of the present study suggested that relying on advanced methods for estimating SSRT may yield meaningful associations between neuropsychological task (i.e., the Stop-Signal Task) and self-report measures of impulsivity. Rather, adopting computationally-intensive methods for estimating the SSRT (i.e., the BEEST methods) did not seem to provide advantages in terms of identifying relationships between neuropsychological tasks (i.e., the Stop-it task and the Go/No Go Task).

As expected, the Stop-Signal Task was significantly associated with the Go/No Go Task on the omission errors, mean reaction time, and on the probability of responding to the signal. Specifically, the omission errors (i.e., the lack of response when the go stimulus appears) of the SST moderately correlated with the omission errors of the Go/No Go Task with Spearman's r coefficient of .31, $p < .01$. In addition, the mean reaction time (i.e., the time between the beginning of the presentation of the stimulus and the beginning of the participants' reaction to it) of both tasks were moderately correlated, $r = .58$, $p < .01$. However, it is important to consider the size of these correlation coefficients in the light of available meta-analytic data on the relationships between neuropsychological measures and self-reports (e.g., Sharma et al, 2014). Lastly, the probability of responding to the stop signal (indicating the possibility to commit commission errors) in the Stop-Signal Task showed a significant relationship with the commission errors (i.e., a response has been executed when no response was necessary) of the Go/No Go Task, $r = .24$, $p < .01$. Despite existent literature is debated about the distinction (Raud et al, 2020) or similarity (e.g., Eagle et al, 2008) between these two tasks, this study showed that the Stop-Signal Task and the Go/No Go Task can be both used as reliable computerized task to assess the principal aspect of disinhibition behavior (i.e., reaction time, omission, and commission errors) and response inhibition, at least in this sample. This study confirmed the work of Eagle and colleagues (2008) on the overlap between the two tasks, while suggesting that they may also map onto different domains (e.g., attention).

In addition, Stop-Signal Reaction Time estimates, assessed with both non-parametric and parametric approach, showed moderately and significant associations with the reaction time of the Go/No Go Task. Moreover, differently from previous studies (e.g., Skippen et al, 2019) non-parametric approach yielded higher correlations than parametric approach, suggesting that the use of non-parametric estimations may be also valuable. Specifically, in this sample, the mean and the integration methods significantly showed correlations with the reaction time of the Go/No Go Task. In addition, also BEEST methods significantly showed associations with the mean reaction time of the Go/No Go Task for BEEST without trigger failure and BEEST with trigger failure, respectively. Indeed, it should be observed that Skippen and colleagues (2019) relied on a longer version of the Stop Signal Task (i.e., 700 trials).

According to present literature (e.g., Sharma et al, 2014) the correlation between behavioral tasks and self-reports is modest. However, it is also important to consider different types and more advanced estimation as demonstrated in previous studies (e.g., Skippen et al, 2019). In this sample, advanced estimation (i.e., BEEST with trigger failure) yielded significant correlation with lack of planning, as operationalized with BIS-11, r = .21, $p < .05$, indicating that people with difficulties in planning action present difficulties in inhibiting their response when it is required. Moreover, different estimations of the SSRT produced different correlations with the facets of the PID-5. Specifically, considering the mean method, SSRT was negatively and significantly correlated with Impulsivity and Risk-Taking, r = -.21, and r = -.29, all $p < .05$. Considering, the integration method, SSRT was negatively and significantly correlated with Distractibility, Impulsivity, Risk-Taking and Irresponsibility, r = -.26, r = -.24, r = -.21, and r = -.22, all $p < .05$. Lastly, only BEEST with trigger failure showed significant associations with three facets of PID-5, r = -.28, r = -.36, and r = -.42, all $p < .05$, namely Distractibility, Impulsivity and Risk-Taking. All these results indicated that participants with high level of disinhibition, showed a lower reaction time in the Stop-Signal Task, and thus a faster response. The present results are consistent with previous data (Gialdi et al, 2021; Skippen et al, 2019), demonstrating that better relationships between self-report and laboratory tasks may be better identified with advanced estimation of SSRT, as shown with increasing correlations found moving from non-parametric estimations to parametric estimations of SSRT.

5.6 Limitations

The results of the present study should be considered in the light of several important limitations. Although it was larger than previous reports (e.g., Skippen et al, 2019), the sample size of my study is quite limited. Indeed, administering neuropsychological tasks requires a great deal of time (i.e., more than one hour per participants), and resources (i.e., all participants were administered the measures in the laboratory) (e.g., Crawford & Garthwaite, 2002). Moreover, it should be observed that Bayesian parametric estimation requires extremely long estimation time and meeting stringent inclusion criteria to allow model convergence (e.g., Matzke et al, 2019). Accordingly, I tried to find a balance between having an adequate sample size and allowing parametric estimation of SSRT.

A subsample of participants was derived from a different study (see Gialdi et al, 2020), and included only university students; although this represented a limitation of the current study, I collected those data in the early phase of the project. Additionally, my sample was more a convenient study group than a sample representative of the Italian population. Of course, this can limit the generalizability of the findings to other populations (i.e., clinical, or forensic sample).

Notably, different stop-signal paradigms are available. As mentioned before, I relied on an open-source software for administering the stop-it task (Verbruggen et al, 2019). Of course, changing the task parameters and relying on different stop-signal paradigms may result in different findings. I computed SSRT based on the integration method with replacement of go omissions because it showed to be less biased in previous simulation studies (e.g., Verbruggen et al, 2019). Relying on different non-parametric methods for estimating the SSRT may yield different findings. Despite it suffers from several limitations (Verbruggen et al, 2019), I considered also the mean method for computing SSRT. This method choice was related to the popularity of the mean method; in my opinion, its inclusion could be useful for comparison purposes.

Finally, it should be observed that I did not compute SSRT based on the EXG3 model (Matzke et al, 2019). Indeed, Matzke and colleagues (2019) proposed a parametric framework that extends the standard 2-runner race model to account for go errors, and hence expand the scope of the stop-signal paradigm to the study of response inhibition in the context of difficult choices (Heatcote et al, 2019). This approach is based on Bayesian approach based on the ex-Gaussian distribution – the EXG3 model (Heathcote et al, 2018; Matzke et al, 2019). Interestingly, Matzke and colleagues (2019) showed that the EXG3 approach can be successfully applied to stop-signal tasks with high error rates; however, this model requires novel stop-signal data with high error rates and a manipulation of task difficulty to enable researchers to study difficult-choice inhibition. Because of the large number of trials needed for the EXG3 estimation (e.g., 700 trials in Skippen and colleagues' [2019] study) and because it was meant to extends the scope and applicability of the stop-signal paradigm to the study of response inhibition in the context of difficult choices (i.e., a more demanding go task), I did not rely on this estimation method, which is not consistent with common stop-signal paradigm (Verbruggen et al, 2019). Indeed, previous studies showed that healthy participants have low trigger failure rates (< 9%; Verbruggen et al, 2020), low inter-individual variability and substantial ceiling effects (e.g., Matzke, et al, 2017) on "traditional" stop-it tasks (i.e., the tasks I was interested in). Indeed, making the task longer and more difficult result in obtaining the variability in trigger failure that allows for the analysis of individual differences in response inhibition parameters; however, the resulting tasks is quite different from the results on the stop-it tasks used in previous studies (e.g., Sharma et al, 2014)

As to the other measures, I relied only on self-reports to assess impulsive behaviors and only on the Disinhibition domain of PID-5. Although all the selected self-report measures used in the present study are reliable and valid measures of impulsivity, using different measures, including interviews may yield different results.

6. Study 2

The second study of this first section on the Stop-Signal Task focused on the reliability of the mentioned task. Assessing reliability of behavioral tasks, is important for both clinical applications and research practice, since measures with low reliability are unsuitable and cannot predict clinical outcomes. Indeed, Sharma and colleagues (2014) stated that if laboratory tasks cannot identify a particular construct and are provided with low reliability, it critical to rely on these tasks both in research and clinical application. Moreover, theoretical conclusion on a particular construct can be modified if clinicians and researchers taking into account low reliability estimates (Hedge et al, 2018). It is therefore important to consider this aspect as well when selecting a task.

A meta-analysis of Enkavi and colleagues (2019) showed that the psychometric proprieties of the Stop-Signal Task and others neuropsychological tasks, are under studied. Moreover, the studies that consider the reliability of the tasks are usually conducted with children (e.g., Alderson et al, 2008; Soreni et al, 2009). Reliability is an important criterion of any psychological measure (Anastasi & Urbina, 1997) and necessary for its validity.

In line with neuropsychological research, in this study I assessed reliability estimations in terms of temporal stability (i.e., Spearman *r* coefficient) and Intraclass Correlation Coefficient (ICC) for absolute agreement based on random-effect one-way ANOVA. However, in this study I relied also on internal consistency (i.e., Cronbach's alpha and Omega coefficient) coefficients, considering neuropsychological tasks as psychological tests, and consequently considering each Stop-Signal Task block as a part of a test.

Current literature, in addition to be poor, also has contrast findings. Kindlon and colleagues (1995) and more recently Soreni and colleagues (2009), have found good test-retest reliability of SSRT in the Stop-Signal Task; on the contrary, Wöstmann and colleagues (2013) have found poor test-reliability. However, all these studies have been carried out on children.

Considering present literature and the increasing number of studied on the Stop-Signal Task, it is important to assess its temporal stability, looking for the best way to test reliability for this type of neuropsychological task.

6.1 Aim

The aim of the present study was to assess test-retest reliability of the Stop-Signal Task, considering the Stop-Signal Reaction Time estimated with both non-parametric and parametric approach, and internal consistency reliability across the six blocks of the Stop-Signal Task with a three-months test-retest paradigm in a un sub-sample of participants (n = 114), who agreed to take part at a three-months follow-up assessment.

In this study, Spearman's r coefficient, Cronbach's α and McDonald omega coefficients have been calculated.

In line with previous study (Kindlom et al, 1995; Sharma et al, 2014; Soreni et al, 2009; Wöstmann et al, 2013), I expected that the Stop-Signal Task has moderately low temporal stability when tested with Spearman's r coefficient. However, since the Stop-Signal Task is provided with good criterion validity, I also expected that internal consistency reliability was adequate in the two different time of administration.

6.2 Material and Methods

6.2.1 Participants

The sample was composed of 114 Italian community dwelling adult participants with a mean age of 25.71 years (*SD* = 4.89 years; age range: 19 years – 46 years), who agreed to take part to the three-months follow-up. In my sample 51 participants (44.7%) were male, and 62 participants (54.4%) were female; one participant (0.9%) refuse to disclose his/her gender. Three participants (2.6%) were left-handed. The sample was composed of one-hundred twelve (98.2%) unmarried participants, and 2 (1.8%) married participants. Six participants (5.3%) had a junior high school degree, 49 (43.0%) had a high school degree, 54 (47.4%) had a university degree, and 5 (4.4%) had a post-lauream degree. Seventy-six participants (66.7%) were students, 18 (15.8%) were blue collars, 2 (1.8%) were white collars, 10 (8.8%) were managers, and 8 (7.0%) were liberal arts practitioners.

Participants who completed the follow-up of the Stop-Signal Task were significantly younger then participants who did not completed the three-months administrations of the Stop-Signal Task, $t(205) = 2.59$, $p < .00$, $d = .35$. However, the two sample did not significantly differ on gender, $\chi^2(2) = 1.68$, $p > .43$, Cramer $V = .09$, civil status, $\chi^2(2) = 4.08$, $p > .13$, Cramer $V = .14$, educational level, $\chi^2(3) = 5.47$, $p > .14$, Cramer $V = .16$, and occupation, $\chi^2(5) = 4.00$, $p > .55$, Cramer $V = .14$.

To be included in the sample, participants had to agree to the written informed consent in which the study was extensively described. To avoid cultural and lexical bias in questionnaire responses, to participate in the present study, participants were required to speak Italian as their first language. All participants were treated in accordance with the Ethical Principles of Psychologists and Code of Conduct.

6.2.2 Measures

- Stop-Signal Task (Verbruggen et al, 2019). See paragraph 5.2.2 and 2.2.3 for an extended description.

6.2.3 Procedure

Participants who agreed to take part to the second part of this study completed only the computerized Stop-Signal Task, in its Italian translation. In order to match the task results and to maintain anonymity, each participant included in the sample created the originally alphanumeric ID code, following the same instructions in the two different time of administrations (i.e., first letter of the mother's name, first letter of the father's name, number of letters of the surname, day of birth).

Participants volunteered to take part in the second part of the study receiving no economic incentive or academic credit for their participation. The Stop-Signal Task was administered using a laptop computer in individual session, and each session lasted on average two hours per participant.

Written informed consent was obtained prior to study participation; all participants were of adult age and volunteered to take part in the present study after it was extensively described. Institutional Review Board was obtained for all aspects of the study.

6.3 Data Analysis

Spearman *r* coefficient and Intraclass Correlation Coefficient (ICC) were used to assess temporal stability. Correlations of .70 or higher indicate good test-reliability, and correlations between .40 and .60 indicate moderate test-retest reliability. Good test-retest reliabilities are a necessary prerequisite for the validity of any measure (Becser et al, 1998; Kuntsi et al, 2005). *t*-test comparisons were also assessed to evaluate the differences in mean in the two times of administration. Cohen's *d* statistic was used to evaluate the effect size of the *t*-tests.

Whereas Intraclass Correlation Coefficients (ICC) was used with a one-way random effects model for absolute agreement. Moreover, to assess reliability I relied also on Cronbach's α coefficient and Omega (ω) coefficient for the internal consistency of the Stop-Signal Task. High internal consistency of a measure indicates high homogeneity of the scale. To this aim, in this study I considered the 6 different blocks of the Stop-Signal Task as part of a test.

As mentioned for Study 1 (see paragraph 5.3 of this Section for an extended description) I relied on both parametric and non-parametric methods to estimate SSRT (see paragraph 3 and 4 of this Section). For non-parametric method I computed the mean method, even it is known to biased, by subtracting mean SSD from mean RT on go trials (Verbruggen et al, 2019), and the integration method with replacement of go omissions. In this second method, according to Verbruggen and colleagues' (2019) simulation study, I reported SSRT obtained by subtracting mean SSD from the n^{th} RT. To this aim, all go trials with a response were included (including go trials with a choice error and go trials with a premature response). Importantly, go omissions (i.e., go trials on which the participant did not respond before the response deadline) are assigned the maximum RT in order to compensate for the lacking response. Premature responses on unsuccessful stop trials (i.e., responses executed before the stop signal is presented) were also included when calculating *p*(respond|signal) and mean SSD (Verbruggen et al, 2019). On the contrary, parametric methods allow for the estimation of the entire distribution of SSRTs (Matzke et al, 2013a), and are based on Bayesian parametric estimation and on Markov chain Monte Carlo sampling (MCMC; Gilks et al, 1996; Gamerman & Lopes, 2006) to obtain posterior distributions for the go and stop parameters.

In the presents study I relied on two different BEESTS models (see paragraph 4.1 of this Section), namely, the BEESTS method (Matzke et al, 2013a), and the BEESTS method with trigger failure (Matzke et al, 2017b). Similarly, to the Study 1 (see paragraph 5.3 of this Section), in the present study I relied on a hierarchical estimation (e.g., Matzke et al, 2013a; Matzke et al, 2017b), so that the estimation of each individual's model parameters is informed by data from the entire sample, resulting in more precise and, on average, more accurate estimates of the true parameters (e.g., Farrell & Ludwig, 2008). The hierarchical approach has the potential to provide accurate parameter estimates with relatively few (i.e., 384 trials for participants) observations per participant (Matzke et al, 2013).

6.4 Results

The descriptive statistics, Spearman *r* coefficient, and Intraclass Correlation Coefficient (ICC) for the Stop-Signal Task's mean method and integration method in two time of administrations are reported in Table 6.1 and Table 6.2, respectively. For this analysis I considered both the general estimations of the Stop-Signal Reaction Time (e.g., mean method for the entire task) and the single estimation of each block.

Table 6.1

The Stop-Signal Task test-retest reliability: Descriptive Statistics, Spearman r coefficient, and Intraclass Correlation Coefficient (ICC) in two time of administrations with mean method estimation (n = 114).

	Time 1		Time 2					
Stop-Signal Task	*M*	*SD*	*M*	*SD*	$t(113)$	*d*	*r*	ICC
Mean Method	332.99	92.52	328.66	82.88	0.55	.10	.55**	.55**
Mean Method 1	346.88	96.99	335.54	90.29	1.17	.22	.42**	.39**
Mean Method 2	327.74	93.88	330.63	90.64	-0.33	-.06	.53**	.49**
Mean Method 3	327.60	93.36	330.38	86.96	-0.32	-.06	.49**	.46**
Mean Method 4	325.45	95.25	324.32	94.69	0.11	.02	.44**	.34**
Mean Method 5	329.07	90.71	328.82	90.72	0.03	.01	.53**	.52**
Mean Method 6	327.85	95.97	324.31	87.51	0.38	.07	.47**	.42**

Note. ** $p < .01$

Table 6.2

The Stop-Signal Task test-retest reliability: Descriptive Statistics, Spearman r coefficient, and Intraclass Correlation Coefficient (ICC) in two time of administration with integration method estimation (n = 114).

	Time 1		Time 2					
Stop-Signal Task	*M*	*SD*	*M*	*SD*	$t(113)$	*d*	*r*	ICC
Integration Method	312.80	85.84	314.55	83.199	-0.22	-.04	.52**	.49**
Integration Method 1	334.67	104.88	315.04	91.574	1.19	.22	.44**	.34**
Integration Method 2	319.30	92.31	321.75	96.15	-0.25	-.05	.44**	.39**
Integration Method 3	312.54	95.48	320.24	88.097	-0.77	-.14	.36**	.33**
Integration Method 4	311.25	99.91	319.49	98.085	-0.75	-.14	.41**	.30**
Integration Method 5	314.32	88.76	322.02	96.946	-0.79	-.15	.47**	.38**
Integration Method 6	314.75	93.18	316.39	97.466	-0.16	-.03	.46**	.37**

Note. ** $p < .01$

The descriptive statistics, Spearman *r* coefficient, and Intraclass Correlation Coefficient (ICC) for the Stop-Signal Task's BEEST method in two time of administrations are reported in Table 6.3 and Table 6.4.

For the same reasons presented for Study 1 (see paragraph 5.4 of this Section) the sample for the analysis with BEEST method was reduced to 85 participants (mean age = 25.69 years, SD = 5.11 years). In this sample 41 participants (48.2%) were male, and 43 participants (50.6%) were female; one participant (1.2%) refuse to disclose his/her gender. Two participants (2.4%) were left-handed. Participants to which BEEST could be calculated did not significantly differ on age $t(113) = .06$, $p > .90$, $d = .01$, and on gender $\chi^2(2) = 2.51$, $p > .28$, Cramer $V = .15$.

However, the number of trials for estimating BEEST with trigger failure wasn't enough for all the participants. For this reason, the sample for the estimation of the test-retest reliability with BEEST with trigger failure was reduced again to 49 participants (mean age = 26.41 years, SD = 5.47 years). In this sample 23 participants (46.9%) were male, and 25 participants (51.0%) were female; one participant (2.0%) refuse to disclose his/her gender. Also in this case, participants with enough trials to assess BEEST with trigger failure, did not significantly differ on age $t(83) = -1.51$, $p > .70$, $d = -.33$, and on gender $\chi^2(2) = 78$, $p > .65$, Cramer $V = .10$.

Table 6.3

The Stop-Signal Task test-retest reliability: Descriptive Statistics, Spearman *r* coefficient, and Intraclass Correlation Coefficient (ICC) in two time of administration with BEEST estimations (n = 85).

	Time 1		Time 2					
Stop-Signal Task	*M*	*SD*	*M*	*SD*	*t*(84)	*d*	*r*	ICC
BEEST	330.70	88.91	329.85	87.96	.08	0.02	.51**	.44**

Note. ** $p < .01$

Table 6.4

The Stop-Signal Task test-retest reliability: Descriptive Statistics, Spearman *r* coefficient, and Intraclass Correlation Coefficient (ICC) in two time of administration with BEEST WTF estimations (n = 49).

	Time 1		Time 2					
Stop-Signal Task	*M*	*SD*	*M*	*SD*	*t*(48)	*d*	*r*	ICC
BEEST WTF	358.92	70.72	340.23	73.80	1.58	0.46	.30*	.29*

Note. WTF: with trigger failure

* $p < .05$

The descriptive statistics, Cronbach's α values, Omega (ω) values for the Stop-Signal Task in the two times of administrations for the mean method and the integration method are reported in Table 6.5 and Table 6.6, respectively.

Table 6.5

The Stop-Signal Task test-retest reliability: Descriptive Statistics, Cronbach's α values, Omega (ω) values in two time of administrations with mean method estimation (n = 114).

	Time 1		Time 2	
Stop-Signal Task	*M*	*SD*	*M*	*SD*
Mean Method 1	346.88	96.99	335.54	90.29
Mean Method 2	327.74	93.88	330.63	90.64
Mean Method 3	327.60	93.36	330.38	86.96
Mean Method 4	325.45	95.25	324.32	94.69
Mean Method 5	329.07	90.71	328.82	90.72
Mean Method 6	327.85	95.97	324.31	87.51
α	.96		.96	
ω	.99		.99	

Table 6.6

The Stop-Signal Task test-retest reliability: Descriptive Statistics Cronbach's α values, Omega (ω) values in two time of administrations with integration method estimation (n = 114).

	Time 1		Time 2	
Stop-Signal Task	*M*	*SD*	*M*	*SD*
Integration Method 1	334.67	104.88	315.04	91.574
Integration Method 2	319.30	92.31	321.75	96.15
Integration Method 3	312.54	95.48	320.24	88.097
Integration Method 4	311.25	99.91	319.49	98.085
Integration Method 5	314.32	88.76	322.02	96.946
Integration Method 6	314.75	93.18	316.39	97.466
α	.95		.95	
ω	.98		.98	

6.5 Discussion

As a whole, the results of the present study suggested that test-retest reliability assessed with Spearman *r* coefficient and Intraclass Correlation Coefficient (ICC), yielded modest associations between two times of administrations, also with the advanced estimations of SSRT (i.e., BEEST methods). However, Cronbach's alpha and Omega coefficients, showed good internal consistency in both times of administrations.

Confirming and extending previous results (e.g., Wöstmann et al, 2013), and in contrast to findings from other studies (Kindlon et al, 1995; Soremi et al, 2009), the Stop-Signal Task was provided with moderate test-retest reliability of SSRT, Spearman *r* coefficient ranging from .30 (BEEST WTF estimation) to .55 (mean method estimation), at least in this sample. Moreover, in this analysis it is shown that the mean method is shown to be the most stable method in this three-months test-retest paradigm, showing that the use of advanced computational models of SSRT estimates does not necessarily lead to better results. In my sample, also, the results with BEEST WTF showed even lower correlations. However, this may be due to the fact that being a more complex analysis, more information is lost during data processing. On the other hands, this result can be important, since the use of more complex and longer computational models can be difficult for some researchers to assess (e.g., Hedge et al, 2018) and for some participants to complete.

These results also show that there may be variables between the two times of administration that are not taken into account during classical temporal stability analyses, such as learning effect and individual differences. These results are also consistent with the work of Hedge and colleagues (2018), according to which robust cognitive tasks, widely used is psychology and neuropsychology, may be influenced by between subject variability. However, these controversial aspects are mostly ignored in cognitive and neuropsychology (Hedge et al, 2018).

For this reason, in order to evaluate the reliability of the Stop-Signal Task, I estimated two different internal consistency coefficients (i.e., Cronbach's alpha and Omega), since considering these types of estimations may yield to different conclusion about the reliability of the Stop-Signal Task. In this study I considered the six blocks of the Stop-Signal Task as a part of a test and the I assessed the internal consistency in the two time of administrations with the SSRT estimations evaluated with the mean method and the integration method. The result of this analysis showed that the Stop-Signal Task is highly reliable with Cronbach's alpha of .95 for the integration method and .96 for the mean method in both time of administration. Similarly, the Omega coefficients showed overlapping results, with ω estimations of .98 for the integration methods and .99 for the mean method. High internal consistency of the task in a three-month period of administrations indicates high homogeneity of the Stop-Signal Task.

The present results confirmed the assumption of Hedge and colleagues (2018) that different types of estimation to assess the reliability of cognitive and neuropsychological tasks may be a better way to understand the psychometric proprieties of a task.

6.6 Limitations

The results of the present study should be considered in the light of several important limitations. First of all, for the purpose of this second study, the sample size was quite limited. Although administering neuropsychological tasks requires a great deal of time (i.e., more than one hour per participants), and resources (i.e., all participants were administered the measures in the laboratory) (e.g., Crawford & Garthwaite, 2002), the drop-out of the participants in the two times of administrations was quite elevated. Moreover, the use of Bayesian parametric estimation also caused the loss of additional participants. Lastly, my sample was more a convenient study group than a sample representative of the Italian population. Of course, this can limit the generalizability of the findings to other populations (i.e., clinical, or forensic sample).

As I mentioned before, I relied on an open-source software for administering the stop-it task (Verbruggen et al, 2019), and relying on different stop-signal paradigms may result in different findings. Notwithstanding their limitations, I computed SSRT based on the mean method and on the integration methods (Verbruggen et al, 2019), and relying on different non-parametric methods for estimating the SSRT may yield different findings. However, the results of the present study based on non-parametric and parametric estimations yielded very similar results, confirming Band and colleagues' simulation study (2013) that SSRT estimated with the mean method are most reliable when the tracking procedure is used.

Lastly, a test-retest reliability depends heavily on the length of the retest period (Kaplan & Saccuzzo, 2017; Polit, 2014). Therefore, it is important to consider that underlying processes can change rapidly in a three-month test-retest paradigm, and this can explain poor test-retest reliability

7. Future Work

Based on the results of the two previous studies, future work should replicate these findings in different samples. The possibility of the replication of the results is simplified by the use of the free PEBL platform and the open-source software used to administer the Stop-Signal Task. It would also be important to replicate the results of this study in clinical population, in order to extend the generalizability of the findings and to assess the reliability of the task also in different population (e.g., patients with difficulties in inhibiting responses).

In addition, future studies should also consider the possibility to use different computerized task measures to assess response inhibition (e.g., different stop-signal paradigm). Also, convergent and discriminant validity between the Stop-Signal Task may be assessed with other behavioral tasks, as for example attention task. These correlations may allow researchers to estimate the presence of difficulties in the sustained attention, that can therefore lead to different results also in a task of response inhibition. For example, if future studies will rely on the EXG3 model (Matzke et al, 2019), it will be interesting to study the relationship with a difficult choices' tasks (i.e., a Stop-Signal Task with almost 700 trials: Skippen et al, 2019) and decision-making task.

Most importantly, future studies should rely on a different paradigm to assess test-retest reliability. Based on Hedge and colleagues' study (2018), between two times of administrations there are numerous variables that can influence the performance and are not considered in temporal stability paradigm (e.g., individual differences). According to these colleagues, subject variability may influence the reliability of neuropsychological tasks. Indeed, future researchers should rely on alternative ways to assess the longitudinal stability of the disinhibition.

8. General Conclusion

As a whole, the results of this Section on the Stop-Signal Task suggested that (a) although relying on advanced methods for estimating SSRT may yield meaningful associations between neuropsychological task, this does not result in improved reliability estimates; (b) adopting computationally-intensive methods for estimating the SSRT (i.e., the BEEST methods) did not seem to provide advantages in terms of identifying relationships between neuropsychological tasks (i.e., the Stop-it task and the Go/No Go Task). Moreover, these studies extended the present literature about the temporal stability. Indeed, the second study of this Section showed that test-retest reliability assessed with Spearman *r* coefficient and Intraclass Correlation Coefficient (ICC), yielded modest associations between two times of administrations (i.e., r = .55 with the mean method estimation), also with the advanced estimations of SSRT (i.e., BEEST methods). Rather, the advanced estimations method for SSRT provided similar (i.e., r = .51 with BEEST methods) or even lower (i.e., r = .30 with BEEST with trigger failure methods) correlations. However, my studies showed that relying on a different approach to assess test-retest reliability provided better results. Indeed, relying on Cronbach's alpha and Omega coefficients, to assess the internal stability of the task in two times of administrations, provided good internal consistency in both times.

Thus, the results of this Section suggested that the Stop-Signal Task may be provided with good reliability, assessed with the internal consistency coefficient, and convergent validity, assessed with self-reports of disinhibition and similar computerized behavioral task (i.e., the Go/No-Go Task). Moreover, these results also suggest the importance to evaluate the psychometric proprieties of a behavioral task, mostly used with clinical application, before the evaluation of more complex and advanced cognitive modelling.

SECTION II: IOWA GAMBLING TASK

1. Introduction

The Iowa Gambling Task (IGT) was developed by Bechara and colleagues (1994), to examine decision making impairment. This task allows researchers to assess participants' sensitivity to rewards and losses and their ability to make decision under uncertainty. Moreover, several studies have shown that the IGT measures cognitive impulsivity, such as the inability to delay gratification and evaluate the outcome of a planned action (e.g., Bechara et al, 2000). In this second Section, several works about the Iowa Gambling Task will be presented. Specifically, Chapter 2 presents bibliographic research where risk taking, and decision-making theories are described (Chapter 2.1). Then (Chapter 2.2), risk taking paradigms used in this project are presented and described in detail (i.e., Balloon Analogue Risk Task [Lejuez et al, 2002] and Iowa Gambling Task [Bechara et al, 1994]). Chapter 3 focuses on the most used estimations of the IGT, that include the number of selections from each deck, the number of advantageous and number of disadvantageous selections, and the difference between the disadvantageous choices and the advantageous choices (i.e., net score; Buelow & Suhr, 2009). Two different IGT estimations will be briefly presented in Chapter 4, namely, the Expectancy-Valence Learning Model (Busemeyer & Stout, 2002) and the Prospect Valence Learning Model (Ahn et al, 2008). Finally, three studies on the psychometric proprieties of the Iowa Gambling Task will be presented: the first study aims at assessing the convergent validity between the IGT and self-report measures of risk taking and at testing if one of the different models proposed for estimating the IGT would show larger convergent validity with self-reports of disinhibition (i.e., the UPPS-P Impulsive Behavior Scale [UPPS-P; Cyders & Smith, 2007], Impulsive Sensation Seeking [ImpSS; Zuckerman et al, 1991], Barratt Impulsiveness Scale-11 [BIS-11; Patton et al, 1995] and Personality Inventory for *DSM-5* [PID-5; Krueger et al, 2012]); moreover this study aims at assessing the convergent validity between the IGT and the Balloon Analogue Risk Task (BART). Also in this study, IGT scores will be computed using different measurement models to show their convergence with lab tasks. The second study aims at assessing the test-retest stability of the IGT with a three-months test-retest reliability paradigm. In this study I relied on three different approaches to assess the reliability estimations: test-retest reliability (i.e., Spearman *r* coefficient), Intraclass Correlation Coefficient (ICC), and internal consistency estimations. The last study aimed at assessing the possible relationship between a decision-making task, as it is operationalized with the Iowa Gambling Task, and a impulsivity task, as it is operationalized with the Stop-Signal Task.

This section will be concluded with future direction (Chapter 8) and a general conclusion about the Iowa Gambling Task (Chapter 9).

2. Literature Review

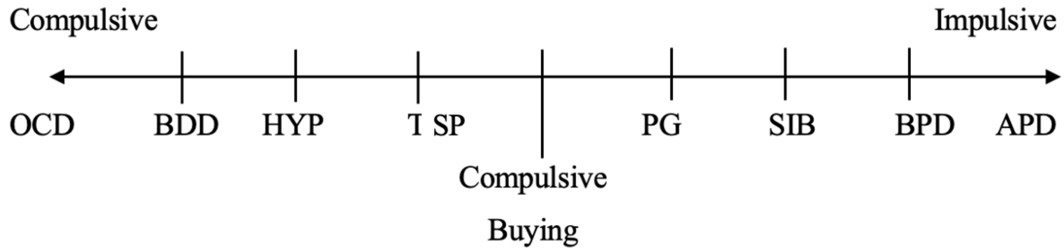
2.1 Impulsivity

The concept of impulsivity has been considered from several authors a multidimensional construct that affect behavior in many life domains, for example, from recreational activities to important decision making (for a review, Sharma et al, 2014). In fact, our life involves numerous decisions that we have to make every day, but the process of the human decision making is not a uniform, easy, or straightforward (Phillips et al, 2016). For example, a lot of people will consider wrong, risky, or rash, others' choices. Moreover, an individual can make a choice based on an erroneous logic, and, anyway, may feel compelled to make that nonoptimal choice.

As mentioned before in Section I for inhibition, even for the concept of impulsivity there are difficulties for a clear definition. In concrete terms, it happened to anyone, to have engaged in impulsive behavior, whether it was taking an extra drink on a party, an extra purchase on online shops, going on a last-minute trip, or more simply, just stopping and chatting to a friend met unexpectedly in the street. Other authors defined impulsivity as a series of actions without conscious judgment (Hinslie & Shatzky, 1940), behaviors without enough thought (Smith, 1952), etc. The two authors Eysenck and Eysenck (1977), originally classified impulsivity in a dimensional system of personality, identifying specific subtraits (i.e., risk-takin, lack of planning, and "making up one's mind quickly"; Moeller et al, 2001).

An interesting work of Hollander and Rosen (2000) describes a continuum between compulsive and impulsive behaviors dimensions (see also, Stein et al, 1994; Stein et al, 1996; Stein, 2000). Thus, behaviors like the previously mentioned can be classified on the impulsive end. The representation of this continuum can be describe as shown in Figure 2.1, where Obsessive-Compulsive Disorder (OCD) is the opposite polar of the Antisocial Personality Disorder (PD). As shown in this figure, other impulsive behaviors can be characterized by self-injurious behaviors (SIB), sexual addictions, pyromania, and impulsive-personality disorders (e.g., Borderline Personality Disorder [BPD]; Hollander, 1998).

Figure 2.1. Continuum between Compulsivity and Impulsivity



Note. Adapted from Hollander and Rosen (2000).
OCD: obsessive-compulsive disorder; BDD: body dysmorphic disorder; HYP: hypochondriasis; SP: skin picking; PG: pathological gambling; SIB: self-injurious behavior; BPD: borderline personality disorder: APD: antisocial personality disorder.

According to a more recent dimensional perspective, Berlin & Hollander (2014) described compulsivity and impulsivity from the perspective of the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (*DSM-5*). As described also in Figure 2.1, the dimensional perspective included different disorders, ranging from obsessive-compulsive disorder and repetitive self-injurious behaviors (e.g., skin picking, trichotillomania), to substance use and behavioral addictions (Starcevic, 2016).

Impulsivity has been linked to many types of behaviors, whether they are healthy or not, with direct impact on daily functioning (Sharma et al, 2014). For example, sensation seeking, a facet of many models of impulsivity (e.g., Whiteside & Lynam, 2001), has been linked to skydiving (Myrseth et al, 2012) or rock climbing (Llewellyn & Sanchez, 2008). Moreover, individuals high in impulsivity are likewise non-planful (Ottaviani & Vandone, 2011; Whiteside & Lynam, 2001; Youzhi & Jing, 2009). According to Bechara (2005), and Whiteside and Lynam (2001) impulsivity, in its high level, can be associated with mental health problems. Other activities that are defined as impulsive include bungee jumping od skydiving (Whiteside & Lynam, 2001). However, although this type of sports requires careful planning, it tends to be classified among other risky, spontaneous, unplanned, and potentially dangerous behaviors, if performed without the right attention and without thinking to the possible consequences.

A different explanation of impulsivity can be made by a comparison between behavioral and cognitive impulsivity (see also, Bari & Robbins, 2013). These two definitions of impulsivity are also called "motor" and "choice" impulsivity, respectively. The first one refers to the difficulty to immediately conclusion of an obvious behavior (e.g., acting on impulse). On the other hand, choice, or cognitive impulsivity, represents a difficulty to ponder the consequences of specific actions, future events, and delay gratification (e.g., pathological gambler).

But, even if anyone can easily identify examples of impulsive behavior, there is considerably more difficulty in defining impulsivity precisely. For this reason, a broad, multidimensional construct of impulsivity might better explain research findings. A number of definitions of impulsivity and related concepts are listed in Table 2.1.

Table 2.1. Various definition of impulsivity

Author	**Definition**
APA[1]	Describing or displaying behavior characterized by little or no forethought, reflection, or consideration of the consequences of an action, particularly one that involves taking risks.
DSM-5[2]	Rushed actions that occur instantly, without premeditation, and which have a high potential for harm to the individual (e.g., crossing the street without looking). It may reflect a desire for immediate reward or an inability to delay gratification.
PID-5[3]	Acting on the spur of the moment in response to immediate stimuli; acting on a momentary basis without a plan or consideration of outcomes; difficulty establishing and following plans; a sense of urgency and self-harming behavior under emotional distress.
UPPS-P[4]	Consist of five distinct but related constructs: sensation seeking, lack of premeditation, lack of persistence, positive urgency, and negative urgency.
BIS-11[5]	Consist of three distinct but related traits: motor impulsivity, attention impulsivity, and non-planning impulsivity.
IMPSS[6]	It refers to lack of planning and the tendency to act impulsively without thinking and the seeking of excitement, novel experiences, and the willingness to take risks for these types of experiences.

[1]: *American Psychological Association (APA) Dictionary of Psychology, online version*
[2]: *Diagnostic and Statistical Manual of Mental Disorders Fifth Edition (DSM-5; APA, 2013)*
[3]: *Personality Inventory for DSM-5 (Krueger et al, 2012)*
[4]: *UPPS-P Impulsive Behavior Scale (Lynam et al, 2006)*
[5]: *Barratt Impulsiveness Scale-11 (Patton et al, 1995)*
[6]: *Zuckerman-Kuhlman Personality Questionnaire Impulsive Unsocialized Sensation Seeking Scale (Zuckerman et al, 1991)*

As Evenden (1999) pointed out in his review, there are not one unitary definition for "impulsivity" and there are also numerous types of impulsive behavior. It is more accurate to say that impulsivity is a term that refers to several related phenomena, which are usually classified together as impulsivity; these phenomena lead to different forms of impulsive behavior. This view is supported by several psychological studies on human personality traits, which show that impulsivity is made up of several factors' interactions coupled with qualitatively different behavior aspects (e.g., Sharma et al, 2014). Some of the facets of existent impulsivity models are decision making and risk-taking. In the next two paragraphs (Paragraph 2.1.1 and 2.2.2 of this Section) a review of decision-making and risk-taking theories will be presented.

2.1.1 Decision Making

Several studies showed that impulsivity has been studied in the context of decision-making (e.g., Kieres et al, 2004; Winstanley et al, 2004). In fact, the capacity to plan and/or control its impulsive behavior and responses is the principal characteristic for the ability to make a decision. Choosing between options is part of the decision-making process. It is possible to make a choice, or a decision, by deliberately calculating of risks and benefits ("cold" decision making), by using (at least in part) emotional reactions and gut instincts about each option ("hot" decision making; Buelow et al, 2015), or by combining these two factors (Buelow et al, 2015).

A well-known laboratory-based simulated gambling paradigm developed by Bechara and colleagues (i.e., The Iowa Gambling Task; 1994) has experienced great interest in its use in several fields (Busemeyer & Stout, 2022). In fact, this task has been used to experimentally study deficits in decision making exhibited by populations with brain damage, psychopathology, antisocial personality, or drug abuse problems (Busemeyer & Stout, 2022).

2.1.2 Risk-Taking

Impulsive behaviors are also influenced by the presence of some personality traits (Eysenck & Eysenck, 1985). Among these being extroverted and risk-oriented are the most studied (Chico, 2000; Eysenck & Eysenck, 1985). Also in developmental studies, impulsivity was positively correlated to risk-taking in children (Levin & Hart, 2003; García et al, 2004).

Risk-taking has been defined in various different ways, since it can include the engagement in a wide behavior that usually individuals defined as "risky". As mentioned by Boyer (2006) risk-taking can be viewed as the voluntary participation in any behavior, that can lead to negative consequences (see also Beyth-Marom & Fischhoff, 1997; Beyth-Marom et al, 1993; Byrnes, 1998; Furby & Beyth-Marom, 1992; Irwin, 1993). Among the risk-taking behaviors, there are behaviors that are both illegal and risky for one's health (Magar et al, 2008). For example, frequently risky behaviors could be alcohol and tobacco consumption, dangerous driving, unsafe sexual activity, etc. A substantial literature identified the emerging and increasing of a large amount of risk-taking behavior in adolescence (Arnett, 1992; Arnett, 1999; Donovan & Jessor, 1985; Donovan et al, 1988; Gottfredson & Hirschi, 1990; Gullone et al, 2000; Jessor, 1991; Laird et al, 2003; Moffitt, 1993; Rai et al, 2003).

Lastly, numerous studies have been carried out to understand why people are more inclined to participate in risky behaviors and which are the individual differences in personality and risk-taking propensity (Lauriola & Levin, 2001; Nicholson et al, 2005; Schwebel et al, 2006). These mentioned studies have found association between risky behaviors and high level of extraversion and openness (ibidem).

2.2 Decision Making and Risk-Taking Tasks

In order to evaluate decision making and risk taking in laboratory setting, in this project two different tasks were used: the Balloon Analogue Risk Task (Lejuez et al, 2002) and the Iowa Gambling Task (Bechara et al, 1994). These two tasks were administered with the Psychology Experiment Building Language platform (PEBL; Muller, 2013; Mueller & Piper, 2014; an exhaustive description of the platform is presented in Chapter 2.2.1 of Section I).

2.2.1 Balloon Analogue Risk Task

The Balloon Analogue Risk Task (BART; Lejuez et al, 2002) was designed to provide a context in which researchers could examine actual risky behavior. During this task, the reward of participants' behavior (i.e., by gaining money) measures propensity toward risk. The rewarding phase lasts for few trials, after which continued engagement in that behavior leads to poorer outcomes (i.e., the loss of the collected money; Lejuez et al, 2002).

Throughout the task, the computer screen showed a small, simulated balloon accompanied by a balloon pump, which says "Press this button to pump up the balloon". In the first presentation of the different components of the task, participants will see a permanent money-earned display labeled "Total Earned", and a second display listing the money earned on the last balloon and labeled "Last Balloon" (Lejuez et al, 2002), as shown in Figure 2.2.

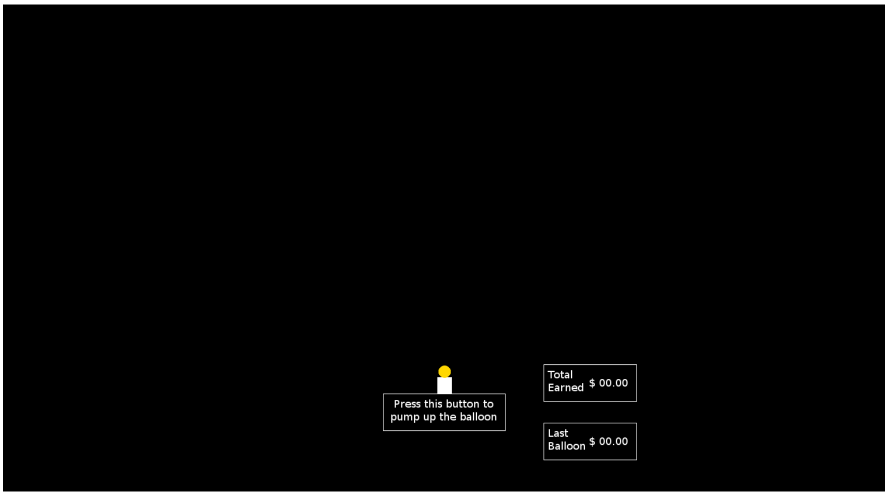


Figure 2.2. Presentation screen of the BART.

After the first pump on the balloon, a reset button labeled "Press to Collect $$$" will appear, to allow participants to collect the money earnt with the present balloon (see Figure 2.3).

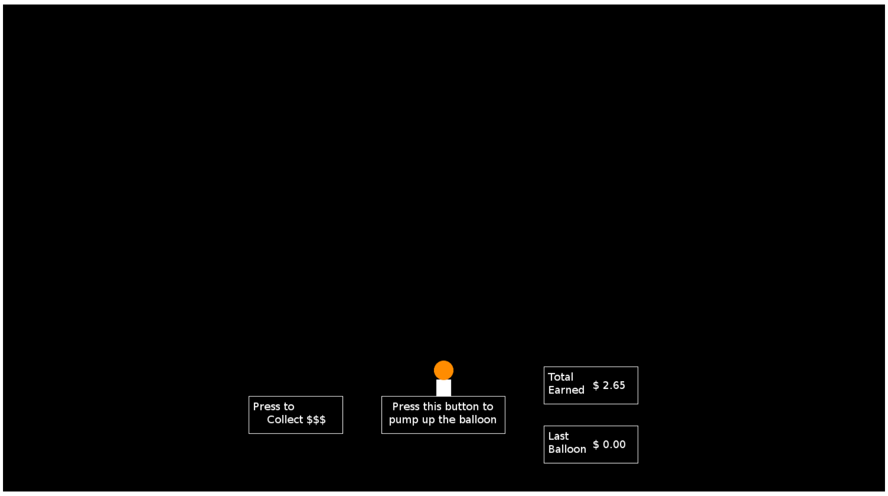


Figure 2.3. Complete presentation screen of the BART with all the buttons.

With each pump (i.e., click on the button), 5 digital cents were accumulated in a temporary reserve. Participants do not see the amount of money in the temporary reserve. At any point during each balloon trial, the participant could stop pumping the balloon and click the "Collect $$$" button. By clicking this button, the task transfers all the money from the temporary reserve to the permanent reserve (i.e., "Total Earned"). This new reserve will increase cent by cent with each pump.

When a balloon explodes, all money in the temporary reserve is lost, and the next uninflated balloon appears on the screen. The participant's exposure to a balloon end, after each balloon explosion or money collection, and a new balloon appears until the total of balloons (i.e., trials) has been completed.

PEBL software allows researchers to choose between two different versions of the task: 30 trials or 90 trials. Both versions comprised 3 different balloon types (i.e., orange, yellow, and blue). Each balloon color had a different probability of exploding (Lejuez et al, 2002). Specifically, the balloon that burst most easily are the orange ones. Participants receive no detailed information about the probability of an explosion, and they are not informed that different balloon colors had different probabilities of exploding (Lejuez et al, 2002). Instructions presented at the beginning of the administration, only say that at some point each balloon would explode, and this can happen at the first pump or when the balloon had expanded to fill the entire computer screen (Lejuez et al, 2002). As the blue balloon allowed the widest range of possible number of pumps (see for example, Figure 2.4), the number of pumps on this balloon can be considered as a primary dependent measure (Lejuez et al, 2002), since they can capture the greatest amount of individual variability and risk-taking behaviors.

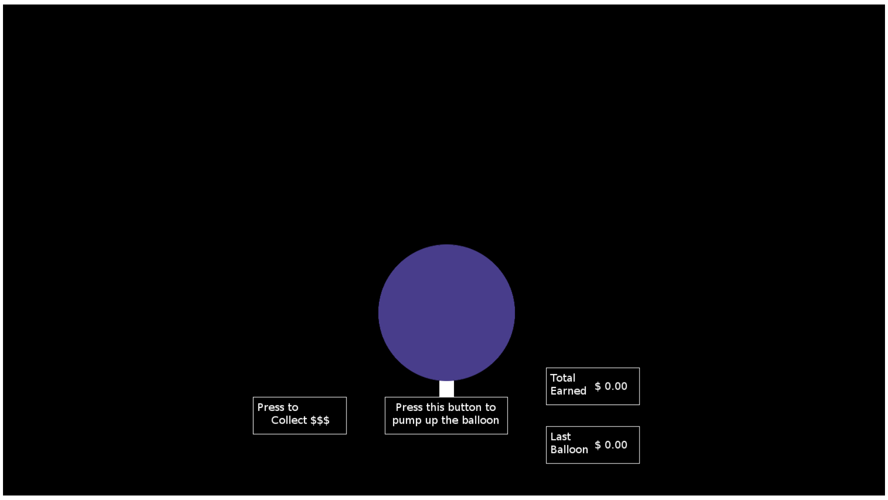


Figure 2.4. A possible number of pumps with the blue balloon.

Riskier performance on this task has been shown in individuals endorsing substance abuse and general risk-taking behaviors (e.g., Aklin et al, 2005; Fernie et al, 2010; Hopko et al, 2006; Lejuez et al, 2002; Lejuez et al, 2003a; Lejuez et al, 2003b), indicating the measure's ecological validity. The BART has shown to correlate strongly with self-reports of behavioral risk-taking and also to distinguish between smokers and non-smokers (Lejuez et al, 2002; Lejuez et al, 2003a).

2.2.2 Iowa Gambling Task

As previously mentioned in paragraph 2.1.1 (of the present Section), the Iowa Gambling Task (IGT) has been commonly used to assess decision making. The Figure below show the wide fields where the Iowa Gambling Task is used.

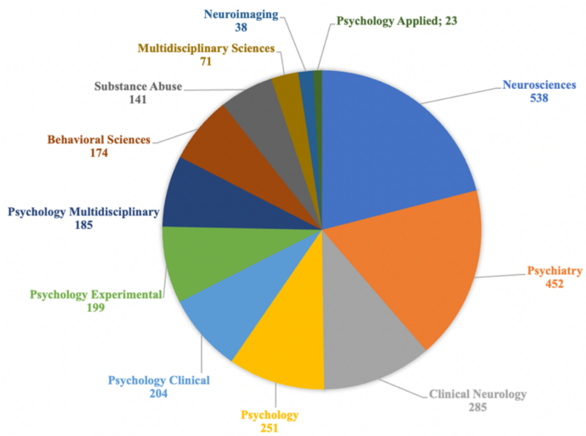


Figure 2.5. The number of stop-signal publications per area
Source: Web of Science, 14/09/2022. Search term: "topic = Iowa gambling task".

The Iowa Gambling Task (IGT; Bechara et al, 1994) was designed as a clinical assessment instrument to measure deficits in decision making among clinical populations; particularly, this task assess the notion of outcome impulsivity (see paragraph 2.1 of the present Section; Buelow & Suhr, 2009). Designed to mimic the decision made in real life, this type of behavioral task is useful at assessing participants' ability to make decision under uncertain condition and their sensitivity to potential gains and losses in the environment.

The Iowa Gambling Task is a decision-making task in which participants are asked to organize a sequence of actions towards a goal (i.e., maximizing the starting amount of money) and inhibit response sequences (i.e., by selecting card from advantageous decks) to achieve the initial goal to maximize the amount of money.

Executive Function can be defined in a variety of ways; however, one widely accepted definition refers to organizing a sequence of actions toward a goal. Another different definition refers to activating and inhibiting response sequences guided by internal neural representations. Thus, the IGT is a good measure of executive functions. Furthermore, this task is widely used in neuroscience to measure impairment in decision-making. Bechara and colleagues (2000) suggested that the Iowa Gambling Task measures cognitive impulsivity, such as the inability to delay gratification and evaluate the outcome of a planned action (see paragraph 2.1 of the present Section). A consensus regarding what type of decision-making is assessed with the IGT has not yet been reached, limiting its clinical utility. Moreover, decisions made during this type of task are based on an interplay of several basic processes (e.g., cognitive and/or motivational processes; Busemeyer & Stout, 2002). Thus, behavioral deficits cannot be attributed to any specific process (e.g., choice mechanisms), since these decisions are the result of complex cognitive-motivational interactions (Moutoussis et al, 2021). For this reason, existent literature should continue to grow, to increase the knowledge on this task.

The Iowa Gambling Task involves four decks of cards: A, B, C and D (Figure 2.6).

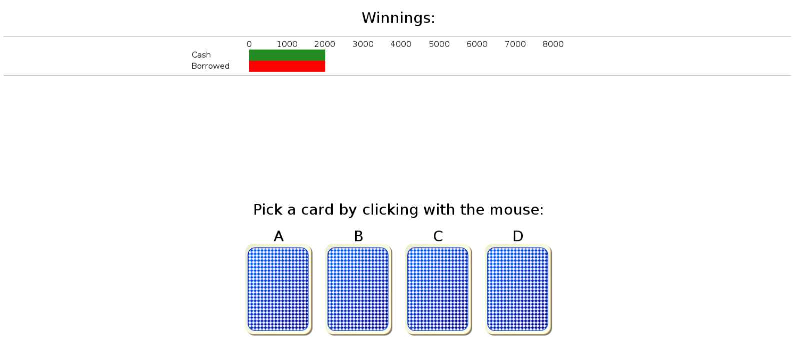


Figure 2.6. Presentation of the four decks in the Iowa Gambling Task

Using the mouse, subjects must choose one card at a time from one of the four decks. The selection of a card in two decks (i.e., A and B), is accompanied by a high gain of play money (Figure 2.7 - Upper part), but at unpredictable points it is accompanied by a high penalty (Figure 2.7 - Down part).

Figure 2.7. Gains and losses for the Deck A and Deck B
Upper part: the selection of these two decks is followed by a high gain of money
Down part: the selection of these two decks is also followed by a high penalty

For the other two decks (i.e., C and D), the immediate gain is smaller (Figure 2.8 - Upper part), but the future loss is also smaller (Figure 2.8 - Down part).

Figure 2.8. Gains and losses for the Deck C and Deck D
Upper part: the selection of these two decks is followed by a small gain of money
Down part: the selection of these two decks is also followed by a small penalty

The goal of the IGT, is to maximize their long-term returns. Successful completion of the IGT requires participants to learn that two of the decks will be disadvantageous in the long term (i.e., A and B; high immediate returns but long-term losses) while the remaining two decks will be advantageous (i.e., C and D; low immediate win amounts but long-term gains). As shown in Figure 2.9, at the end of the task, participants received feedback for their performance in the task.

You've completed the game. After repaying your loan, your net winnings are $1150.

Thank you for participating in the experiment. Please notify the researcher that you have finished.

Figure 2.9. Conclusive feedback of the IGT

As shown in Table 2.2 the gain-loss frequency is balanced, with 14 gains and 6 losses between the bad (A and B) and the good decks (C and D) in the first ten trials. This Figure represents only the first 40 trials, that will be repeated until the maximum number of trials for the task (i.e., 100 trials). In this project I relied on a 100 trials version of the Iowa Gambling Task, with five consecutive blocks of 20 trials each.

Table 2.2. Gain-loss frequency for the first 40 trials on the IGT

Trials	**Deck A**	**Deck B**	**Deck C**	**Deck D**
1	100	100	50	50
2	100	100	50	50
3	100, **-150**	100	50, **-50**	50
4	100	100	50	50
5	100, **-300**	100	50, **-50**	50
6	100	100	50	50
7	100, **-300**	100	50, **-50**	50
8	100	100	50	50
9	100, **-250**	100, **-1250**	50, **-50**	50
10	100, **-350**	100	50, **-50**	50, **-300**
11	100	100	50	50
12	100, **-350**	100	50, **-25**	50
13	100	100	50, **-75**	50
14	100, **-250**	100, **-1250**	50	50
15	100, **-200**	100	50	50
16	100	100	50	50
17	100, **-300**	100	50, **-25**	50
18	100, **-150**	100	50, **-75**	50
19	100	100	50	50
20	100	100	50, **-50**	50, **-250**
21	100	100, **-1250**	50	50
22	100, **-300**	100	50	50
23	100	100	50	50
24	100, **-350**	100	50, **-50**	50
25	100	100	50, **-25**	50
26	100, **-200**	100	50, **-50**	50
27	100, **-250**	100	50	50
28	100, **-150**	100	50	50
29	100	100	50, **-75**	50, **-250**
30	100	100	50, **-50**	50
31	100, **-350**	100	50	50
32	100, **-200**	100, **-1250**	50	50
33	100, **-250**	100	50	50
34	100	100	50, **-25**	50
35	100	100	50, **-25**	50, **-250**
36	100	100	50	50
37	100, **-150**	100	50, **-75**	50
38	100, **-300**	100	50	50
39	100	100	50, **-50**	50
40	100	100	50, **-75**	50
Total (100 trials)	-2500	-2500	+2500	+2500

Note. Bold numbers represent the losses.

Subjects with non-risky behavior, avoid the decks with high an immediate gain, which differs from clinical populations, such as pathological gambler, obsessive–compulsive patients, and people with psychopathic tendencies, or schizophrenia (de Visser et al, 2011). For this reason, in the context of the Iowa Gambling Task, outcome impulsivity is defined as the tendencies to focus more on overall outcomes, rather than winning and losing frequencies (Buelow & Suhr, 2009).

Three assumptions about healthy participants' performance on the IGT are based on Steingroever and colleagues' study (2013).

1. According to the first assumption, healthy participants learn to prefer the good options over the bad ones, since they understand that these decks produce positive long-term effects, despite rather small immediate rewards (Bechara et al, 1994).
2. The second assumption stated that a healthy participant's choice performance is implied to be homogeneous (Steingroever et al, 2013).
3. The last assumption, it is explicitly assumed that healthy participants explore different options (i.e., explorative phase), before settling down the most profitable ones (i.e., exploitation phase) (Steingroever et al, 2013).

Figure 2.10 shows deck selection of participant LL627 (female, 25 years old): in this figure is evident the preference of the subject for the good decks (more specifically for Deck C) over the bad ones. The preference was settled down after exploring all the options, showing, afterwards, a homogeneous choice for Deck C.

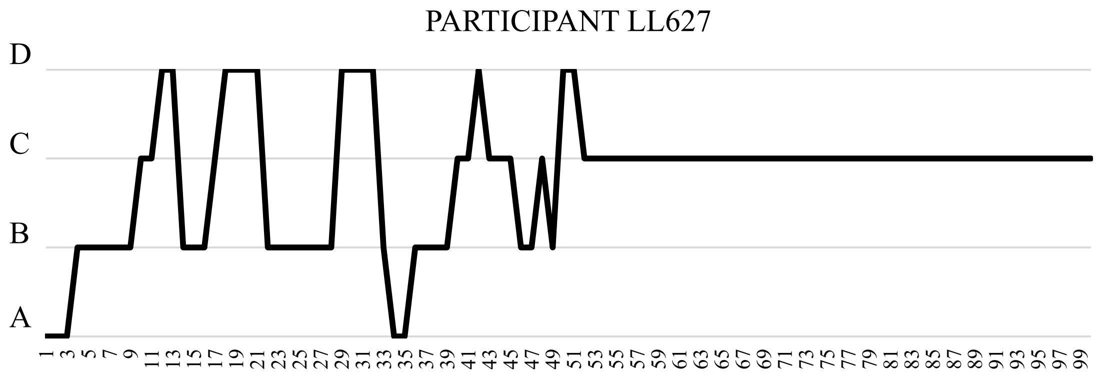


Figure 2.10. Decks selection over the administration on the Iowa Gambling Task, supporting the three assumptions.
Note. For a better presentation of the data, the axis shows trials with interval of two.

To support their assumptions about the characteristic choice behavior of healthy participants across trials, Bechara and colleagues (1994) presented deck selection profiles of two typical control participants. These profiles illustrate that these control participants "*initially sampled all decks..., but eventually switched to more and more selections from the good Decks C and D, with only occasional returns to decks A and B*" (Bechara et al, 1994; p. 12; in Steingroever et al, 2013).

Nevertheless, the results of Steingroever and colleagues' study (2013) showed that all three assumptions may not be true. For instance, they demonstrated that healthy participants tend to prefer decks with few losses (first assumption; Figure 2.11a); healthy participants exhibit idiosyncratic choice behaviors (second assumption; Figure 2.11b); and there is no systematic decrease in the number of switches across trials in healthy participants (third assumption: Figure 2.11c).

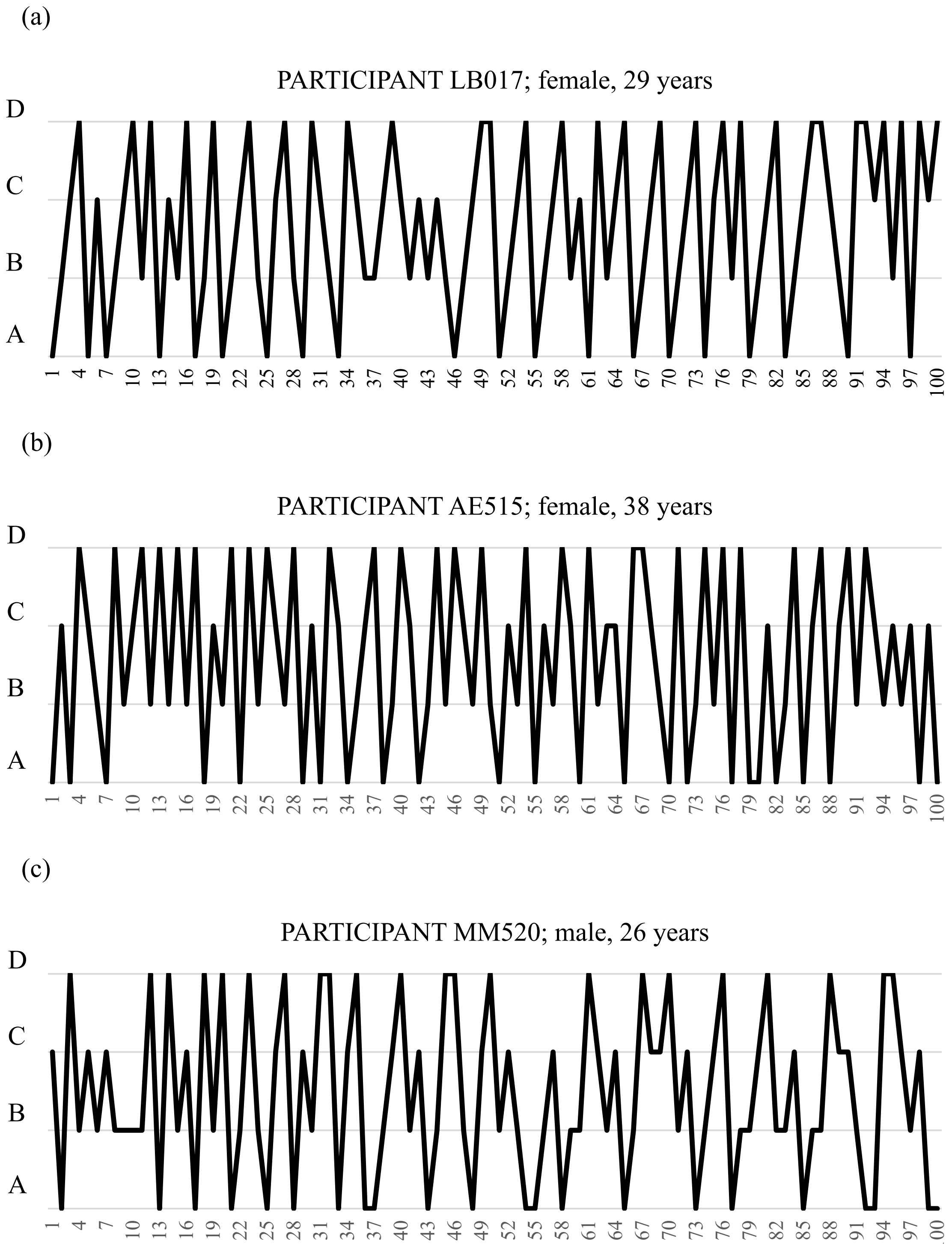


Figure 2.11. Decks selection of three participants over the administration on the Iowa Gambling Task, that invalid Bechara and colleagues' assumption (1994).

In the field of clinical research, multiple, independent studies showed an association between IGT scores, and substance use relapse (De Wilde et al, 2013a; De Wilde et al, 2013; Goudriaan et al, 2011; Kasar et al, 2010; Nejtek et al, 2013; Passetti et al, 2008; Radat et al, 2013; Salgado et al, 2009; Wang et al, 2013). Moreover, the results of Kjome and colleagues' study (2010) showed that patients with cocaine addiction made more disadvantageous choice on the IGT than controls. In several control studies with schizophrenic participants (e.g., Kester et al, 2006; Shurman et al, 2005), patients' performance was worse than control for IGT net score (see also Beninger et al, 2003; Ritter et al, 2004). However, other studies did not find any differences between patients and normal controls on IGT performance (e.g., Cavallaro et al, 2003; Evans et al, 2005).

IGT performance can be identified in two simple, and standard ways: (1) by examining the overall net return after a specified number of trials, or (2) by comparing the frequency of choices between advantageous and disadvantageous decks in blocks throughout the administration of the task (Brown et al, 2012; Lamers et al, 2006; Poletti et al, 2011; Stout et al, 2004).

However, other studies have demonstrated that these standard measures have questionable validity (e.g., Buelow & Suhr, 2009; Lin et al, 2013; Steingroever et al, 2013). More pertinently, these mentioned studies importantly noted that both net return and frequency of deck choice are composite of multiple decision-making processes, making it difficult to claim that poor performance is a result of impulsive behavior alone. For instance, a participant with lack in attention span may be disadvantaged on the IGT, since, in order to make the best choices in the future, it has to remember multiple outcomes over time.

In the next Chapter (Chapter 3 of this Section) several methods to estimate performance on IGT will be presented, with their strengths and weaknesses.

3. Frequentist estimations of IGT

According to different studies of Bechara and colleagues (Bechara et al, 1994; Bechara et al, 1997), participants who face are sensitive to the long-term consequences. As demonstrated in their work (Bechara et al, 1994), the gradual shift from A and B to C and D decks, produces long-term benefits for decision makers.

In this Chapter the most popular estimation methods of the Iowa Gambling Task are discussed, with their respective limits. These methods focus exclusively on obtaining summary and frequentist measures of the number of selections from the decks. The methods described in this paragraph are: (a) the number of selections from each deck, (b) the number of advantageous (C + D) and number of disadvantageous (A + B) selections separately, that allows researchers to analyze the difference between long-term losses (i.e., disadvantageous decks A + B) compared with long-term gains (i.e., advantageous decks C + D), and (c) the net score across all the trials.

For a better description and representation of the methods, I will present the data of some participants of the present project.

3.1 Number of selections from each deck

The first estimation method of the Iowa Gambling Task is the number of selections from each deck. As discussed before (see paragraph 2.2.2 of the present Section), community dwelling participants usually prefer the advantageous decks (i.e., C and D) over the disadvantageous ones. As shown in the Figure 3.1, this participant (SP718, female, 24 years old) have understood the difference between the four decks and preferred the advantageous one. In this case, the estimations for the subject are easily deduced from the output provided by PEBL platform and are A = 7; B = 18; C = 25; D = 50.

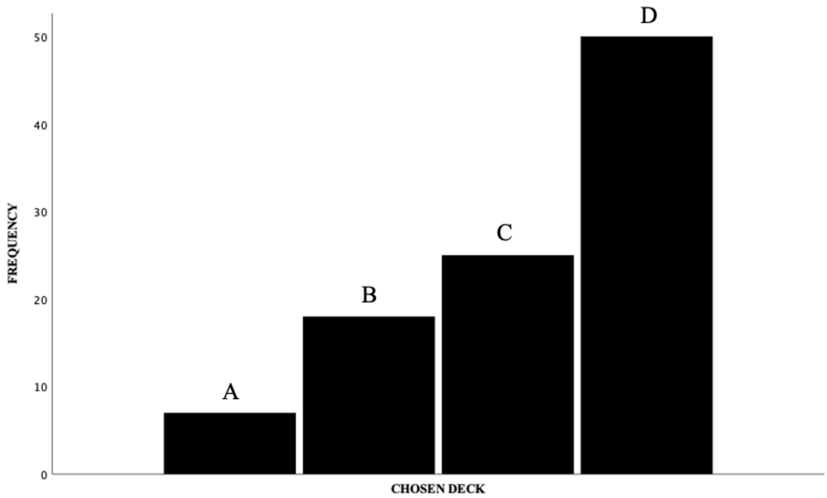


Figure 3.1. Number of selections from each deck, participants SP718

The overall preference for the good options and the exploration of the participant before choosing the two advantageous desks is better visualized in Figure 3.2. This Figure represents the five blocks of the administration of the task, with 20 trials each. Participant SP718 used the firsts 40 trials to explore the four decks (Figure 3.2a and Figure 3.2b). From the 41st trial, instead, participant has understood the difference between advantageous and disadvantageous decks (Figure 3.2c). Finally, in block 4 and 5 (Figure 3.2d and Figure 3.2e, respectively), this participant has begun to avoid the selection of the bad decks.

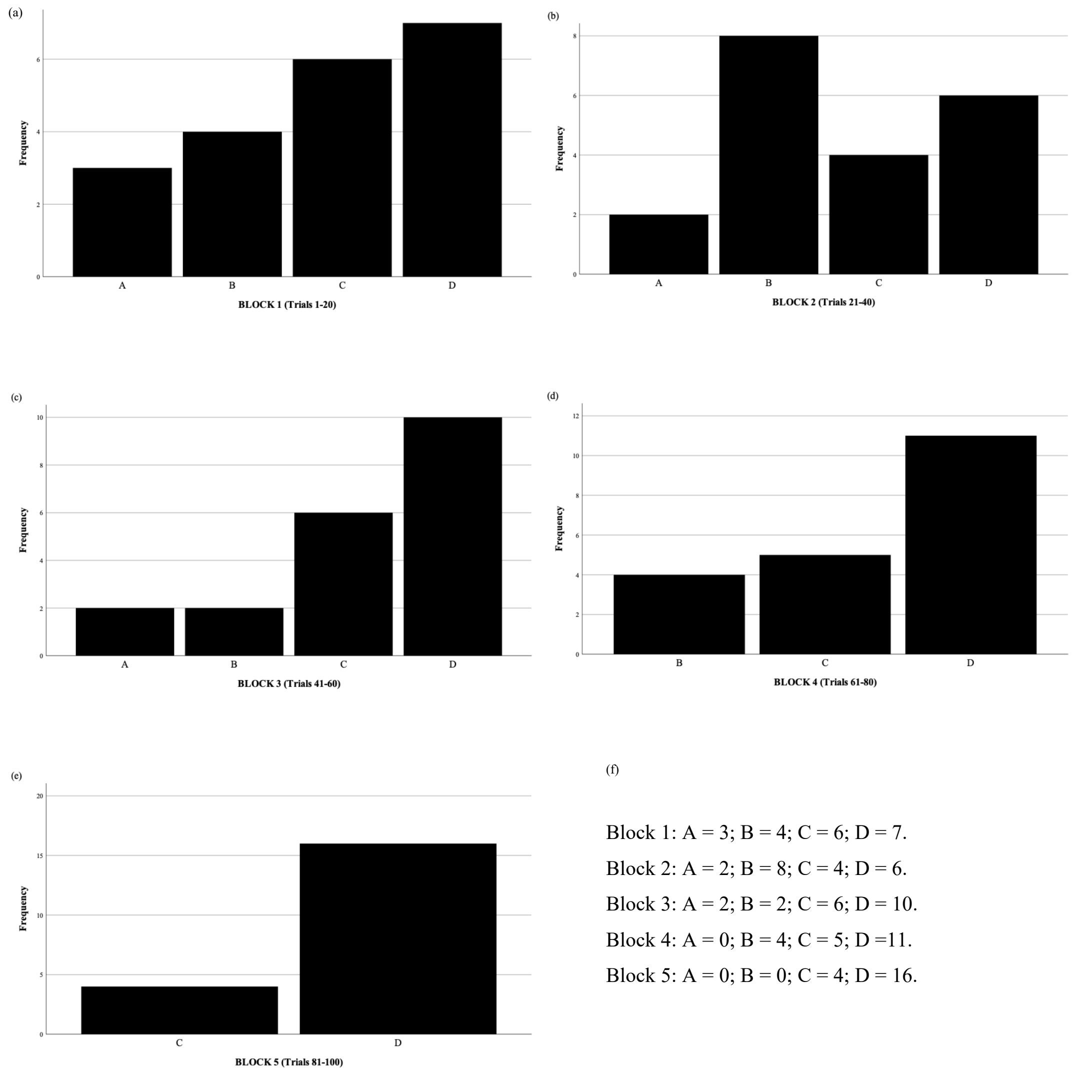


Figure 3.2. Changing of the preference decks across trials and number of selections
Note. (a): trials 1-20; (b): trials 21-40; (c): trials 41-60; (d): trials 61-80; (e): trials 81-100; (f): number of selections in each block.

However, in the previous paragraph (i.e., Paragraph 2.2.2 of the present Section) it was suggested that Bechara's assumptions are not always correct (Steingroever et al, 2013), and participants' preferences within the advantageous and disadvantageous decks on the IGT are not always uniform.

In fact, also in the participants of this project, and as shown in Figure 3.3, participant MG721 (female, 40 years old) showed a preference for the two decks with infrequent losses (decks B and D) regardless of the long-term value of selecting from these decks. This conclusion is also supported by Steingroever and colleagues' findings (2013), that reviewed 17 studies and, out of these 17 studies, 13 preferred Deck B (e.g., Fridberg et al, 2010; Rodríguez-Sánchez et al, 2005). In this case, the frequencies were the opposite as previous participant: A = 8; B = 44; C = 7; D = 41.

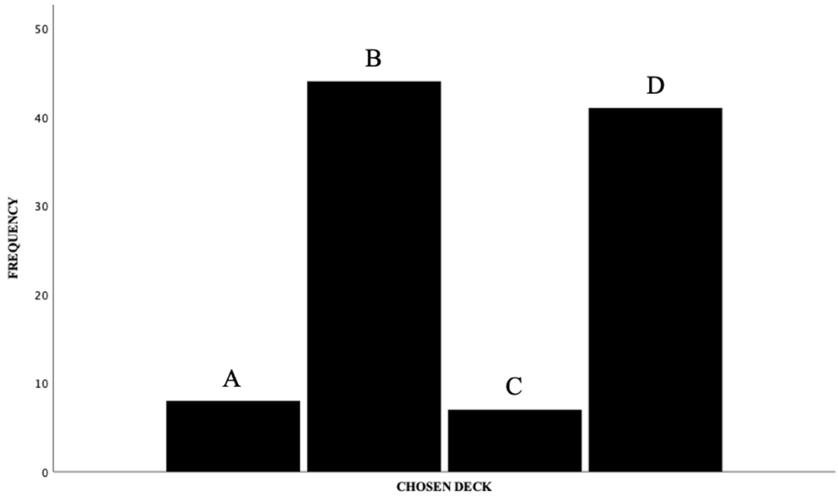


Figure 3.3. Number of selections from each deck, participants MG721

3.2 Number of advantageous and number of disadvantageous selections

A different way to estimate the selections of the deck is to consider the number of advantageous decks (C + D) and the number of disadvantageous decks (A + B) separately. This comparison allows researchers to understand if participants has understood the difference between advantageous and disadvantageous decks.

In the following Figure, this participant (AA517, male, 34 years old) preferred the two decks that allow him to maximize the long-term gains (A + B = 15; C+ D = 85) over the others two decks. A different way to mention these pairs of decks is consider the advantageous decks as the long-term gains, and the disadvantageous decks as the long-term losses. In this case, participant AA517 preferred and maximize long-term gains.

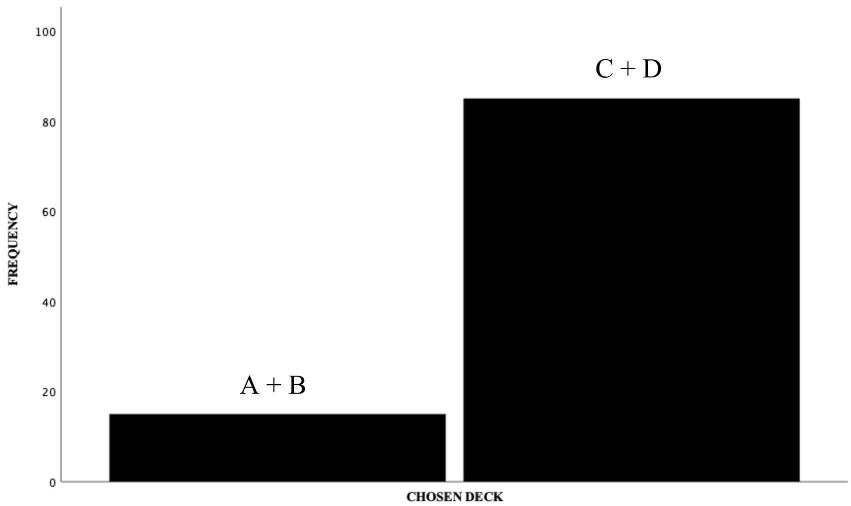


Figure 3.4. Number of selections from good and bad decks, participants AA517

However, when considering the number of selections for advantageous and disadvantageous decks, more specific information about the choices that participants has made throughout the 100 trials, can be lost. As Steingroever and colleagues (2013) has shown, this type of estimation can hide the frequency of gains and losses throughout the administration, which can give specific information on the participant's behavior.

3.3 Subtracting disadvantageous from advantageous selection

The third estimation method is the standard measure of performance on the IGT and allows researchers to estimate the net score of the Iowa Gambling Task. Specifically, net score is calculated by subtracting the number of cards selected from decks A and B (i.e., the disadvantageous decks) from the number selected from the advantageous decks (i.e., C and D), resulting [(C + D) - (A + B)]. If the net score is positive, participants preferred the advantageous decks, whereas a negative net scores indicate a preference for the disadvantageous ones.

To describe this estimation, participant's result will be presented. Participant AG715 (male, 47 years old) showed a preference of 33 choices for the disadvantageous decks, and 67 choices for the advantageous decks. So, in this case the net score is + 34.00, showing a clear preference for the advantageous decks.

Additionally, net score can be calculated across the five blocks of twenty trials for each participant. This allows researchers to understand if there has been a learning rate and to calculate it, by looking at the change in participants' net scores across blocks. For the example presented, participant AG715 showed an increasing in the net score between the five blocks, as shown in Figure 3.5, starting from a negative net score (- 6.00), up to a positive net score (+ 20.00). This example showed that participants can differentiate the good decks from the bad decks, to maximize the gains, after exploring the different options.

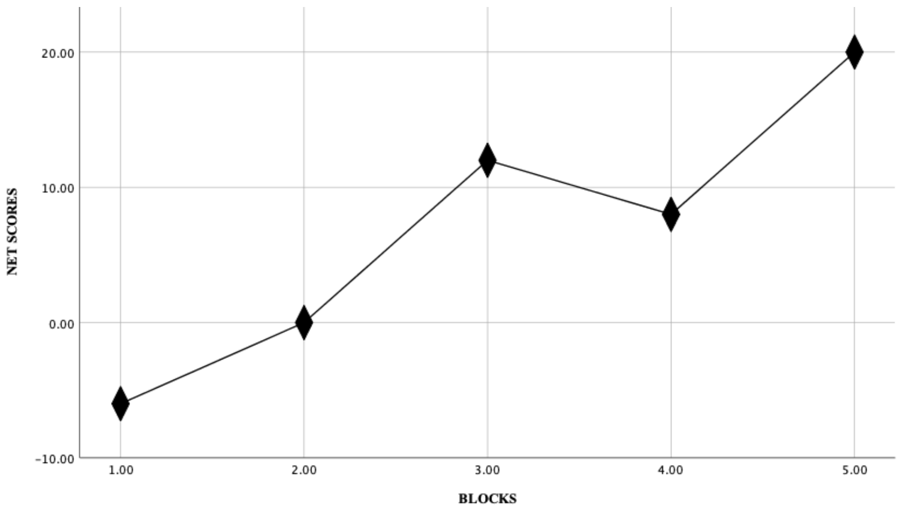


Figure 3.5. Net score across five blocks, participants AG715

However, this assumption is not applicable for all the participants, as shown for participant TR724 (male, 41 years old) in Figure 3.6.

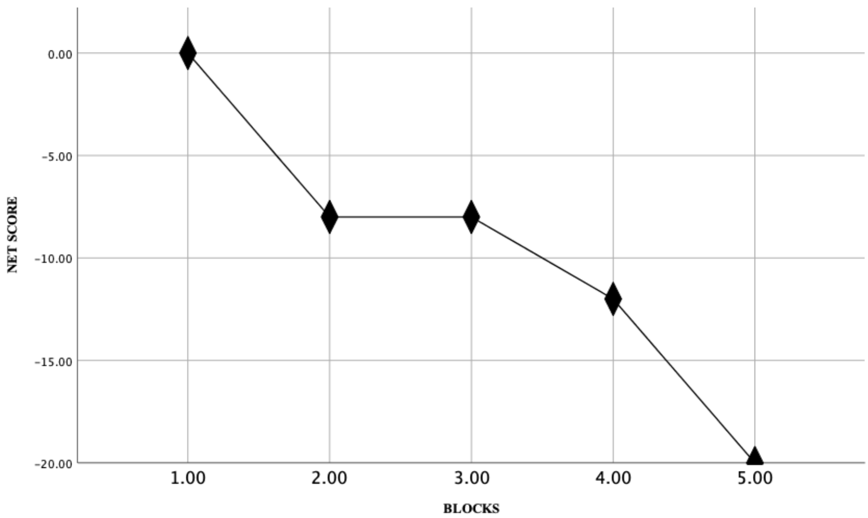


Figure 3.6. Net score across five blocks, participants TR724

Though, a limitation concerns the fact that if researchers consider the mean net score in the entire sample, results are different. As shown in Figure 3.7, although the presence of opposite scores (i.e., participant AG715 and TR724), the overall trend of net scores improves during the administrations.

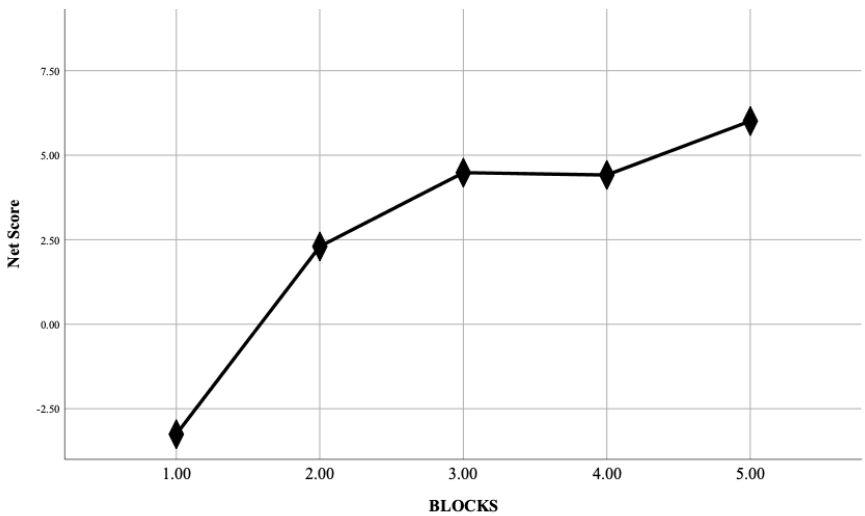


Figure 3.7. Mean net score across 20-trials blocks for each session in the entire sample

3.4 Conclusion

Participants' performance cannot be analyzed only on the basis of the mean number of cards selected in each session. The consequences of these types of estimations can be the loss of lots of information about participants performance. As mentioned before, also MacPherson and colleagues (2002) and Wilder and colleagues (1998), for example, sowed that more cards were selected from decks B and D in their analysis. However, as shown from the participants of this project performance can vary between participants.

Thus, in a session, the number of cards selected from any individual deck is not indicative of the success of the session. For this reason, there are a lot of studies (e.g., Busemeyer & Stout, 2002; Moutoussis et al, 2021) that stated that there is the possibility that decisions made during the Iowa Gambling Task may be influenced by complex cognitive-motivational interactions, which make it difficult to identify the specific processes responsible for the observed behavioral deficits.

Therefore, to investigate psychological processes that may be underlie participants performance, researchers (e.g., Ahn et al, 2008; Busemeyer & Stout, 2002; Busemeyer et al, 2003; Steingroever et al, 2018; Yechiam et al, 2008) prefer using Reinforcement-Rearning (RL) models, that assumes that card selection on the IGT results from an interaction between psychological processes including motivation, memory, and response consistency (Steingroever et al, 2017).

In the next paragraphs (i.e., Paragraph 4, 4.1, and 4.2 of this Section) the most frequently used models will be briefly discussed.

4. Reinforcement-Learning Models of IGT

As I already anticipated in the previous paragraphs, poor performance on the Iowa Gambling Task has been attributed to a failure to anticipate the long-term negative consequences of disadvantageous choices. Using only mean number of cards selected cannot identify the specific processes responsible for the observed behavioral deficits caused by IGT decisions, which are the result of complex cognitive-motivational interactions (see Paragraph 3.4 of this Section; Busemeyer & Stout, 2002; Moutoussis et al, 2021). To identify and measure the underlying psychological processes driving observed performance in complex tasks, Reinforcement-Learning (RL) models provide a promising additional analysis approach (Steingreover et al, 2013a; Steingreover et al, 2018). By using these models, it is possible to gain a deeper understanding of behaviors resulting from a combination of different cognitive and motivational processes.

Two of the most frequently used models include the Expectancy Valence model (EV; Busemeyer & Stout, 2002), and the Prospect Valence Learning model (PVL; Ahn et al, 2008). The most important aspect of these models is that psychological processes such as motivation, learning/memory, and response consistency are reflected in these models (Busemeyer et al, 2003; Steingroever et al, 2013a).

A wide range of methods have been used to test and compare RL models for the IGT, focusing on different aspects (for a review see Steingroever et al, 2013a). In addition, several studies have carefully investigated the ability of these models to recover the data-generating parameters (Ahn et al, 2011; Ahn et al, 2014; Steingroever et al, 2013b; Wetzels et al, 2010).

In this paragraph, the most popular RL models for the IGT will be discussed: EV and PVL (Steingroever et al, 2013a).

The idea underlying these cognitive models is that participants evaluate the rewards and the losses by considering the chosen card, through the application of a utility function after each choice (Steingroever et al, 2013a). The result of this function is used to update expectancies about the following decks. Thus, on every trial, participants adjust their expectations of the decks based on the new experience.

In the next two paragraphs, the Expectancy Valence model (EV; Paragraph 4.1 of this Section) and the Prospect Valence Learning model (PVL; Paragraph 4.2 of this Section) will be deeply presented, describing the specific parameters of each model.

4.1 Expectancy-Valence Learning Model

The Iowa Gambling Task was developed to assess the interplay between cognitive and motivational processes (Busemeyer & Stout, 2002). Based on the IGT theory (Bechara et al, 1994), poor performance on this task indicates a difficulty to differentiate advantageous and disadvantageous choices and a failure to anticipate the long-term consequences. When researchers tried to understand the exact causes of the decisions made during this task, however, problems arise, because identifying the specific processes underlying participants' behavior based only on frequentist estimations (see Paragraph 3 of this Section) it becomes difficult. For this reason, cognitive models provide a theoretical basis at identifying and measuring the underlying processes that occurs during this type of task.

For a better comprehension of the description of the model, a list of the definitions and notations is presented in Table 4.1.

Table 4.1. Definitions and Notations for the IGT Cognitive Models.

Notation	Definition
D	Deck
D_1 and D_2	Disadvantageous decks
D_3 and D_4	Advantageous decks
t	Trials (i.e., number of the chosen card)
$D(t)$[1]	Deck chosen on trial t
$\mathrm{R}[\mathrm{D}(t)]$	Reward on Deck D
$\mathrm{L}[\mathrm{D}(t)]$	Loss on Deck D
$\mathrm{u}_{\mathrm{D}}(t)$	Utility of deck D in trial t
v	Valences
$v(t)$	Valence of a trial t
w	Attention weight
E	Expectancy
a	Updating rate
Pr	Probabilistic choice
θ	Sensitivity
c	Consistency
ϕ	Shape
λ	Loss aversion

Note. 1: e.g., $D(t) = D_3$, the third advantageous deck was chosen on trial t.

The Cognitive Decision Model most widely used is the Expectancy-Valence Learning Model (Busemeyer & Stout, 2002). This model integrates learning and decision-making processes into a unified model (Busemeyer & Myung, 1992; Erev & Roth, 1998). After choosing a card from deck *D* on trial *t*, participants calculate a mean of the experienced rewards and losses. This process can be described as a utility of a specific deck *D* in trial *t*, $u_D(t)$. Consequently, according to this, participants combine the gains and the losses they experience on each trial of the task into an affective reaction called *valence*. By using an adaptive learning mechanism, expectancies about the valence produced by each deck are learned by the decision maker. These expectancies then are used as inputs to a probabilistic choice mechanism that selects the deck for each trial (Busemeyer & Stout, 2002).

Thus, Busemeyer and Stout (2002) identified three parameters to formalize the assumptions about Expectancy-Valence Learning Model to describe participants' performance on the IGT. For additional details about these three parameters see https://osf.io/tudmw/ (Gialdi, 2022).

1. *Valences (v).* As previously mentioned the gains and losses experienced after a selection of a deck produce the affective reaction called *valence*, *v*(*t*), represented by the following Equation:

$$v(t) = \{(1 - w \ \cdot R[D(t) + w \ \cdot \ L[D(t)]\} \quad (1)$$

   The parameter *w* represents the importance that participants assign to losses and to rewards, and it is named *attention weight parameter (*Steingroever et al, 2013a). This parameter, ranging from 0 to 1, allows participants to give different amounts of attention to gains and losses (Steingroever et al, 2013a; Worthy et al, 2012).

2. *Expectancy learning (E).* The decision maker, during the administration, create expectancies about the valences that are produced after each deck selection, *Ev*[*D*|*t*] (Busemeyer & Stout, 2002), as described by the following Equation

$$Ev[D|t] = (1 - a) \cdot Ev[D|t + 1] + a \cdot v(t) \qquad (2)$$

This model produces expectancies based on past valences, and the weight given to each past valence decreases as a function of the experience (Busemeyer & Stout, 2002). Passed experienced valences received less weight than recently experienced valences (Steingroever et al, 2013a). In this equation [Equation (2)], the parameter *a* represents the updating rate. As the previous one, also this parameter range is between 0 and 1, and quantifies the memory for rewards and losses (Steingroever et al, 2013a).

3. *Probabilistic choice (Pr).* Combining these parameters, the participants' choice made on each trial is a probabilistic function of the expectancies of each deck (Steingroever et al, 2013a):

$$\Pr[D|t + 1] = \frac{e^{Ev[D|t]\theta(t)}}{\sum e^{Ev[D|t]\theta(t)}} \qquad (3)$$

Based on this function, Steingroever and colleagues (2013) stated that the probability of choosing deck *D* is an increasing function of the expectancy for that deck and a decreasing function of the expectancies for the other decks (Steingroever et al, 2013a). The parameter $\theta(t)$ in Equation (3) is called the *sensitivity* parameter, and it determines the sensitivity of the choice probabilities to the expectancies (Busemeyer & Stout, 2002). The decision maker's sensitivity should change with the experience of each deck. For healthy individuals, for example, the sensitivity may initially start out at a low value, since choices are almost random while exploring the different options. During the administration, sensitivity should increase, based on past experience, and choices should be influenced by expectancies (Busemeyer & Stout, 2002). These changes can be explained by the following function:

$$\theta(t) = (t/10)^c \qquad (4)$$

To understand the changes of the participants, parameter *c* can have positive or negative values. If *c* is positive, there is an increase in the sensitivity and choices are less random and determined by the expectancy (Steingroever et al, 2013b; Steingroever et al, 2016; Worthy et al, 2012); on the contrary, negative values of *c* indicate random choices, maybe due to boredom or fatigue during the task (Steingroever et al, 2013b; Steingroever et al, 2016; Worthy et al, 2012).

In sum, the Expectancy–Valence model has three parameters: (a) the attention weight, *w*, that indicates the weight of rewards and losses over the administration; (b) the updating rate parameter, *a,* that represents the memory for past expectancies; and (c) the response consistency parameter *c*, that indicates the changes in the sensitivity over the performance and determines the balance between random and deterministic choices (Steingroever et al, 2013b; Steingroever et al, 2016; Worthy et al, 2012).

4.2 Prospect Valence Learning Model

A second model that allows researchers to estimate the possible underlying processes in the performance of the Iowa Gambling Task, is the Prospect Valence Learning Model (Steingroever et al, 2013a). This model can provide an even better fit to IGT data than Expectancy Valence Model, but it is less used. The difference between the Expectancy Valence model and the Prospect Valence Learning model is about the computing of the expectancies. Indeed, this model also assumes that participants maintain an expectancy for each deck (Worthy et al, 2012). Another difference is that the PVL model uses four parameters to formalize its assumptions about participants', instead of three (Steingroever et al, 2013a). This model assumes that decision makers process the outcome only after choosing a card from deck *D* on trial *t* (Steingroever et al, 2016). Differing from the linear utility function of the EV model, the PVL model uses a non-linear utility function from prospect theory (i.e., the Prospect Utility function; Tversky & Kahneman, 1992).

The utility, $u_D(t)$, on trial *t*, of each net outcome, is:

$$u_D(t) = \begin{cases} D(t)^{\phi} & \textit{if}\ D(t) \geq 0 \\ -\lambda \mid D(t)|^{\phi} & \textit{if}\ D(t) < 0 \end{cases} \qquad (5)$$

The Prospect Utility function contains two different parameters: the shape parameter ϕ, ranging from 0 to 1, and that determines the shape of the utility function, and the loss aversion parameter λ, that determines the sensitivity of losses and gains (Worthy et al, 2012; Worthy et al, 2013). The loss aversion parameter λ, ranging from 0 to 5, indicates the impact of the losses and gains to the utility u_D. For an extended description of the parameter see https://osf.io/tudmw/ (Gialdi, 2022).

Moreover, the PVL model assumes that, on every trial *t*, participants update the expected utilities of every deck according to the Decay learning rule (Steingroever et al, 2013a). According to this rule, the expectancies of all decks may decay or be discounted, over time, depending on the recency parameter *a*, creating a new utility function:

$$Ev(t + 1) = a \cdot Ev(t) + \delta_D(t) \cdot v(t) \qquad (6)$$

The variable δ_D represents the learning rule, and it ensures that the new utility of the chosen deck *D* is added to the expectancy of that deck (Steingroever et al, 2013b; Steingroever et al, 2016).

The PVL model also assumes a sensitivity parameter θ, which depends on the response consistency *c*, based on the following equation:

$$\theta = 3^c - 1 \qquad (7)$$

As described in the EV model, small values of *c* indicate a random choice pattern, whereas large values of *c* cause a deterministic choice pattern (Steingroever et al, 2013b; Steingroever et al, 2016; Worthy et al, 2012).

In sum, based on Steingroever and colleagues study (2016) the PVL model has four parameters: (a) the shape parameter ϕ, which determines the shape of the utility function; (b) the loss aversion parameter λ, which represent the weight of net losses and gains; (c) the recency parameter α, which determines the memory for past expectancies; and (d) the response consistency parameter *c*, which determines the balance between random and deterministic choice patterns (Steingroever et al, 2016; Worthy et al, 2012).

4.3 Conclusion

Existent literature showed that the search for a best IGT model is far from a definitive end (for a review see Steingroever et al, 2013b). Previous studies (see for example Steingroever et al, 2013b) have failed to find advantage for the EV model or the PVL model. These authors demonstrated that either of them should be accepted as the default model for the IGT data. Also, in the study of Steingroever and colleagues (2013b) the EV model provided poor fitting data to choices indicating a preference for infrequent losses decks, while the PVL model provided poor fitting data to choices with a preference for the bad decks.

A possible explanation for the inconstancy results of the previous studies, might be that they considered data with different choice patterns (Steingroever et al, 2013b). In fact, participants in Ahn and colleagues' study (2008) showed a preference for the advantageous decks (i.e., deck C and D), whereas participants in Fridberg and colleagues' study (2010) and Yechiam and Busemeyer's study (2005) indicated a preference for deck with infrequent losses (i.e., deck B and D). For this reason, the conclusion in these studies might appears so different. Moreover, it also appears clear that different models estimate more accurately specific choice patterns than others. For instance, the EV model seems to fail at predicting preferences for deck B (Yechiam & Busemeyer, 2005) and to over-predict the proportions of selections from deck A and C (i.e., the high-frequent losses decks; Fridberg et al, 2010).

Thus, the present literature indicates the need for a deeper knowledge of decision making, as it is defined by these two cognitive models (i.e., Expectancy Valence model and Prospect Valence Learning model).

Unfortunately for this project, neither the EV model nor the PVL model has been evaluated. This limitation therefore did not allow me to evaluate the associations between the various measures that have been administered (i.e., self-reports and other behavioral computerized tasks) and these cognitive models. Moreover, it was not possible to assess the presence of underlying cognitive processes in this sample of community dwelling participants.

5. Study 1

The first study of this Section aimed at assessing convergent validity correlations between the Iowa Gambling Task and the computerized version of the Balloon Analogue Risk Task. Existent research showed controversial results when comparing these two behavioral tasks; for example, Aklin and colleagues (2005) and Lejuez and colleagues (2003) found no relationship between these two tasks, and, on the contrary Skeel and colleagues (2007) found significant relationships between the IGT and BART. A critical aspect of the IGT is about the way participants made decision across trials. For instance, when created the task, Bechara and colleagues (1994) proposed the hypothesis of the somatic marker: emotions and feeling at conscious or unconscious level, may influence the decision across the administration (Bechara 2005; Bechara et al, 1997). However, cognitive components (Dunn et al, 2006) and implicit knowledge (Brand et al, 2007; Guillaume et al, 2009; Maia & McClelland, 2004; Persaud et al, 2007) also play an important role, during decision under ambiguity. At present, the IGT manual refers to the task as assessing risky decision making (Bechara, 2008).

In contrast, during the Balloon Analogue Risk Task, that has been indicated with measures' ecological validity (e.g., Aklin et al, 2005; Fernie et al, 2010; Hopko et al, 2006; Lejuez et al, 2003), participants can express their propensity to risk from the first balloon, whereas during the IGT participants first have to learn the differences between the Decks. Consequently, these two tasks seem to have similar goal (i.e., maximize the amount of money, by learning the difference between "good" balloon as well as "good" deck), but the way they are operationalized is different. This difference between tasks can be the explanation of difficulties in experimental research to find an association between the IGT and the BART (see also Bishara et al, 2009).

In addition, this first study aimed also at assessing correlations between the IGT and the impulsivity traits, as they were operationalized in the Barratt Impulsiveness Scale - 11, UPPS-P, ImpSS, and Disinhibition facet of the Personality Inventory for *DSM-5* (PID-5). As described in the previous Section for the Stop-Signal Task (see Study 1 of the Section I), it is important for clinical research to assess the correlations between two types of measurement of the same construct (for a review, Sharma et al, 2014).

5.1 Aim

Starting from these considerations, this first study aimed at evaluating the convergent validity correlations between the Iowa Gambling Task (IGT) and Balloon Analogue Risk Task (BART). The Iowa Gambling Task net score was evaluated for the total number of selections and for each block separately, by subtracting the number of cards selected from decks A and B from the number selected from cards selected from decks C and D. Based on previous findings (e.g., Lejuez et al, 2003), I expected no correlations between the BART and the IGT performance. Other studies (e.g., Upton et al, 2011) showed that disadvantageous choices in the IGT may be correlated with the BART performance, only on the later trials. These authors have differentiated trials in the IGT in early and later trials, highlighting that participants developed the explicit understanding of the IGT disadvantageous decks, through the administration.

Secondly, this study aimed at testing the correlations between the Iowa Gambling Task and different self-reports of impulsivity (i.e., BIS-11, UPPS-P, ImpSS, and PID-5). Based on modest literature, there is a poor overlap between self-report and measure of impulsivity (e.g., Cyders & Coskunpinar, 2011; Sharma et al, 2014). For this reason, I expected low correlations between IGT net scores and self-report impulsivity measures.

5.2 Material and Methods

5.2.1 Participants

The sample was composed of 174 Italian community dwelling adult participants with a mean age of 27.57 years (*SD* = 6.72 years; age range: 19 years – 60 years). In my sample seventy-nine participants (45.1%) were male, and 94 participants (53.7%) were female; one participant (0.6%) refuse to disclose his/her gender. Three participants (1.7%) were left-handed. The sample was composed of one hundred one (97.7%) unmarried participants, 3 (1.7%) married participants, and one participant (0.6%) refuse to disclose his/her civil status. Nine participants (5.1%) had a junior high school degree, 67 (38.3%) had a high school degree, 87 (49.7%) had a university degree, and 11 (6.3%) had a post-lauream degree; one participant (0.5%) refuse to report his/her educational level. Ninety-eight participants (56.0%) were students, 38 (21.7%) were blue collars, 4 (2.3%) were white collars, 16 (9.1%) were managers, 16 (9.1%) were liberal arts practitioners, and 2 (1.1%) were unemployed; one participant (0.6%) refuse to disclose his/her occupation.

To be included in the sample, participants had to document that they were of adult age (i.e., 18 years of age or older), they had no psychiatric or neurological disorders, and had normal or corrected-to-normal vision, and to agree to the written informed consent in which the study was extensively described. To avoid cultural and lexical bias in questionnaire responses, to participate in the present study, participants were required to speak Italian as their first language. All participants were treated in accordance with the Ethical Principles of Psychologists and Code of Conduct.

5.2.2 Measures

- Iowa Gambling Task (Behara et al, 1994). In the Iowa Gambling Task, participants are asked to choose a card from four different decks (namely, A, B, C, D). Two of the decks are bad decks (i.e., A and B), because they result in negative long-term outcomes (i.e., the immediate gain is high, but also the future loss is high). The remaining decks (i.e., C and D) are the good ones, because of their positive long-term outcomes (i.e., the immediate gain is smaller, but also the future loss is smaller). For an extensive description see paragraph 2.2.2 of the present Section.

- Balloon Analogue Risk Task (Lejuez et al, 2002). During this task, participants are rewarded (i.e., by gaining money) with each balloon that do not explode. This measure provides a useful tool to assess propensity toward risk. Participants are asked to pump 30 balloons (i.e., the shorter version of the BART) with three different colors. With each click on the pump, participants gain 5 cents, that eventually can be lost if the balloon exploded. For an extensive description see paragraph 2.2.1 of the present Section.

- Barratt Impulsiveness Scale – 11 (BIS-11; Patton et al, 1995). The BIS-11 is a 30-items *self-report* questionnaire designed to assess three facets of impulsivity. Items are rated on a 4-point Likert-type scale (*1 = Rarely/Never* to *4 = Almost Always/Always)*. This measure has three facets of impulsiveness: motor impulsivity, attention impulsivity, and non-planning impulsivity both in original (Patton et al, 1995) and Italian version (Fossati et al, 2001). These three facets are summed to produce a total score, and the higher the BIS-11 total score, the higher impulsivity level.

- UPPS-P Impulsive Behavior Scale (UPPS-P; Lynam et al, 2006). The UPPS-P is a 59 items *self-report* questionnaire, with Liket-type scale ranging from *1 = Agree Strongly* to *4 = Disagree Strongly*. This questionnaire was designed to measure five dimensions of impulsive behavior: Negative Urgency, Premeditation, Perseverance, Sensation Seeking, and Positive Urgency. The UPPS-P showed adequate psychometric properties (Cyders & Smith, 2007; Whiteside & Lynam 2001) also in its Italian translation (Fossati et al, 2016; Gialdi et al, 2021).

- Zuckerman-Kuhlman Personality Questionnaire Impulsive Unsocialized Sensation Seeking Scale (ImpSS; Zuckerman et al, 1993). The ImpSS is a 19 true-false items *self-report* questionnaire. This scale measures lack of planning and tendency to act on impulse without thinking, need for excitement, change and novelty, and preference for unpredictable situation. The ImpSS was provided with adequate psychometric proprieties also in its Italian translation (Carlotta et al, 2003).

- Personality Inventory for *DSM-5* (PID-5; Krueger et al, 2012). The PID-5 is a 220-items self-report with a 4-point response scale (from *0 = Very False or Often False to 3 = Very True or Often True*). The PID-5 was designed to assess the *DSM-5* traits presented in the Alternative Model of Personality Disorder (AMPD), provided in Section III (APA, 2013). The PID-5 has 25 scales that can be summed to generate five higher order dimensions (Krueger et al, 2012), which represents dysfunctional variants of the Five-Factor Model personality dimensions (APA, 2013). Specifically, the five domains of the PID-5 are: Negative Affectivity, Detachment, Antagonism, Disinhibition, and Psychoticism. The psychometric proprieties of the Italian translation of the PID-5 in nonclinical adults have been previously published (Fossati et al, 2013).

  For this project, in order to avoid a long and inaccurate compilation from participants, I relied only on the Disinhibition domain (i.e., orientation toward immediate gratification and impulsive behavior) with only 46 items, and its corresponding facets (i.e., Distractibility, Impulsivity, Rigid Perfectionism, Risk Taking, Irresponsibility).

5.2.3 Procedures

Participants were randomly organized to complete before the questionnaires and following the computerized tasks or vice versa. In the whole project, all measures and tasks were administered in their Italian translation. In order to match the self-reports scores and tasks results and to maintain anonymity, each participant included in the sample created an alphanumeric ID code.

Participants completed the study online using Online Surveys Jisc, an online survey tool designed for academic research (https://www.onlinesurveys.ac.uk/); participants volunteered to take part in the study receiving no economic incentive or academic credit for their participation. Self-report measures were administered in random order and scored blind to the computerized task results. The computerized tasks were administered using a laptop computer in individual session and each session lasted on average two hours per participant.

Written informed consent was obtained prior to study participation; all participants were of adult age and volunteered to take part in the present study after it was extensively described. Institutional Review Board was obtained for all aspects of the study.

5.3 Data Analysis

Descriptive statistics were used to describe the sample. Cronbach's alpha coefficient, Omega (ω) coefficient, and mean inter-item correlation (MIC) were used to estimate the internal consistency reliability of the self-report measures of impulsivity (i.e., BIS-11, UPPS-P, ImpSS, PID-5) in the whole sample. The limited size of the sample strongly suggested to rely on non-parametric statistics for hypothesis testing. Spearman *r* coefficient. Due to the limited number of participants in this study, to evaluate the correlations between the Balloon Analogue Risk Task and the Iowa Gambling Task and estimate the convergent validity, and to assess the correlation between the Iowa Gambling Task and self-report questionnaires, I relied on the Spearman's *r* coefficient.

5.4 Results

Detailed description of the sample who took part to this study is reported in Table 5.1.

Table 5.1

Description of the Sample: Demographic Variables ($N = 173$[1]).

Demographic Variables	*n / M*	*% / SD*
Civil Status:		
Unmarried	171	97.7
Married	3	1.7
Education Level:		
Junior High School	9	5.1
High School	67	38.3
University Degree	87	49.7
Post-graduate Degree	11	6.3
Occupation:		
Student	98	56.0
Blue Collar	38	21.7
White Collar	4	2.3
Manager	16	9.1
Liberal arts practitioners	16	9.1
Unemployed	2	1.1
Age (years)	27.57	6.72

Note. 1: one participant (0.6%) refused to disclose his/her civil status, educational level, and his/her occupation.

The descriptive statistics, the Cronbach's α values, Omega (ω) values, and item-total correlation values, for the BIS-11 subscale and total score in this sample are reported in Table 5.2.

Table 5.2

The BIS-11 Scales: Descriptive Statistics, Internal Consistency Reliability (i.e., Cronbach's α and McDonald ω Coefficient) Estimates, and Scale Intercorrelation (N = 174).

BIS-11 subscale	*M*	*SD*	α	ω	MIC
Motor impulsivity	19.87	4.77	.74	.87	.23
Attention impulsivity	22.73	4.02	.47	.61	.08
Non-planning impulsivity	17.45	3.92	.72	.88	.24
BIS-11 Total Score	60.06	9.50	.76	.88	.11

Note. MIC = Mean Inter-Item Correlation

The Cronbach's α values, Omega (ω) values, and mean inter-item correlation (MIC) for the UPPS-P subscale in this sample are listed in Table 5.3.

Table 5.3

The UPPS-P Scales: Descriptive Statistics, Internal Consistency Reliability (i.e., Cronbach's α and McDonald ω Coefficient) Estimates, and Scale Intercorrelation (N = 174).

UPPS-P subscale	*M*	*SD*	α	ω	MIC
Negative Urgency	26.95	6.96	.86	.91	.34
Premeditation	20.03	4.92	.84	.91	.35
Perseverance	17.89	5.03	.85	.93	.39
Sensation Seeking	28.17	8.40	.88	.93	.39
Positive Urgency	24.35	7.93	.91	.94	.44

Note. MIC = Mean Inter-Item Correlation

Descriptive statistics, Cronbach's α values, Omega (ω) values, and mean inter-item correlation (MIC) for the ImpSS scale total score are reported in Table 5.4.

Table 5.4

The ImpSS Scale: Descriptive Statistics, Internal Consistency Reliability (i.e., Cronbach's α and McDonald ω Coefficient) Estimates, and Scale Intercorrelation (N = 174).

ImpSS Scale	*M*	*SD*	α	ω	MIC
ImpSS Total Score	6.22	4.60	.87	.91	.25

Note. MIC = Mean Inter-Item Correlation

Finally, Cronbach's α values, Omega (ω) values, and mean inter-item correlation (MIC), and descriptive statistics for the PID-5 Disinhibition subscales in this sample are reported in Table 5.5.

Table 5.5

The PID-5 Facets: Descriptive Statistics, Internal Consistency Reliability (i.e., Cronbach's α and McDonald ω Coefficient) Estimates, and Scale Intercorrelation (N = 174).

PID-5 facet	*M*	*SD*	α	ω	MIC
Distractibility	0.88	0.67	.92	.96	.55
Impulsivity	0.74	0.62	.88	.95	.53
Rigid Perfectionism	1.54	0.70	.90	.95	.49
Risk Taking	1.17	0.61	.92	.95	.45
Irresponsibility	0.49	0.44	.72	.88	.28
Disinhibition	0.95	0.39	.89	.95	.61

Note. MIC = Mean Inter-Item Correlation

To assess the correlation between IGT and BART, total net score and net scores of each five block for the IGT performance and mean blue balloon for the BART performance are considered. As previously mentioned, the blue balloons are those the burst less frequently. For this reason, they considered as a dependent variable. However, no significant correlations were found between the total net score, net scores of each block (i.e., Total Net Score, Block 1, Block 2, Block 3, Block 4, Block 5) and the mean blue balloon, r = .06, p > .40, r = -.17, p > .10, r = .00, p > .90; r = .06, p > .40, r = .08, p > .25, respectively. Only the fifth block of the Iowa Gambling Task showed a negligible but still significant correlation with the BART, r = .17, p < .05.

The Spearman r values for the associations between the IGT estimates and the BIS-11, UPPS-P, ImpSS and PID-5 scale scores are summarized in Table 5.6.

In this sample, the total IGT net score showed significant and negative correlations with ImpSS Total score, r = -.19, and PID-5 Risk-Taking facet, r = -.20, all p < .05. Moreover, Block 5 IGT net score showed negative and significant associations with UPPS-P Sensation Seeking, r = -.19, p < .05. Negative and significant associations were also found between Block 5 IGT net score and ImpSS Total Score, r = -.20, and PID-5 Risk Taking, r = -.20, all p < .01

Table 5.6

The Iowa Gambling Task: Correlations (i.e., Spearman's *r* coefficient) with the disinhibition self-report measures (*N* = 174).

	Iowa Gambling Task									
	Deck A	Deck B	Deck C	Deck D	Net Score	Net Score 1	Net Score 2	Net Score 3	Net Score 4	Net Score 5
Self-report	*r*	*r*	*r*	*r*	*r*	*r*	*r*	*r*	*r*	*r*
BIS-11 Motor	.02	.02	-.03	-.08	-.04	.02	.06	-.07	-.02	-.09
BIS-11 Attention	-.02	.07	-.09	-.05	-.09	-.09	.04	-.07	-.09	-.09
BIS-11 Non-Planning	.03	.02	-.12	.05	-.04	.04	.05	-.06	-.06	-.06
BIS-11 Total Score	.06	.07	-.12	-.05	-.11	-.02	.05	-.12	-.10	-.13
UPPS-P NU	.03	.05	-.07	-.12	-.08	-.04	.06	-.10	-.04	-.11
UPPS-P Prem	<u>.16</u>	.08	-.07	-.05	-.12	-.04	-.01	-.14	-.06	-.13
UPPS-P Pers	-.06	.00	.00	-.03	.01	-.07	.05	-.04	.07	.05
UPPS-P SS	.13	.11	-.09	-.02	-.14	-.05	-.01	-.09	-.04	<u>-.19</u>
UPPS-P PU	-.09	-.06	.03	-.02	.07	-.02	.13	.03	.10	.03
ImpSS Total Score	**.23**	.10	-.14	-.09	<u>-.19</u>	-.04	.00	**-.20**	-.09	**-.20**
PID-5 Distractibility	-.06	-.01	.00	-.06	-.01	-.03	.08	-.05	.02	.01
PID-5 Impulsivity	<u>.16</u>	.02	-.10	-.05	-.09	-.06	.04	-.15	-.05	-.10
PID-5 Rigid Perfectionism	-.11	-.02	.07	-.02	.03	.10	-.03	-.01	.03	.02
PID-5 Risk Taking	**.25**	.13	-.09	-.09	<u>-.20</u>	-.08	-.05	<u>-.15</u>	-.10	**-.20**
PID-5 Irresponsibility	-.05	.04	-.05	-.06	-.04	-.14	.11	-.08	.00	.00
PID-5 Disinhibition	.15	.08	-.11	-.07	-.04	-.13	.05	-.13	-.06	-.12

Note. Underlined correlations: p < .05; bold correlations: p < .01. NU: Negative Urgency, Prem: Premeditation; Pers: Perseveration; SS: Sensation Seeking; PU: Positive Urgency.

Lastly, a comparison on the net scores of the IGT between males and females has been evaluated and results are presented in Table 5.7.

Table 5.7

Males (n = 79) versus females (n = 94): Detailed Comparisons on IGT Net Scores.

	Male (n = 79)		Female (n = 94)				
	M	*SD*	*M*	*SD*	*t*	*df*	*d*
Net score	11.59	33.98	15.40	32.65	-.75	171	-0.11
Net score 1	-3.70	8.57	-2.79	6.43	-.80	171	-0.12
Net score 2	1.90	7.62	2.66	7.25	-.67	171	-0.10
Net score 3	4.00	8.98	4.62	8.92	-.45	171	-0.07
Net score 4	3.85	10.01	4.91	10.44	-.68	171	-0.10
Net score 5	6.05	11.20	6.00	10.82	.30	171	0.05

Note. d: Cohen's d coefficient. N = 173, because one participant refused to disclosure his/her gender.

5.5 Discussion

The present study examined the relationship between two tasks that apparently should measure the same construct of impulsivity, such as the Iowa Gambling Task and the Balloon Analogue Task. During both tasks participants are asked to gain as much money as possible, by identifying the disadvantageous deck in the IGT (i.e., decks C and D) and the balloon that burst after more pumps (i.e., blue balloon). Performance on the IGT can be evaluated by calculating the total net score, that is the differences between the advantageous decks and the disadvantageous ones. On the other hand, performance on the BART can be evaluated by considering the mean of the click on three balloons. However, only the blue balloons can describe more precisely risk behavior, since they are the balloon that burst after a high number of click. In this study to assess the correlation between IGT and BART, total net score and net scores of each five blocks for the IGT performance and mean blue balloon for the BART performance are considered.

Confirming previous findings (Lejuez et al, 2003; Upton et al, 2011), no significant correlations were found between the total net score, and the mean blue balloon, Spearman's $r = .06$, $p > .40$. However, previous studies (e.g., Buelow & Blaine, 2015) showed that IGT performance can be distinguished in early and later trials: the firsts measure decision making under ambiguity, while the seconds measure decision making under risk. Thus, I tried to measure the correlations between the net scores calculated in each block with the mean blue balloon. However, also in this case no significant correlations were found, $r = -.17$, $p > .10$, $r = .00$, $p > .90$; $r = .06$, $p > .40$, $r = .08$, $p > .25$, for Block 1, Block 2, Block 3, Block 4, respectively. Only the last block (i.e., Block 5) yielded small and negligible, but still significant, correlation with the BART, $r = .17$, $p < .05$. To better understand these results, Spearman's r correlations between the BART mean blue balloon and the frequencies of each deck in each block were calculated. Even these results did not show significant correlations, r min. = $|.09|$ (mean blue balloon and frequency deck C in Block 1), r max. = $-|.13|$ (mean blue balloon and frequency deck D in Block 4), all $p > .08$. Only frequency of Deck B on Block 1, was significantly associated with the mean of blue balloon, $r = .15$, $p < .05$. This result was not consistent with previous findings of early versus later trials (Buelow & Blaine, 2015), but can explain that at the beginning of the IGT administration, participants initially preferred the decks with infrequent losses and with high immediate gain of money, that correlated with a high number of clicks to pump the blue balloon in the BART. However, during the task participants has begun to understand that Deck B is not an advantageous deck, and correlations with decks and the mean of click number on the blue balloon started decreasing.

As stated in the description of study (see paragraph 5 of the present Section), in the IGT participants must experience more cards' selection, compared to the BART, where the risky balloon is understandable from the beginning (see also Buelow & Blaine, 2015). For this reason, it is possible that disadvantageous choices on the IGT may not be related to BART mean click on the blue balloon. Concluding, the present results are consistent with previous studies and add to the existing perspectives on these tasks that IGT measures a different type of decision making than BART (Aklin et al, 2005; Buuelow & Blaine, 2015; Lejuez et al, 2003; Wood & Bechara, 2014).

In this study I also assessed the correlation coefficients between the performance on the IGT, considering the frequencies of selection in each Deck, the total net score, and the net scores of each block, and self-report of impulsivity. These results showed significant and negative correlations between the total IGT net score and the ImpSS Total Score, and more specifically between the last block (i.e., Block 5) on the IGT and the ImpSS Total Score, $r = -.19$, $p < .05$, and $r = -.20$, $p < .01$, respectively. The ImpSS was intended to measure lack of planning and these results are consistent with previous research (e.g., Vasconcelos et al, 2014), indicating that individuals with difficulty in planning tend to choose disadvantageous cards during the entire administration and specifically on IGT final block. Moreover, to confirm previous results, the analysis between the ImpSS and the IGT showed positive and significant association also with the number of selections of Deck A, $r = .23$, $p < .01$. Moreover, Block 5 IGT net score showed negative and significant associations also with UPPS-P Sensation Seeking sub-scale, $r = -.19$, $p < .05$. Thus, individuals with high tendency to seek excitement are more incline to choose disadvantageous decks.

Lastly, the same analyses were assessed between the IGT performance and the PID-5 Disinhibition facets. These results showed a positive and significant correlations between the number of selections in Deck A and PID-5 Risk Taking, $r = .25$, $p < .05$, confirming that participants who tend to choose disadvantageous decks have higher risk-taking characteristics. More specifically, also the PID-5 Risk Taking facet showed negative and significant correlations with the total IGT net score and Block 5 IGT net score, $r = -.20$, $p < .05$, and $r = -.20$, $p < .01$, respectively.

For a better comprehension of the performance on the IGT, a formal comparison between the performance between males and females has been calculated. Actually, present literature shows contrasting results with no differences between men and women regarding decision making tasks (e.g., Lighthall et al, 2009; Lighthall et al, 2011; Starcke et al, 2008). However, based on previous findings on the Iowa Gambling Task, I expected to find differences in the total net scores between subjects, with a higher total net score for men compared to women performance (e.g., Bolla et al, 2004; Overman, 2004; van den Bos et al, 2007; van den Bos et al, 2012; etc.). As a consequence, the final budget of the IGT performance for women should be lower than the total budget of men. However, in my sample, no significant differences between genders on the total net score and the net scores across five blocks was found. Actually, even if the differences are not significant, in my sample women performed better than men with a total net score slightly higher: $M_F = 15.40$ ($SD = 32.65$), $M_M = 11.59$ ($SD = 33.98$). These results were consistent with d'Acremont and Van der Linden results (2006) and Overman and Pierce (2013), where they have found small but significant differences between gender, where girls tend to choose more advantageous choices and earn more money.

5.6 Limitations

The results of the present study should be considered in the light of several limitations. Firstly, the sample was not well balanced between males and females, and this can limit the results obtained with the comparison between gender. However, existent literature is very controversial and with different results (e.g., Bolla et al, 2004; Lighthall et al, 2009; Lighthall et al, 2011; Overman, 2004; van den Bos et al, 2012; etc.). Although it was larger, the sample size of my study was quite limited. Indeed, administering neuropsychological tasks requires a great deal of time (i.e., more than 1 hour per participants), and resources (i.e., all participants were administered the measures in the laboratory) (e.g., Crawford & Garthwaite, 2002). For this reason, this sample should be considered more as a convenient study group, than a sample representative of the Italian population, and this can limit the generalizability of the findings to other populations (i.e., clinical, or forensic sample).

Moreover, it should be observed also that I did not compute Cognitive Model of IGT (e.g., EV model and PVL model). Indeed, based on Bishara et al (2009), these approaches may offer new evidence for the relationship between the Iowa Gambling Task and the Balloon Analogue Risk Task, since the use of this model may provide relationships at the model parameter level. These results, in fact, suggest the important to use more advanced model to estimate possible underlying processes in decision-making tasks, since these cognitive models (specifically the EV model) have been already used to understand decision-making behaviors in clinical population [i.e., cocaine abuser (Stout et al, 2004), Parkinson's and Huntington diseases (Busemeyer & Stout, 2002), and others (Yechiam et al, 2005)]. The presence of underlying process during a task has been described in several studies (e.g., Hedge et al, 2018), which stated the importance not only to measure them but also to always take in consideration the influence that they might have in different performance.

In addition, it is possible the administer more than the classical 100 trials to evaluate if a difference in performance would occur. Indeed, previous study (Overman & Pierce, 2013) have found that the overall IGT performance improved during the administration of the task (i.e., increasing of the selection of advantageous decks). Otherwise, like usually happen for neuropsychological task, it is possible to administer a practice phase before the experimental phase, since the comprehension of the task might be not immediately clear for the participants.

Lastly, I relied only on self-reports to assess impulsive behaviors and only on the Disinhibition domain of PID-5; however, using different measures of impulsivity, including interview, that specifically assess behaviors related to reduce impulse control (e.g., pathological gambling, substance abuse, etc.), or different self-report measures that assess for decision-making style (e.g., General Decision-Making Style; Scott & Bruce, 1995) may yield different results.

6. Study 2

The second study of this section on the Iowa Gambling Task focused on the test-retest reliability. As mentioned for the second study of the first Section (i.e., paragraph 6 of the previous Section), assessing reliability of any psychological measure is important for both research and clinical practice. In addition, for what concern the Iowa Gambling Task the temporal stability of the task may be invalidated by practice effects on later administrations and on multiple administrations (Buelow & Suhr, 2009). For example, the study of Ernst and colleagues (2003a) showed the presence of a learning effect in control participants and substance use disorders participants, with a one-week test-retest paradigm (see also, Ernst et al, 2003b). However, a one-week test-retest paradigm may be too short to really assess the temporal stability of a measure. Nevertheless, current literature has contrast findings. For example, Xu and colleagues (2013) has found moderately temporal stability with a two-weeks temporal stability; but, on the other hand, more recently Schmitz and colleagues (2020) found lower reliability estimates.

Thus, in line with neuropsychological research, in this study I assessed reliability estimations in terms of temporal stability (i.e., Spearman *r* coefficient) and Intraclass Correlation Coefficient (ICC) for absolute agreement based on random-effect one-way ANOVA. However, in this study I relied also on internal consistency (i.e., Cronbach's alpha and Omega coefficient), considering neuropsychological tasks as psychological tests, and consequently considering each Iowa Gambling Task block as a part of a test.

The inconsistency between studies regarding reliability of the Iowa Gambling Task may limits its applicability in research and clinical practice (e.g., during pre/post intervention studies; Buelow & Shur (2009). Considering all these aspects, it is important to assess Iowa Gambling Task temporal stability, looking for the best way to test reliability for this type of neuropsychological task

6.1 Aim

The aim of the present study was to assess test-retest reliability of the Iowa Gambling Task, considering the general net score, and internal consistency reliability across decks and the five blocks of the Iowa Gambling Task with a three-months test-retest paradigm in a un sub-sample of participants (*n* = 134), who agreed to take part at a three-months follow-up assessment.

In this study, Spearman's *r* coefficient, Cronbach's *α* and McDonald omega coefficients have been calculated.

In line with previous study (Buelow & Suhr, 2009; Ernst et al, 2003; Ernst et al, 2003b; Schmitz et al, 2020), I expected that the Iowa Gambling Task has moderately low temporal stability when tested with Spearman's *r* coefficient. However, since the Iowa Gambling Task is provided with good criterion validity, I also expected that internal consistency reliability was adequate in the two different time of administration.

6.2 Material and Methods

6.2.1 Participants

The sample was composed of 134 Italian community dwelling adult participants with a mean age of 26.15 years (SD = 5.82 years; age range: 19 years – 55 years), who agreed to take part to the three-months follow-up. In my sample 57 participants (42.5%) were male, and 76 participants (56.7%) were female; one participant (0.7%) refuse to disclose his/her gender. Three participants (2.2%) were left-handed. The sample was composed of one-hundred thirty-one (97.8%) unmarried participants, and 3 (2.2%) married participants. Seven participants (5.2%) had a junior high school degree, 57 (42.5%) had a high school degree, 65 (48.5%) had a university degree, and 5 (3.7%) had a post-lauream degree. Eighty-six participants (64.2%) were students, 24 (17.9%) were blue collars, 2 (1.5%) were white collars, 12 (9.0%) were managers, and 10 (7.5%) were liberal arts practitioners.

Participants who completed the follow-up of the Iowa Gambling Task did not significantly on age, $t(172) = 5.54$, $p > .10$, $d = .84$, gender, $\chi^2(2) = 2.14$, $p > .34$, Cramer $V = .11$, civil status, $\chi^2(1) = .91$, $p > .34$, $\varphi = .07$, then participants who did not completed the follow-up. However, the two samples were significantly different on educational level, $\chi^2(3) = 8.91$, $p < .03$, Cramer $V = .23$, and occupation, $\chi^2(5) = 20.80$, $p < .01$, Cramer $V = .35$.

To be included in the sample, participants had to agree to the written informed consent in which the study was extensively described. To avoid cultural and lexical bias in questionnaire responses, to participate in the present study, participants were required to speak Italian as their first language. All participants were treated in accordance with the Ethical Principles of Psychologists and Code of Conduct.

6.2.2 Measures

- Iowa Gambling Task (Bechara et al, 1994). See paragraph 5.2.2 and 2.2.2 for an extended description.

6.2.3 Procedure

Participants who agreed to take part to the second part of this study completed only the computerized Iowa Gambling Task, in its Italian translation. In order to match the task results and to maintain anonymity, each participant included in the sample created the originally alphanumeric ID code, following the same instructions in both administration (i.e., first letter of mother's name, first letter of father name, number of letters of the surname, date of birth).

Participants volunteered to take part in the second part of the study receiving no economic incentive or academic credit for their participation. The Iowa Gambling Task was administered using a laptop computer in individual session and each session lasted on average two hours per participant.

Written informed consent was obtained prior to study participation; all participants were of adult age and volunteered to take part in the present study after it was extensively described. Institutional Review Board was obtained for all aspects of the study.

6.3 Data Analysis

Spearman r coefficient and Intraclass Correlation Coefficient (ICC) were used to assess temporal stability. Correlations of .70 or higher indicate good test-reliability, and correlations between .40 and .60 indicate moderate test-retest reliability. Good test-retest reliabilities are a necessary prerequisite for the validity of any measure (Becser et al, 1998; Kuntsi et al, 2005).

Moreover, to assess reliability I relied also on Cronbach's α coefficient and Omega (ω) coefficient for the internal consistency of the Iowa Gambling Task. High internal consistency of a measure indicates high homogeneity of the scale. In this study I considered the 6 blocks of Iowa Gambling Task as part of a test.

6.4 Results

Descriptive statistics and mean comparison between participants who completed the follow-up (n = 134) and participants who did not complete the follow-up at three months (n = 40) of the Iowa Gambling Task, are presented in Table 6.1.

Table 6.1.

Participants who completed the follow-up (n = 134) and participants who did not complete the follow-up at three months (n = 39) of the Iowa Gambling Task: Detailed Comparisons on Demographic Variables.

	Follow-up (n = 134)		No Follow-up (n = 39[1])				
Demographic Variables	*n / M*	*% / SD*	*n / M*	*% / SD*	χ^2 / t	*df*	φ/*V*/*d*
Civil Status:							
Unmarried	131	97.8	40	97.6			
Married	3	2.2	--	--	.91	1	.07
Education Level:							
Junior High School	7	5.2	2	4.9			
High School	57	42.5	10	24.4			
University Degree	65	48.5	22	53.7			
Post-lauream Degree	5	3.7	6	14.6	8.91	3	.23
Profession: [2]							
Student	86	64.2	12	39.3			
Blue Collar	24	17.9	14	34.1			
White Collar	2	1.5	2	4.9			
Manager	12	9.0	4	9.8			
Liberal art practitioner	10	7.5	6	14.6			
Retyred	--	--	2	4.9	20.80	5	.35
Gender							
Male	57	42.5	22	53.7			
Female	76	56.7	18	43.9	2.14	2	.11
Age (years) [3]	26.15	5.82	32.35	7.39	5.54	172	.84

Note. 1: one participant (0.6%) refused to disclose his/her civil status, educational level, and his/her occupation. --: values not present.

φ: Phi; V: Cramer's V coefficient; d: Cohen's d coefficient.

The descriptive statistics, Spearman r coefficient, and Intraclass Correlation Coefficient (ICC) for the Iowa Gambling Task decks' selection and net scores are reported in Table 6.2 and Table 6.3, respectively.

Table 6.2

The Iowa Gambling Task test-retest reliability: Descriptive Statistics, Spearman *r* coefficient, and Intraclass Correlation Coefficient (ICC) in two time of administrations (*n* = 134) with the number of selections for each deck.

	Time 1		Time 2					
Iowa Gambling Task	*M*	*SD*	*M*	*SD*	*t*(133)	*d*	*r*	ICC
Deck A	15.85	6.90	13.87	7.96	2.92**	0.51	.46**	.44**
Deck B	27.01	12.46	23.34	13.45	2.87**	0.50	.39**	.35**
Deck C	24.05	13.88	27.10	19.12	-2.03*	-0.35	.56**	.46**
Deck D	33.09	15.13	35.70	19.39	-1.45	-0.25	.24**	.28**

Note. D: Cohen's d coefficient; $* p < .05$; $** p < .01$

Table 6.3

The Iowa Gambling Task test-retest reliability: Descriptive Statistics, Spearman *r* coefficient, and Intraclass Correlation Coefficient (ICC) in two time of administrations (*n* = 134) with the total net score and net scores across the five blocks.

	Time 1		Time 2					
Iowa Gambling Task	*M*	*SD*	*M*	*SD*	*t*(133)	*d*	*r*	ICC
Total Net Score	14.28	31.16	25.60	35.91	-3.80**	-0.66	.49**	.47**
Net Score 1	-3.13	6.85	-1.64	7.53	-1.73	-0.30	.02	.04
Net Score 2	2.52	7.25	4.97	8.61	-2.77*	-0.48	.09	.18*
Net Score 3	4.40	8.54	6.73	9.12	-2.83*	-0.49	.42**	.42**
Net Score 4	4.75	9.85	7.36	10.05	-2.87*	-0.50	.45**	.44**
Net Score 5	6.04	10.10	8.18	10.70	-2.14*	-0.37	.40**	.39**

Note. D: Cohen's d coefficient; $* p < .05$; $** p < .01$

These differences between two times of administrations can be explained with the first assumption of Bechara and colleagues (1994). According to this assumption participants, during the administration, learn to prefer the good options over the bad options, resulting in significant differences between the first and the last block. These results are presented in Table 6.4.

Table 6.4
The Iowa Gambling Task mean differences across trials: Descriptive Statistics and *t*-test in two time of administrations (n = 134) considering the changing across blocks.

	Block 1		Block 5			
Iowa Gambling Task	*M*	*SD*	*M*	*SD*	*t*(133)	*d*
Time 1						
Deck A	4.28	1.85	2.69	2.67	5.49**	0.95
Deck B	7.29	3.52	4.39	4.23	6.15**	1.07
Deck C	3.87	2.22	5.84	5.03	-4.48**	-0.78
Deck D	4.56	3.06	7.09	5.07	-5.35**	-0.93
Time 2						
Deck A	4.26	2.13	1.92	2.09	10.86**	1.88
Deck B	6.56	2.97	3.99	4.50	5.48**	0.95
Deck C	4.47	2.78	6.36	5.86	-3.93**	-0.68
Deck D	4.71	2.86	7.73	6.01	-5.72**	-0.99

Note. ** $p < .001$

The descriptive statistics, Cronbach's α values, Omega (ω) values for the Iowa Gambling Task in the two times of administrations with the net scores across the five blocks are reported in Table 6.5.

Table 6.5

The Iowa Gambling Task test-retest reliability: Descriptive Statistics Cronbach's α values, Omega (ω) values in two time of administrations with the net scores (n = 134).

	Time 1		Time 2	
Iowa Gambling Task	*M*	*SD*	*M*	*SD*
Net Score 1	-3.13	6.85	-1.64	7.53
Net Score 2	2.52	7.25	4.97	8.61
Net Score 3	4.40	8.54	6.73	9.12
Net Score 4	4.75	9.85	7.36	10.05
Net Score 5	6.04	10.10	8.18	10.70
α	.75		.83	
ω	.95		.96	

6.5 Discussion

As a whole, the results of the present study suggested that test-retest reliability assessed with Spearman *r* coefficient and Intraclass Correlation Coefficient (ICC), yielded low to modest associations between two times of administrations, across decks and the five blocks of the Iowa Gambling Task. However, Cronbach's alpha and Omega coefficients, showed good internal consistency in both times of administrations.

Confirming and extending previous results (e.g., (Buelow & Suhr, 2009; Ernst et al, 2003; Ernst et al, 2003b; Schmitz et al, 2020), and in contrast to findings from other studies (Xu et al, 2013), the Iowa Gambling Task was provided with low to moderate test-retest reliability between decks, Spearman *r* coefficient ranging from .24 (Deck D) to .56 (Deck C). With the exception of Deck D, also *t-test* analysis showed that the differences between the two time of administrations were significantly different with a decrease of the selections in disadvantageous decks (e.g., M_1 = 27.01, SD_1 = 12.46, M_2 = 23.34, SD_2 = 13.45, for Deck B) and an increase of selection in advantageous choices. Even if the difference between Deck D in the two time of administrations, the difference in mean selection of the Deck increase after three months. Similar results were shown when I considered the net scores, with Spearman *r* coefficient ranging from .02 (Net score 1) to .45 (Net score 4).

As previously mentioned in the results, these differences between two times of administrations can be explained with the first assumption of Bechara and colleagues (1994), where participants learn to prefer the good options over the bad options. This assumption has been confirmed analyzing the differences in all four decks between the first block and the last block. This result showed that participants on average start with the disadvantageous decks, because of the high immediate gain, but after few blocks they learn that Deck C and Deck D may be a better solution to increase the initial amount of money. This knowledge remains also after three months. These findings support previous studies (e.g., Buelow & Suhr, 2009; Ernst et al, 2003a; Ernst et al, 2003b) where a learning effect was supposed in the performance of the Iowa Gambling Task. For example, Verdejo-Garcia and Perez-Garcia (2007) have found that learning effects was present both in normal control and in abstinent substance users (i.e., cocaine and marijuana) after

25 days. This result extends previous literature about the occurrence of external variables that con influence the performance of the subject in two time of administration (Hedge et al, 2018).

Based on these findings, it appears very important to find a different estimation to assess the reliability of this task. For this reason, in order to evaluate the reliability of the Iowa Gambling Task, I estimated two different internal consistency coefficients (i.e., Cronbach's alpha and Omega), since considering these types of estimations may yield to different conclusion about the reliability of the Iowa Gambling Task. In this study I considered the five blocks of the Iowa Gambling Task as a part of a test and the I assessed the internal consistency in the two time of administrations with the net score estimations evaluated with the difference between the advantageous decks and the disadvantageous decks. The result of this analysis showed that the Iowa Gambling Task is highly reliable with Cronbach's alpha of .75 in Time 1 and .83 in Time 2. Similarly, the Omega coefficients showed overlapping results, with ω estimations of .95 and .96, for Time 1 and Time 2 respectively. High internal consistency of the task in a three-month paradigm indicates high homogeneity of the Iowa Gambling Task.

The present results confirmed the assumption of Hedge and colleagues (2018) that different types of estimation to assess the reliability of cognitive and neuropsychological tasks may be a better way to understand the psychometric proprieties of a task.

6.6 Limitations

The results of the present study should be considered in the light of several important limitations. First of all, for the purpose of this second study, the sample size was quite limited. Although administering neuropsychological tasks requires a great deal of time (i.e., more than one hour per participants), and resources (i.e., all participants were administered the measures in the laboratory) (e.g., Crawford & Garthwaite, 2002), the drop-out of the participants in the two times of administrations was quite elevated. Thus, my sample was more a convenient study group than a sample representative of the Italian population. Of course, this can limit the generalizability of the findings to other populations (i.e., clinical, or forensic sample).

As I mentioned before, I did not compute Cognitive Model of IGT (e.g., EV model and PVL model), that can be provide information about underlying processes, that may influence the performance between two time of administrations. Indeed, Hedge and colleagues (2018) stated that individual differences or other processes may influence the reliability of neuropsychological task. In the case of the Iowa Gambling Task, a possible process that influence the performance might be the learning effect.

Lastly, a test-retest reliability depends heavily on the length of the retest period (Kaplan & Saccuzzo, 2017; Polit, 2014). Therefore, it is important to consider that underlying processes can change rapidly in a three-month test-retest paradigm, and this can explain poor test-retest reliability

7. Study 3

7.1 Aim

The last study aimed at assessing the possible relationship between a decision-making task, as it is operationalized with the Iowa Gambling Task, and an impulsivity task, as it is operationalized with the Stop-Signal Task. Based on previous studies (Heathcote et al, 2019; Matzke et al, 2019), and on previous conclusion of this project (see paragraph 7 of the previous Section), the relationship between these two tasks might be useful at assessing whether the Stop-Signal Task with an elevated number of trials (i.e., 384 trials) can be considered a difficult choices task or a decision-making task.

Based on previous studies (Verbruggen et al, 2013), positive associations between SSRT and IGT scores are hypothesized. Indeed, previous studies (Noël et al, 2007; Shuster & Toplak, 2009) have found that poor response inhibition was associated to riskier decision in the Iowa-Gambling Task.

7.2 Material and Methods

7.2.1 Participants

The sample was composed of 158 Italian community dwelling adult participants with a mean age of 27.44 years (*SD* = 6.00 years; age range: 19 years – 47 years). In my sample seventy-three participants (46.2 %) were male, and 84 participants (53.3%) were female; one participant (0.6%) refuse to disclose his/her gender. Three participants (1.9%) were left-handed. The sample was composed of one hundred-six (98.7%) unmarried participants, and 2 (1.3%) married participants. Eight participants (5.1%) had a junior high school degree, 63 (39.3%) had a high school degree, 76 (48.1%) had a university degree, and 11 (7.0%) had a post-lauream degree. Eighty-nine participants (56.3%) were students, 34 (21.5%) were blue collars, 4 (2.5%) were white collars, 15 (9.5%) were managers, 14 (8.9%) were liberal arts practitioners, and 2 (1.3%) were unemployed.

To be included in the sample, participants had to document that they were of adult age (i.e., 18 years of age or older), they had no psychiatric or neurological disorders, and had normal or corrected-to-normal vision, and to agree to the written informed consent in which the study was extensively described. To avoid cultural and lexical bias in questionnaire responses, to participate in the present study, participants were required to speak Italian as their first language. All participants were treated in accordance with the Ethical Principles of Psychologists and Code of Conduct.

7.2.2 Measures

- Iowa Gambling Task (Bechara et al, 1994). See paragraph 5.2.2 and 2.2.2 of the present Section for an extended description.
- Stop-Signal Task (Verbruggen et al, 2019). See paragraph 5.2.2 and 2.2.3 of Section I for an extended description.

7.2.3 Procedures

Participants who agreed to take part to the present study completed randomly the computerized Iowa Gambling Task and the Stop-Signal Task, in its Italian translation, relying on PEBL platform (see paragraph 2.2.1 of Section I) and on the open-source paradigm of the Stop-Signal Task provided by Verbruggen and colleagues (2019; see paragraph 2.2.3 of Section I). In order to match the task results and to maintain anonymity, each participant included in the sample created an alphanumeric ID code.

Participants volunteered to take part in the study receiving no economic incentive or academic credit for their participation. The Iowa Gambling Task and Stop-Signal Task were administered using a laptop computer in individual session and each session lasted on average two hours per participant.

Written informed consent was obtained prior to study participation; all participants were of adult age and volunteered to take part in the present study after it was extensively described. Institutional Review Board was obtained for all aspects of the study.

7.3 Data analysis

Spearman *r* coefficient was used to assess bivariate associations between the two neuropsychological tasks.

In the present study, since I did not assess the advanced cognitive model of the Iowa Gambling Task (see paragraph 4 of the present Section), I relied on non-parametric methods to estimate SSRT (see paragraph 3 of the previous Section). Specifically, I computed the mean method to assess if it provides association with the frequentist estimation of the Iowa Gambling Task (i.e., net scores). Moreover, I also computed non-parametric SSRT estimates based on the integration method with replacement of go omissions. The overall estimates of both Iowa Gambling Task and Stop-Signal Task were assessed. Moreover, the estimations of SSRT and net score were also assessed in the specific blocks for the Stop-Signal Task and the Iowa Gambling Task, respectively.

7.4 Results

Table 7.1 represents the correlations coefficient (i.e., Spearman's *r* coefficient) between the Stop-Signal Task estimations the Iowa Gambling Task estimations. However, as detailed shown from the Table, no significant correlation was found between these two tasks

Table 7.1

The Iowa Gambling Task and the Stop-Signal Task: Correlations (i.e., Spearman's r coefficient) between tasks (N = 158).

	Iowa Gambling Task									
	Deck A	Deck B	Deck C	Deck D	Net Score	Net Score 1	Net Score 2	Net Score 3	Net Score 4	Net Score 5
Stop-Signal Task	*r*	*r*	*r*	*r*	*r*	*r*	*r*	*r*	*r*	*r*
Mean Method	0.07	-0.04	-0.03	0.03	0.01	0.05	-0.08	0.05	0.05	-0.06
Mean Method 1	0.01	0.05	-0.09	-0.01	-0.04	-0.01	-0.15	0.02	-0.01	-0.06
Mean Method 2	0.07	-0.03	0.03	-0.04	-0.01	0.08	-0.08	0.05	0.04	-0.10
Mean Method 3	0.11	-0.04	-0.06	0.01	-0.03	-0.02	-0.06	0.03	0.05	-0.10
Mean Method 4	0.06	-0.07	0.00	0.02	0.03	0.08	-0.05	0.03	0.05	-0.02
Mean Method 5	0.03	-0.05	-0.04	0.06	0.02	0.00	-0.01	0.08	0.09	-0.07
Mean Method 6	0.05	-0.05	-0.02	0.06	0.02	0.02	-0.05	0.05	0.09	-0.01
Integration Method	0.04	-0.03	-0.05	0.06	0.02	0.05	-0.07	0.05	0.06	-0.02
Integration Method 1	0.02	-0.02	-0.09	0.05	0.02	0.06	-0.11	0.07	0.06	-0.05
Integration Method 2	0.05	-0.02	0.01	-0.01	-0.01	0.04	-0.07	0.03	0.03	-0.03
Integration Method 3	0.06	-0.05	-0.05	0.03	0.02	-0.02	-0.04	0.06	0.08	-0.05
Integration Method 4	0.08	-0.07	0.00	0.03	0.02	0.09	-0.08	0.03	0.06	-0.03
Integration Method 5	0.01	-0.04	-0.11	0.13	0.04	-0.02	-0.02	0.08	0.09	-0.01
Integration Method 6	0.03	-0.01	-0.04	0.03	-0.01	0.00	-0.08	0.03	0.06	-0.02

7.5 Discussion

As a whole these findings suggested the evidence of no associations between Stop-Signal Task estimations and Iowa Gambling Task net scores. These results are in contrast with previous findings (Noël et al, 2007; Shuster & Toplak, 2009; Verbruggen et al, 2013) that have found positive associations between these two tasks.

Two possible explanations for these results concern the type of task used and the type of study carried out. Indeed, in order to consider the Stop-Signal Task a decision-making task, perhaps more trials are needed than those used in this study (i.e., 384 trials). In their study, in fact, Skippen and colleagues (2019), consider the Stop-Signal Task as a task of difficult choices, a task with more than 700 trials, which became difficult to administer for this project. In addition, it is important to consider that for the study of Skippen and colleagues (2019), participants received a reward of 20 dollars per hours. This also leads to a different motivation in the performance of the task (Gerstein et al, 2004) between paid and nonpaid participants.

A second explanation relates to the fact that the tasks used in previous studies (e.g., Verbruggen et al, 2013) were a combination of the Stop-Signal Task with decision-making tasks, while in the present study the two tasks were used separately to evaluate their possible associations. In fact, in the study of Verbruggen and colleagues (2013), the aim was to investigate whether there was an effect in gambling as a result of an inhibition of response. In this study, however, the two tasks were administered separately and randomly.

7.6 Limitations

The results of the present study should be considered in the light of several important limitations. Firstly, the sample size was quite limited. It is important however to notice that administering neuropsychological tasks requires a great deal of time (i.e., more than one hour per participants), and resources (i.e., all participants were administered the measures in the laboratory) (e.g., Crawford & Garthwaite, 2002). For this reason, my sample can be considered more a convenient study group than a sample representative of the Italian population. Of course, this can limit the generalizability of the findings to other populations (i.e., clinical, or forensic sample).

As I previously mentioned in the discussion, the limited number of trials used in the Stop-Signal Task and the use of two separate tasks might have led to different results from previous studies. Indeed, in this project the Stop-Signal Task was provided with "only" 384 trials. However, it is important to noticed that I had manipulated the structure of the original task (Verbruggen et al, 2019) by adding two additional blocks, that can change the applicability of the Stop-Signal Task to the context of difficult choices (Heathcote et al, 2019; Skippen et al, 2019).

Lastly, a test-retest reliability depends heavily on the length of the retest period (Kaplan & Saccuzzo, 2017; Polit, 2014). Therefore, it is important to consider that underlying processes can change rapidly in a three-month test-retest paradigm, and this can explain poor test-retest reliability

8. Future Work

Based on the results of these studies, future work should replicate these findings in different samples. The possibility of the replication of the results is simplified by the use of the free PEBL platform (see paragraph 2.2.1 in the previous Section). It would also be important to replicate the results of this study in clinical population, in order to extend the generalizability of the findings and to assess the reliability of the task also in different population (e.g., patients with difficulties decision-making).

In addition, future studies should also consider the possibility to use different computerized task measures to assess response inhibition. For example, the convergent validity between the Iowa Gambling Task may be assess with other behavioral tasks. Indeed, previous studies have shown differences between hypothetical and real monetary reward by comparing the two performances in both Iowa Gambling Task and Balloon Analogue Risk Task (e.g., Bowman & Turnbull, 2003; Hu et al, 2019). An additional difference in these two tasks is that during the Balloon Analogue Risk Task participants received immediately “money” from each balloon, whereas during the Iowa Gambling Task participants receive “money” only at the end of the administrations. Future studies should also consider these differences when administrating a task. The same consideration can be done with self-report measures. Indeed, as previously mentioned, in literature are present several assessment measures, more specific for the decision-making (e.g., General Decision-Making Style; Scott & Bruce, 1995). Relying on different measures to assess the convergent validity may yield to different results from the present study.

Based on the results obtained with the temporal stability of the Iowa Gambling Task, that might indicate the presence of underlying processes between the two time of administrations (i.e., possible learning effect), future studies should consider the possibility to assess the performance on the Iowa Gambling Task with advanced cognitive model (i.e., reinforcement learning model, including EV model and PVL model). Indeed, these two advanced computational estimations may explain not only the modest correlation found in this project, but also what types of processes underlying the performance in this task.

Most importantly, future studies should rely on a different paradigm to assess test-retest reliability. Based on Hedge and colleagues' study (2018), between two times of administrations there are numerous variables that can influence the performance and are not considered in temporal stability paradigm (e.g., individual differences and learning effects). According to these colleagues (Hedge et al, 2018), subject variability may influence the reliability of neuropsychological tasks. Indeed, future researchers should rely on alternative ways to assess the longitudinal stability of the disinhibition.

9. General Conclusion

As a whole, the results of this Section on the Iowa Gambling Task suggested that this task was provided with modest relationship with self-report of impulsivity. Specifically, the Iowa Gambling Task estimated showed correlation with the ImpSS and the subscale of Sensation Seeking, as operationalized with the UPPS-P. However, this result confirmed and extended previous results (e.g., Vasconcelos et al, 2014) indicating that individuals with difficulties in planning tend to choose disadvantageous. These findings were consistent also with the associations found between the net scores of the overall performance of the Iowa Gambling Task and the Risk-Taking scale of PID-5, suggesting that this task may provide information about risky choices.

However, no significant correlations were found between the total net score, and the Balloon Analogue Risk Task, that was intended to evaluate risky decision. Nevertheless, these findings were consistent with previous research (Lejuez et al, 2003; Upton et al, 2011), suggesting that these two tasks might measure different behaviors. A possible explanation of these results, as previously mentioned, might be the difference in the timing of the collection of the money: with the Balloon Analogue Risk Task, money is earned after each trial, whereas in the Iowa Gambling Task, money is earned only at the end of the task.

Lastly, these studies extended the present literature about the temporal stability. Indeed, the second study of this Section showed that test-retest reliability assessed with Spearman *r* coefficient and Intraclass Correlation Coefficient (ICC), yielded modest associations between two times of administrations (i.e., r_{mean} = .41). These findings support previous studies (e.g., Buelow & Suhr, 2009; Ernst et al, 2003a; Ernst et al, 2003b) where a learning effect was supposed in the performance of the Iowa Gambling Task. However, my studies showed that relying on a different approach to assess test-retest reliability provided better results. Indeed, relying on Cronbach's alpha and Omega coefficients, to assess the internal stability of the task in two times of administrations, provided good internal consistency in both times.

Thus, the results of this Section suggested that the Iowa Gambling Task may be provided with good reliability, assessed with the internal consistency coefficient, and convergent validity, assessed with self-reports of disinhibition (i.e., Sensation Seeking and Risk-Taking scales). Moreover, these results also suggest the importance to evaluate the psychometric proprieties of a behavioral task, mostly used with clinical application, before the evaluation of more complex and advanced cognitive modelling.

SECTION III: DISCUSSION & CONCLUSION

In my opinion, the overall contribution of the present research project lies in its attempt at examining the reliability and validity of two neuropsychological tasks (i.e., the stop-it task and the Iowa gambling task). In doing this, I tried to consider recent advanced cognitive models for computing stop-signal reaction time in order to establish their contribution in improving the reliability of SSRT estimates. The findings of my studies suggested that relying on advanced cognitive models for assessing the duration of the stop process may be useful in research contexts for advancing our knowledge on the ability to stop ongoing responses that are no longer appropriate (e.g., Matzke et al, 2017). However, when the focus is the study of individual differences in cognitive paradigms, the parametric estimation methods did not seem to represent the best choice due to the computation time and associated convergence problems, at least according to the results of the present studies.

Assessing the reliability of neuropsychological tasks represents a relevant research issue (see, for instance, Elliott et al, 2020), also because the reliability of a measure reduces the correlation that can be observed between the target measure and a theoretically relevant external construct (Nunnally, 1994; Spearman, 1904). Accordingly, low reliability limits the usefulness of the tasks as research and clinical measures (see Hedge et al, 2018). In line with the findings reported here, test-retest paradigm may underestimate the reliability of tasks, whereas considering the blocks composing tasks as test unit may allow for the assessment of internal consistency.

Notably, in 2008 the National Institute of Mental Health launched the Research Domain Criteria (RDoC) as a research framework for studying mental disorders in order to obtain better diagnosis and intervention in the future. In his overview of the project's first decade Cuthberth (2022) underscore that one of the challenges experienced by the RDoC is related to the issues of measurement and psychometrics. Indeed, in their report on tasks and measures, the National Advisory Mental Health Council Workgroup on Tasks and Measures for Research Domain Criteria (2016) underlined that all RDoC domain subgroups encountered a particular challenge: the absence of psychometric data. Accordingly, I think that the results of the present research project might advance our knowledge in the field, providing additional data and further analysis useful to understand the basic psychometric properties of two tasks that could be used for assessing one of the RDoC Cognitive System Domain (i.e., Response Selection, Inhibition/Suppression Subconstruct).

References

Achenbach TM, Edelbrock CS (1979) The child behavior profile: II. Boys aged 12- 16 and girls aged 6-11 and 12-16. *J Consult Clin Psychol* 47: 223–233.

Ahn WY, Busemeyer JR, Wagenmakers EJ, Stout JC (2008) Comparison of decision learning models using the generalization criterion method. *Cogn Sci* 32(8): 1376-1402.

Ahn WY, Krawitz A, Kim W, Busemeyer JR, Brown JW (2011) A model-based fMRI analysis with hierarchical Bayesian parameter estimation. *J Neurosci Psychol Econ* 4: 95–110.

Ahn WY, Vasilev G, Lee SH, Busemeyer JR, Kruschke JK, Bechara A, Vassileva J (2014) Decision-making in stimulant and opiate addicts in protracted abstinence: Evidence from computational modeling with pure users. *Front Psychol* 5: 849.

Aklin WM, Lejuez CW, Zvolensky MJ, Kahler CW, Gwadz M (2005) Evaluation of behavioral measures of risk taking propensity with inner city adolescents. *Behav Res Ther* 43: 215–228.

Alderson RM, Rapport MD, Hudec KL, Sarver DE, Kofler MJ (2010) Competing core processes in attention-deficit/hyperactivity disorder (ADHD): do working memory deficiencies underlie behavioral inhibition deficits? *J Abnorm Child Psychol* 38: 497–507.

Alderson RM, Rapport MD, Sarver DE, Kofler MJ (2008) ADHD and behavioral inhibition: A re-examination of the stop-signal task. *J Abnorm Child Psychol* 36(7): 989-998.

Almasy L, Blangero J (2001) Endophenotypes as quantitative risk factors for psychiatric disease: rationale and study design. *Am J Med Genet* 105: 42–44.

American Psychiatric Association (2013) *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. American Psychiatric Association, Washington, DC.

Anastasi A, Urbina S (1997) *Psychological testing*. Prentice Hall/Pearson Education, New Jersey, USA.

Arnett JJ (1992) Reckless behavior in adolescence: a developmental perspective. *Dev Rev* 12: 339–373.

Arnett JJ (1999) Adolescent storm and stress, reconsidered. *Am Psychol* 54(5): 317.

Aron AR (2007) The neural basis of inhibition in cognitive control. *Neuroscientist* 13: 214–228.

Aron AR, Poldrack RA (2005) The cognitive neuroscience of response inhibition: relevance for genetic research in attention-deficit/hyperactivity disorder. *Biol Psychiatry* 57: 1285–1292.

Aron AR, Poldrack RA (2006) Cortical and subcortical contributions to stop signal response inhibition: Role of the subthalamic nucleus. *J Neurosci* 26: 2424–2433.

Badcock JC, Michie P, Johnson L, Combrinck J (2002) Acts of control in schizophrenia: Dissociating the components of inhibition. *Psychol Med* 32: 287–297.

Band GP, van der Molen MW, Logan GD (2003) Horse-race model simulations of the stop-signal procedure. *Acta Psychol* 112: 105–142.

Band GP, van der Molen MW, Overtoom CC, Verbaten MN (2000) The ability to activate and inhibit speeded responses: Separate developmental trends. *J Exp Child Psychol* 75(4): 263-290.

Bari A, Robbins TW (2013) Inhibition and impulsivity: behavioral and neural basis of response control. *Progr Neurobiol* 108: 44-79.

Barkley RA (1991) The ecological validity of laboratory and analogue assessment methods of ADHD symptoms. *J Abnorm Child Psychol* 19(2): 149-178.

Barkley RA (1997) Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychol Bull* 121: 65–94.

Barratt ES, Patton JH (1983) Impulsivity: cognitive, behavioral, and psychophysiological correlates. In *Biological Basis of Sensation Seeking, Impulsivity, and Anxiety*, Zuckerman M (ed) pp 77–116. Hillsdale, NJ: Erlbaum.

Bartzokis G, Lu P, Beckson M, Rapoport B, Grant S, Wiseman E, Londond E (2000) Abstinence from cocaine reduces high-risk responses on a gambling task. *Neuropsychopharmacology* 22: 102–103.

Bechara A (2005) Decision making, impulse control and loss of will- power to resist drugs: A neurocognitive perspective. *Nat Neurosci* 8: 1458–1463.

Bechara A (2008) *Iowa gambling task professional manual*. Psychological Assessment Resources, Lutz, FL.

Bechara A, Damasio AR, Damasio H, Anderson SW (1994) Insensitivity to future consequences following damage to human prefrontal cortex. *Cogn* 50(1-3): 7-15.

Bechara A, Damasio H, Damasio AR (2000) Emotion, decision making and the orbitofrontal cortex. *Cereb Cortex* 10: 295–307.

Bechara A, Damasio H, Damasio AR, Lee GP (1999) Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *J Neurosci* 19: 5473–5481.

Bechara A, Damasio H, Tranel D, Anderson SW (1998) Dissociation of working memory from decision making within the human prefrontal cortex. *J Neurosci* 18: 428–437.

Bechara A, Damasio H, Tranel D, Damasio AR (1997) Deciding advantageously before knowing the advantageous strategy. *Science* 275: 1293–1295.

Bechara A, Tranel D, Damasio H (2000) Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain* 123: 2189-2202.

Becser N, Sand T, Zwart JA (1998) Reliability of cephalic thermal thresholds in healthy subjects. *Cephalalgia* 18(8): 574-582.

Bedard AC, Nichols S, Barbosa JA, Schachar R, Logan GD, Tannock R (2002) The development of selective inhibitory control across the life span. *Dev Neuropsychol* 21(1): 93-111.

Belin TR, Rubin DB (1995) The analysis of repeated-measures data on schizophrenic reaction times using mixture models. *Statist Med* 14: 747–768.

Beninger RJ, Wasserman J, Zanibbi K, Charbonneau D, Mangels J, Beninger BV (2003) Typical and atypical antipsychotic medications differentially affect two nondeclarative memory tasks in schizophrenic patients: a double dissociation. *Schizophr Res* 61: 281–292.

Bentler PM (2009) Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika* 74: 137–143.

Berlin GS, Hollander E (2014) Compulsivity, impulsivity, and the DSM-5 process. *CNS Spectr* 19(1): 62-68.

Betsch C, Kunz JJ (2008) Individual strategy preferences and decisional fit. *J Behav Decis Mak* 21: 532–555.

Beyth-Marom R, Austin L, Fischho VB, Palmgren C, Quadrel MJ (1993) Perceived consequences of risky behaviors: adults and adolescents. *Dev Psychol* 29: 549–563.

Beyth-Marom R, FischhoVB (1997) Adolescents decisions about risks: a cognitive perspective. In *Health Risks and Developmental Transitions in Adolescence*, Schulenberg J, Maggs JL (ed) pp 110-135. New York, NY: Cambridge University Press.

Bezdjian S, Baker LA, Lozano DI, Raine A (2009) Assessing inattention and impulsivity in children during the Go/NoGo task. *Br J Dev Psychol* 27:365-383.

Bishara AJ, Pleskac TJ, Fridberg DJ, Yechiam E, Lucas J, Busemeyer JR, Finn PR, Stout JC (2009) Similar processes despite divergent behavior in two commonly used measures of risky decision making. *J Behav Decis Mak* 22: 435– 454

Bissett PG, Hagen MP, Jones HM, Poldrack RA (2021) Design issues and solutions for stop-signal data from the Adolescent Brain Cognitive Development (ABCD) study. *Elife* 10: e60185.

Bissett PG, Logan GD (2011) Balancing cognitive demands: Control adjustments in the stop-signal paradigm. *J Exp Psychol Learn Mem Cogn* 37*:* 392–404.

Block J (2002) *Personality as an Affect-Processing System: Toward an Integrative Theory.* Lawrence Erlbaum, Mahwah, NJ.

Boldini A, Russo R, Avons SE (2004) One process is not enough! A speed-accuracy tradeoff study of recognition memory. *Psychon Bull Rev* 11: 353–361.

Bolla KI, Eldreth DA, Matochik JA, Cadet JL (2004) Sex-related differences in a gambling task and its neurological correlates. *Cereb Cortex* 14(11): 1226-1232.

Boucher L, Palmeri TJ, Logan GD, Schall JD (2007) Inhibitory control in mind and brain: an interactive race model of countermanding saccades. *Psychol Rev* 114(2):376.

Bowman CH, Turnbull OH (2003) Real versus facsimile reinforcers on the Iowa Gambling Task. *Brain Cogn 53*(2): 207-210.

Boyer TW (2006) The development of risk-taking: A multi-perspective review. *Dev Rev* 26(3): 291-345.

Brand M, Recknor EC, 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. *J Clin Exp Neuropsychol* 29: 86 –99.

Breese BB (1899) On inhibition. *Psychol Monogr* 3: 1–65.

Brown WS, Anderson LB, Symington MF, Paul LK (2012) Decision-making in individuals with agenesis of the corpus callosum: expectancy-valence in the Iowa gambling task. *Arch Clin Neuropsychol* 27(5): 532–544.

Bruner JS (1957) Neural mechanisms in perception. *Psychol Rev* 64: 340–358.

Buelow MT, Blaine AL (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. *Psychol Assess* 27(3): 777.

Buelow MT, Suhr JA (2009) Construct validity of the Iowa gambling task. *Neuropsychol Rev* 19(1): 102-114.

Busemeyer JR, Myung IJ (1992) An adaptive approach to human decision making: Learning theory, decision theory, and human performance. *J Exp Psychol Gen* 121: 177–194.

Busemeyer JR, Stout J, Finn P (2003) Using computational models to help explain decision making processes of substance abusers. In *Cognitive and affective neuroscience of psychopathology*, Barch D (ed) pp 1-41. New York, NY: Oxford University Press.

Busemeyer JR, Stout JC (2002) A contribution of cognitive decision models to clinical assessment: decomposing performance on the Bechara gambling task. *Psychol Assess* 14(3): 253.

Byrnes JP (1998) *The nature and development of decision making: a self-regulation model*. Lawrence Erlbaum, Mahwah, NJ.

Carlotta D, Borroni S, Maffei C, Fossati A (2011) The role of impulsivity, sensation seeking and aggression in the relationship between childhood AD/HD symptom and antisocial behavior in adolescence. *Neurol Psychiatry Brain Res* 17(4): 89-98.

Carlson AA (2008) The role of impulsivity and compulsivity in disordered eating, self-harm, and obligatory exercise in a nonclinical sample. *Diss Abstr Int* 69(3-B): 1944.

Carpenter RH, Williams MLL (1995) Neural computation of log likelihood in control of saccadic eye movements. *Nature* 377(6544): 59-62.

Carpenter RHS (1981) Oculomotor Procrastination. In *Eye movements: Cognition and visual perception*, Routledgs (ed) pp 237-246. Mahwah, NJ: Erlbaum.

Carver CS, White TL (1994) Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *J Pers Soc Psychol* 67(2): 319–333.

Casey BJ, Trainor RJ, Orendi JL, Schubert AB, Nystrom LE, Giedd JN, Castellanos FX, Haxby JV, Noll DC, Cohen JD *et al* (1997) A developmental functional MRI study of prefrontal activation during performance of a go-no-go task. *J Cogn Neurosci* 9(6): 835-847.

Castellanos FX, Sonuga-Barke EJ, Milham MP, Tannock R (2006) Characterizing cognition in ADHD: beyond executive dysfunction. *Trends Cogn Sci* 10: 117–123.

Cavallaro R, Cavedini P, Mistretta P, Bassi T, Angelone SM, Ubbiali A, Bellodi L (2003) Basal-corticofrontal circuits in schizophrenia and obsessive-compulsive disorder: a controlled, double dissociation study. *Biol Psychiatry* 54: 437–443.

Chamberlain SR, Blackwell AD, Fineberg NA, Robbins TW, Sahakian BJ (2005) The neuropsychology of obsessive compulsive disorder: the importance of failures in cognitive and behavioural inhibition as candidate endophenotypic markers. *Neurosci Biobehav* 29: 399–419.

Chambers CD, Bellgrove MA, Stokes MG, Henderson TR, Garavan H, Robertson IH, Morris AP, Mattingley JB (2006) Executive ''brake failure'' following deactivation of human frontal lobe. *J Cogn Neurosci* 18: 444– 455.

Cheng AS, Lee HC (2012) Risk-taking behavior and response inhibition of commuter motorcyclists with different levels of impulsivity. *Transp Res F: Traffic Psychol Behav* 15(5): 535-543.

Chevalier N, Chatham CH, Munakata Y (2014) The practice of going helps children to stop: The importance of context monitoring in inhibitory control. *J Exp Psychol Gen* 143: 959 –965.

Chico E (2000) Intensidad emocional y su relación con extraversión y neuroticismo. *Psicothema* 12: 568-573.

Chmielewski M, Watson D (2009) What is being assessed and why it matters: The impact of transient error on trait research. J Pers Soc Psychol 97: 186–202.

Cinaz B, Vogt C, Arnrich B, Tröster G (2012) Implementation and evaluation of wearable reaction time tests. *Pervasive Mob Comput* 8(6): 813-821.

Clark DG, Kar J (2011) Bias of quantifier scope interpretation is attenuated in normal aging and semantic dementia. *J Neurolinguistics* 24(4): 401-419.

Cohen J (1988) *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen JD (2017) Cognitive control: Core constructs and current considerations. In *The Wiley handbook of cognitive control*, Egner T (ed) pp 3–28. Hoboken, USA: Wiley-Blackwell.

Cohen JR, Poldrack RA (2008) Automaticity in motor sequence learning does not impair response inhibition. *Psychon Bull Rev* 15: 108–115.

Collerton J, Collerton D, Arai Y, Barrass K, Eccles M, Jagger C, Newcastle 85+ Study Core Team (2007) A comparison of computerized and pencil-and-paper tasks in assessing cognitive function in community-dwelling older people in the Newcastle 85+ Pilot Study. *J Am Geriatr Soc* 55: 1630-1635.

Colonius H (1990) A note on the stop-signal paradigm, or how to observe the unobservable. *Psychol Rev* 97(2): 309.

Colonius H, Özyurt J, Arndt PA (2001) Countermanding saccades with auditory stop signals: testing the race model. *Vision Res* 41(15): 1951-1968.

Congdon E, Mumford JA, Cohen JR, Galvan A, Canli T, Poldrack RA (2012) Measurement and reliability of response inhibition. *Front Psychol* 3: 37.

Cortina JM (1993) What is coefficient alpha? An examination of theory and applications. *J Appl Psychol* 78: 98–104.

Crawford JR, Garthwaite PH (2002) Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia* 40(8): 1196-1208.

Crone EA, van der Molen MW (2004) Developmental changes in real life decision making: Performance on a gambling task previously shown to depend on the ventromedial prefrontal cortex. *Dev Neuropsychol* 25(3): 251-279.

Cuthbert BN (2022) Research Domain Criteria (RDoC): Progress and Potential. *Curr Dir Psychol Sci* 31(2): 107-114.

Cyders MA, Coskunpinar A (2011) Measurement of constructs using self-report and behavioral lab tasks: is there overlap in nomothetic span and construct representation for impulsivity? *Clin Psychol Rev* 31(6): 965-982.

Cyders MA, Coskunpinar A (2012) The relationship between self-report and lab task conceptualizations of impulsivity. *J Res Pers* 46(1): 121-124.

Cyders MA, Smith GT (2007) Mood-based rash action and its components: Positive and negative urgency. *Pers Individ Differ* 43:839-850.

Cyders MA, Smith GT (2007) Mood-based rash action and its components: Positive and negative urgency. *Pers Individ Differ* 43(4): 839-850.

d'Acremont M, Van der Linden M (2006) Gender differences in two decision-making tasks in a community sample of adolescents. *Int J Behav Dev* 30(4): 352-358.

Damasio AR (1994) *Descartes' error: Emotion, reason, and the human brain*. Avon, New York, USA.

Danckert J, Stöttinger E, Quehl N, Anderson B (2012) Right hemisphere brain damage impairs strategy updating. *Cereb Cortex* 22(12): 2745-2760.

de Jong R, Coles MG, Logan GD, Gratton G (1990) In search of the point of no return: The control of response processes. *J Exp Psychol Hum Percept Perform* 16: 164–182.

De Leeuw JR (2015) jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behav Res Methods* 47(1): 1-12.

de Visser L, Homberg JR, Mitsogiannis M, Zeeb FD, Rivalan M, Fitoussi A, Galhardo V, van den Bos R, Winstanley CA, Dellu-Hagedorn F (2011) Rodent versions of the iowa gambling task: opportunities and challenges for the understanding of decision-making. *Front Neurosci* 5: 109.

De Wilde B, Bechara A, Sabbe B, Hulstijn W, Dom G (2013a) Risky decision- making but not delay discounting improves during inpatient treatment of polysubstance dependent alcoholics. *Front Psychiatry* 4: 91.

De Wilde B, Verdejo-García A, Sabbe B, Hulstijn W, Dom G (2013b) Affective decision-making is predictive of three-month relapse in polysubstance- dependent alcoholics. *Eur Addict Res* 19(1): 21–28.

De Young CG (2010) Impulsivity as a personality trait. In *Handbook of Self-regulation: Research, Theory and Applications*, Vohs KD, Baumeister RF (ed) pp 485–502. New York: Guilford Press.

Dempster FN, Corkill AJ (1999) Individual differences in susceptibility to interference and general cognitive ability. *Acta Psychol* 101: 395–416.

Dickman S (1985) Impulsivity and perception: individual differences in the process sing of the local and global dimensions of stimuli. *J Pers Soc Psychol* 48: 133–149.

Dickman SJ (1990) Functional and dysfunctional impulsivity: personality and cognitive correlates. *J Pers Soc Psychol* 58: 95–102.

Dickman SJ, Meyer DE (1988) Impulsivity and speed-accuracy tradeoffs in information processing. *J Pers Soc Psychol* 54: 274–290.

Dijkstra AFJ, Timmermans MPH, Schriefers HJ (2000) On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *J Mem Lang* 42: 445–464.

Donders FC (1969) On the speed of mental processes. *Acta Psychol* 30: 412–431.

Donovan JE, Jessor R (1985) Structure of problem behavior in adolescence and young adulthood. *J Consult Clin Psychol* 53: 890–904.

Donovan JE, Jessor R, Costa FM (1988) Syndrome of problem behavior in adolescence: a replication. *J Consult Clin Psychol* 56: 762–765.

Dumais A, Potvin S, Joyal C, Allaire JF, Stip E, Lesage A, Gobbi G, Cote G (2011) Schizophrenia and serious violence: a clinical-profile analysis incorporating impulsivity and substance-use disorders. *Schizophr Res* 130: 234–237.

Dunn BD, Dalgleish T, Lawrence AD (2006) The somatic marker hypothesis: A critical evaluation. *Neurosci Biobehav Rev* 30(2): 239-271.

Dunn BD, Dalgleish T, Lawrence AD (2006) The somatic marker hypothesis: A critical evaluation. *Neurosci Biobehav Rev* 30: 239 –271.

Dunn TJ, Baguley T, Brunsden V (2014) From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *Br J Psychol* 105: 399–412.

Dupuis A, Indralingam M, Chevrier A, Crosbie J, Arnold P, Burton CL, Schachar R (2019) Response time adjustment in the stop signal task: Development in children and adolescents. *Child Dev* 90(2): e263-e272.

Eagle DM, Bari A, Robbins TW (2008) The neuropsychopharmacology of action inhibition: cross-species translation of the stop-signal and go/no-go tasks. *Psychopharmacology* 199: 439–456.

Eagle DM, Bari A, Robbins TW (2008) The neurosychopharmacology of action inhibition: cross-species translation of the stop-signal and go/no-go tasks. *Psychopharmacology (Berl)* 199(3): 439–56.

Eagle DM, Lehmann O, Theobald DE, Pena Y, Zakaria R, Ghosh R, Dalley JW, Robbins TW (2009) Serotonin depletion impairs waiting but not stop-signal reaction time in rats: implications for theories of the role of 5-HT in behavioral inhibition. *Neuropsychopharmacology* 34(5): 1311-1321.

Elandt-Johnson RC, Johnson NL (1980) *Survival models and data analysis*. Wiley, New York, USA.

Elliott ML, Knodt AR, Ireland D, Morris ML, Poulton R, Ramrakha S, Sison ML, Moffitt TE, Caspi A, Hariri AR (2020) What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. *Psychol Sci* 31(7): 792-806.

Elliott ML, Knodt AR, Caspi A, Moffitt TE, Hariri AR (2021) Need for psychometric theory in neuroscience research and training: Reply to kragel et al. (2021). *Psychol Sci* 32(4): 627-629.

Enkavi AZ, Eisenberg IW, Bissett PG, Mazza GL, MacKinnon DP, Marsch LA, Poldrack RA (2019) Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proc Natl Acad Sci USA* 116(12): 5472-5477.

Epstein S (1994) Integration of the cognitive and the psychodynamic unconscious. *Am Psychol* 49: 709–724.

Erev I, Roth AE (1998) Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *Am Econ Rev* 88: 848–881.

Ernst M, Grant SJ, London ED, Contoreggi CS, Kimes AS, Spurgeon L (2003a). Decision making in adolescents with behavior disorders and adults with substance abuse. *Am J Psychiatry* 160: 33–40.

Ernst M, Kimes AL, London ED, Matochik JA, Eldreth D, Tata S, Contoreggi C, Leff M, Bolla K (2003b). Neural substrates of decision making in adults with attention deficit hyperactivity disorder. *Am J Psychiatry* 160: 1061–1070.

Evans CEY, Bowman CH, Turnbull OH (2005) Subjective awareness on the Iowa Gambling Task. The key role of emotional experience in schizophrenia. J *Clin Exp Neuropsychol* 27: 656–664.

Evans JSBT (2008) Dual-processing accounts of reasoning, judgment, and social cognition. *Annu Rev Psychol* 59: 255–278.

Evans JSBT (2010) Intuition and reasoning: A dual-process perspective. *Psychol Inq* 21: 313–326.

Evans JSBT, Stanovich KE (2013) Dual-process theories of higher cognition: Advancing the debate. *Perspect Psychol Sci* 8: 223–241.

Evenden JL (1999) Varieties of impulsivity. *Psychopharmacology* 146(4): 348-361.

Eysenck HJ (1993) The nature of impulsivity. In *The Impulsive Client: Theory, Research, and Treatment*, McCown WG, Johnson JL, Sure MB (ed) pp 57–69. Washington DC: American Psychological Association.

Eysenck HJ, Eysenck MW (1985) *Pers Individ Differ: a natural science approach*. Plenum, New York.

Eysenck HJ, Eysenck SBG (1975) *Manual of the Eysenck Personality Questionnaire*. Hodder & Stoughton, London.

Eysenck SB, Eysenck HJ (1977) The place of impulsiveness in a dimensional system of personality description. ***Br J Soc Clin Psychol*** 16: 57-68.

Farrell S, Ludwig CJH (2008) Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychon Bull Rev* 15: 1209–1217.

Fernie G, Cole JC, Goudie AJ, Field M (2010) Risk-taking but not response inhibition or delay discounting predict alcohol consumption in social drinkers. *Drug Alcohol Depend* 112: 54–61.

Ferrier D (1876) *The functions of the brain*. Smith Elder, London.

Fossati A, Di Ceglie A, Acquarini E, Barratt ES (2001) Psychometric properties of an Italian version of the Barratt Impulsiveness Scale-11 (BIS-11) in nonclinical subjects. *J Clin Psychol* 57(6): 815-828.

Fossati A, Krueger RF, Markon KE, Borroni S, Maffei C (2013) Reliability and validity of the Personality Inventory for DSM-5 (PID-5) predicting DSM-IV personality disorders and psychopathy in community-dwelling Italian adults. *Assessment* 20(6): 689-708.

Fossati A, Somma A, Borroni S, Markon KE, Krueger RF (2018) Executive functioning correlates of DSM-5 maladaptive personality traits: Initial evidence from an Italian sample of consecutively admitted adult outpatients. *J Psychopathol Behav Assess* 40(3): 484-496.

Fossati A, Somma A, Karyadi KA, Cyders MA, Bortolla R, Borroni S (2016) Reliability and validity of the Italian translation of the UPPS-P Impulsive Behavior Scale in a sample of consecutively admitted psychotherapy patients. *Pers Individ Differ* 91: 1-6.

Fridberg DJ, Queller S, Ahn WY, Kim W, Bishara AJ, Busemeyer JR, Stout JC (2010) Cognitive mechanisms underlying risky decision-making in chronic cannabis users. *J Math Psychol* 54: 28–38.

Friedman NP, Miyake A, Young SE, Defries JC, Corley RP, Hewitt JK (2008) Individual differences in executive functions are almost entirely genetic in origin. *J Exp Psychol Gen* 137: 201–225.

Furby L, Beyth-Marom R (1992) Risk taking in adolescence: a decision-making perspective. *Dev Rev* 12: 1–44.

Fuster JM (2008) *The Prefrontal Cortex*. Academic Press, London.

Gall FJ (1835) *On the Functions of the Brain and on Each of its Parts*. Marsh, Capen and Lyon, Boston, USA.

Gamerman D, Lopes HF (2006) *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference.* Chapman & Hall/CRC, Boca Raton, Florida, USA.

García de la Banda G, Martínez Abascal MA, Riesco M, Pérez G (2004) La respuesta al cortisol ante un examen y su relación con otros acontecimientos estresantes y con algunas características de personalidad. *Psicothema* 16: 294-298.

Gelman A, Hill J (2007) *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, Cambridge, England.

Gelman A, Meng X, Stern H (1996) Posterior predictive assessment of model fitness via realized discrepancies. *Stat Sin* 6: 733– 807.

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* 7(4): 457-472.

Gerstein LH, Wilkeson DA, Anderson H (2004) Differences in motivations of paid versus nonpaid volunteers. *Psychol Rep* 94(1): 163-175.

Gialdi G (2022) Hierachical results stop-signal task. Available at: osf.io/w5fpt

Gialdi G (2022) PhD Thesis. Available at: https://osf.io/tudmw/

Gialdi G, Somma A, Borroni S, Fossati A (2021) Factor Structure, Measurement Invariance Across Gender Sub-Groups, and Normative Data for the Italian Translation of the UPPS-P Impulsive Behavior Scale in Italian Community-Dwelling Adults. *Mediterr J Clin Psychol* 9(2).

Gialdi G, Somma A, Manara CV, Fossati A (2020) Alternative models of estimating the Stop-Signal Reaction Time in the Stop-Signal Paradigm and their differential associations with self-reports of impulsivity domains. *BPA Appl Psychol Bull* 68(287).

Gilks WR, Richardson S, Spiegelhalter DJ (1996) *Markov chain Monte Carlo in practice*. Chapman & Hall/CRC, Boca Raton, Florida, USA.

Gillespie SM, Lee J, Williams R, Jones A (2022) Psychopathy and response inhibition: A meta-analysis of go/no-go and stop signal task performance. *Neurosci Biobehav Rev* 142: 104868.

Glöckner A, Betsch T (2008) Modeling option and strategy choices with connectionist networks: Towards an integrative model of automatic and deliberate decision making. *Judgm Decis Mak* 3: 215–228.

Godek J, Murray KB (2008) Willingness to pay for advice: The role of rational and experiential processing. *Organ Behav Hum Decis Process* 106: 77– 87.

Goel V, Buchel C, Frith C, Dolan RJ (2000) Dissociation of mechanisms underlying syllogistic reasoning. *NeuroImag,* 12: 504–514.

Goel V, Dolan RJ (2003) Explaining modulation of reasoning by belief. *Cognition*, 87: B11–B22.

Goldberg MC, Courtney S, Mostofsky SH, Abrams MT, Arnold S, Kaufmann WE, Denckla MB, Pekar JJ (2001) Hybrid block/ event-related paradigm for fMRI of a go/no-go task. *Proc Int Soc Magn Reson Med* 9: 1294.

Gordon B, Caramazza A. (1982) Lexical decision for open- and closed-class words: Failure to replicate differential frequency sensitivity. *Brain Lang* 15: 143–160.

Gottfredson MR, Hirschi T (1990) *A general theory of crime*. Stanford University Press, Palo Alto, California, USA.

Goudriaan AE, Grekin ER, Sher KJ (2011) Decision making and response inhibition as predictors of heavy alcohol use: a prospective study. *Alcohol Clin Exp Res* 35(6): 1050-1057.

Grant S, Contoreggi C, London ED (2000) Drug abusers show impaired performance in a laboratory test of decision making. *Neuropsychology* 38: 1180–1187.

Gray JA (1972) *The Psychophysiological Basis of Introversion-Extraversion: a Modification of Eysenck's Theory*. Academic Press, New York, USA.

Guillaume S, Jollant F, Jaussent I, Lawrence N, Malafosse A, Courtet P (2009) Somatic markers and explicit knowledge are both involved in decision-making. *Neuropsychology* 47: 2120–2124.

Gullo MJ, Stieger AA (2011) Anticipatory stress restores decision-making deficits in heavy drinkers by increasing sensitivity to losses. *Drug Alcohol Depend* 117(2-3): 204-210.

Gullone E, Moore S, Moss S, Boyd C (2000) The adolescent risk-taking questionnaire: development and psychometric evaluation. *J Adolesc Res* 15: 231–250.

Gut-Fayand A, Dervaux A, Olie JP, Loo H, Poirier MF, Krebs MO (2001) Substance abuse and suicidality in schizophrenia: a common risk factor linked to impulsivity. *Psychiatry Res* 102: 65–72.

Guthrie ER (1930) Conditioning as a principle of learning. *Psychol Rev* 37: 412–428.

Hallett PE (1978) Primary and secondary saccades to goals defined by instructions. *Vision Res* 18(10): 1279-1296.

Halperin JM, Wolf L, Greenblatt ER, Young G (1991) Subtype analysis of commission errors on the continuous performance test in children. *Dev Neuropsychol* 7(2): 207-217.

Hanes DP, Carpenter RH (1999) Countermanding saccades in humans. *Vision Res* 39(16): 2777-2791.

Hanes DP, Patterson WF, Schall JD (1998) Role of frontal eye fields in countermanding saccades: visual, movement, and fixation activity. *J Neurophysiol* 79(2): 817-834.

Hanes DP, Schall JD (1995) Countermanding saccades in macaque. *Vis Neurosci* 12: 929–937.

Harnishfeger KK, Bjorklund DF (1994) A developmental perpective on individual differences in inhibition. *Learn Individ Differ* 6: 331–355.

Heathcote A, Lin YS, Reynolds A, Strickland L, Gretton M, Matzke D (2019) Dynamic models of choice. *Behav Res Methods* 51(2): 961-985.

Heathcote A, Popiel SJ, Mewhort DJ (1991) Analysis of response time distributions: An example using the Stroop task. *Psychol Bull* 109: 340–347

Hedge C, Powell G, Sumner P (2018) The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behav Res methods* 50(3): 1166-1186.

Hendry A, Greenhalgh I, Bailey R, Fiske A, Dvergsdal H, Holmboe K (2022) Development of directed global inhibition, competitive inhibition and behavioural inhibition during the transition between infancy and toddlerhood. *Dev Sci* 25(5): e13193.

Hino Y, Lupker SJ (2000) The effects of word frequency and spelling-to-sound regularity in naming with and without preceding lexical decision. *J Exp Psychol Hum Percept Perform* 26: 166–183.

Hinslie L, Shatzky J (1940) *Psychiatric Dictionary*. Oxford University Press, New York, USA.

Hockley WE (1982) Retrieval processes in continuous recognition. *J Exp Psychol Learn Mem Cogn* 8: 497–512.

Hockley WE (1984) Analysis of response time distributions in the study of cognitive processes. *J Exp Psychol Learn Mem Cogn* 10: 598–615.

Hollander E (1998) Treatment of obsessive-compulsive spectrum disorders with SSRIs. *Br J Psychiatry* 173(35): 7-12

Hollander E, Rosen J (2000) Impulsivity. *Journal of Psychopharmacology* 14(2): S39-S44.

Hopko DR, Lejuez CW, Daughters SB, Aklin WM, Osborne A, Simmons BL, Strong D R (2006) Construct validity of the balloon analogue risk task (BART): Relationship with MDMA use by inner-city drug users in residential treatment. *J Psychopathol Behav Assess* 28: 95–101.

Hutton SB, Ettinger U (2006) The antisaccade task as a research tool in psychopathology: a critical review. *Psychophysiology* 43: 302–313.

Irwin CE Jr (1993) Adolescence and risk taking: How are they related? In *Adolescent risk taking*, Bell NJ, Bell RW (ed) pp 7-28. Newbury Park, CA, US: Sage Publications.

James W (1890) *The principles of psychology*. Holt and Company, New York, USA.

Jennings JR, van der Molen MW, Pelham W, Debski KB, Hoza B (1997) Inhibition in boys with attention deficit hyperactivity disorder as indexed by heart rate change. *Dev Psychol* 33(2): 308.

Jentsch JD, Taylor JR (1999) Impulsivity resulting from frontostriatal dysfunction in drug abuse: implications for the control of behavior by reward-related stimuli. *Psychopharmacology* 146: 373–390.

Jessor R (1991) Risk behavior in adolescence: a psychosocial framework for understanding and action. *J Adolesc Health* 12: 597–605.

John OP, Caspi A, Robins RW, Moffitt TE, Stouthamer-Loeber R (1994) The little five: exploring the nomological network of the five-factor model of personality in adolescent boys. *Child Dev* 65: 160–178.

Johnstone SJ, Dimoska A, Smith JL, Barry RJ, Pleffer CB, Chiswick D, Clarke AR (2007) The development of stop-signal and Go/Nogo response inhibition in children aged 7–12 years: performance and event-related potential indices. *Int J Psychophysiol* 63(1): 25-38.

Kahneman D, Frederick S (2002) Representativeness revisited: At- tribute substitution in intuitive judgment. In *Heuristics and biases: The psychology of intuitive judgment*, Gilovich T, Griffin D, Kahneman D (ed) pp. 49–81. New York: Cambridge University Press.

Kalinowska-Łyszczarz A, Pawlak MA, Michalak S, Losy J (2012) Cognitive deficit is related to immune-cell beta-NGF in multiple sclerosis patients. *J Neurol Sci* 321(1-2): 43-48.

Kaplan RM, Saccuzzo DP (2017) *Psychological testing: Principles, applications, and issues*. Pacific Grove, CA: Brooks.

Kasar M, Gleichgerrcht E, Keskinkilic C, Tabo A, Manes FF (2010) Decision- making in people who relapsed to driving under the influence of alcohol. *Alcohol Clin Exp Res* 34(12): 2162–2168.

Kendall PC, Wilcox LE (1979) Self-control in children: development of a rating scale. *J Consult Clin Psychol* 47: 1020–1029.

Kester HM, Sevy S, Yechiam E, Burdick K, Cervellione KL, Kumra S (2006) Decision-making impairments in adolescents with early-onset schizophrenia. *Schizophr Res* 85: 113–123.

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.

Kieres AK, Hausknecht KA, Farrar AM, Acheson A, de Wit H, Richards JB (2004) Effects of morphine and naltrexone on impulsive decision making in rats. *Psychopharmacology* 173: 167-174.

Kindlon D, Mezzacappa E, Earls F (1995) Psychometric properties of impulsivity measures: Temporal stability, validity and factor structure. *J Child Psychol Psychiatry* 36(4): 645-661.

King JA, Colla M, Brass M, Heuser I, von Cramon DY (2007) Inefficient cognitive control in adult ADHD: evidence from trial-by-trial Stroop test and cued task switching performance. *Behav Brain Funct* 3(1): 1-19.

Kinnunen SP, Windmann S (2013) Dual-processing altruism. *Front Psychol* 4: 193.

Kjome KL, Lane SD, Schmitz JM, Green C, Ma L, Prasla I, Swann AC, Moeller FG (2010) Relationship between impulsivity and decision making in cocaine dependence. *Psychiatry Res* 178(2): 299-304.

Kok A (1999) Varieties of inhibition: manifestations in cognition, event-related potentials and aging. *Acta Psychol* 101(2-3): 129-158.

Kramer AF, Humphrey DG, Larish JF, Logan GD (1994) Aging and inhibition: beyond a unitary view of inhibitory processing in attention. *Psychol Aging* 9(4): 491.

Krueger RF, Derringer J, Markon KE, Watson D, Skodol AE (2012) Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychol Med* 42: 1879-1890.

Krueger RF, Derringer J, Markon KE, Watson D, Skodol AE (2012) Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychol Med* 42(9): 1879-1890.

Kuntsi J, Andreou P, Ma J, Börger NA, van der Meere JJ (2005) Testing assumptions for endophenotype studies in ADHD: reliability and validity of tasks in a general population sample. *BMC Psychiatry* 5(1): 1-11.

Laird RD, Pettit GS, Bates JE, Dodge KA (2003a) Parents' monitoring-relevant knowledge and adolescents' delinquent behavior: evidence of correlated developmental changes and reciprocal influences. *Child Dev* 74: 752–768.

Lamers CTJ, Bechara A, Rizzo M, Ramaekers JG (2006) Cognitive function and mood in MDMA/THC users, THC users and non-drug using controls. *Psychopharmacology* 20(2): 302–311.

Lane SD, Cherek DR, Rhoades HM, Pietras CJ, Tcheremissine OV (2003) Relationships among laboratory and psychometric measures of impulsivity: Implications in substance abuse and dependence. *Addict Disord Their Treat* 2(2): 33-40.

Lappin JS, Eriksen CW (1966) Use of a delayed signal to stop a visual reaction-time response. *J Exp Psychol* 72: 805–811.

Lauriola M, Levin IP (2001) Personality traits and risky decision-making in a controlled experimental task: An exploratory study. *Pers Individ Differ* 31(2): 215-22.

Lawrence A, Clark L, Labuzetta JN, Sahakian B, Vyakarnum S (2008) The innovative brain. *Nature* 456: 168–169.

Lee MD (2011) How cognitive modeling can benefit from hierarchical Bayesian models. *J Math Psychol* 55: 1–7.

Lee MD, Wagenmakers EJ (2014) *Bayesian cognitive modeling: A practical course*. Cambridge university press, Cambridge, UK.

Lejuez CW, Aklin WM, Jones HA, Richards JB, Strong DR, Kahler CW, Read JP (2003a) The Balloon Analogue Risk Task (BART) differentiates smokers and nonsmokers. *Exp Clin Psychopharmacol* 11: 26–33.

Lejuez CW, Aklin WM, Zvolensky MJ, Pedulla CM (2003b) Evaluation of the Balloon Analogue Risk Task (BART) as a predictor of adolescent real-world risk-taking behaviours. *J Adolesc* 26: 475–479.

Lejuez CW, Read JP, Kahler CW, Richards JB, Ramsey SE, Stuart GL, Strong DR, Brown RA (2002) Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *J Exp Psychol Appl* 8: 75–84.

Leth-Steensen C, King Elbaz Z, Douglas VI (2000) Mean response times, variability, and skew in the responding of ADHD children: A response time distributional approach. *Acta Psychol* 104: 167– 190.

Levin IP, Hart SS (2003) Risk preferences in young children: early evidence of individual differences in reaction to potential gains and losses. *J Behav Decis Mak* 16: 397-413.

Lieberman MD, Gaunt R, Gilbert DT, Trope Y (2002) Reflexion and reflection: A social cognitive neuroscience approach to attributional inference. In *Advances in experimental social* psychology, Zanna MP (ed) pp 199–249. San Diego: Academic Press.

Lieberman MD, Jarcho JM, Satpute AB (2004) Evidence-based and intuition-based self-knowledge: An FMRI study. *J Pers Soc Psychol* 87: 421–435.

Lighthall NR, Mather M, Gorlick MA (2009) Acute stress increases sex differences in risk seeking in the balloon analogue risk task. *PLoS One* 4(7): e6002.

Lighthall NR, Sakaki M, Vasunilashorn S, Nga L, Somayajula S, Chen EY, Samii N, Mather M (2012) Gender differences in reward-related decision processing under stress. *Soc Cogn Affect* 7(4): 476-484.

Lijffijt M, Bekker EM, Quik EH, Bakker J, Kenemans JL, Verbaten MN (2004) Differences between low and high trait impulsivity are not associated with differences in inhibitory motor control. *J Atten Disord* 8(1): 25-32.

Lijffijt M, Kenemans JL, Verbaten MN, van Engeland H (2005) A meta-analytic review of stopping performance in attention-deficit/hyperactivity disorder: deficient inhibitory motor control? *J Abnorm Psychol* 114(2): 216.

Lin CH, Song TJ, Chen YY, Lee WK, Chiu Y (2013) Reexamining the validity and reliability of the clinical version of the Iowa gambling task: evidence from a normal subject group. *Front Psychol* 4: 220.

Lin H, Werner KM, Inzlicht M (2021) Promises and perils of experimentation: The mutual-internal-validity problem. *Perspect Psychol Sci* 16(4): 854–863.

Lindley DV, Smith AFM (1972) Bayes estimates for the linear model. *J R Stat Soc Series B Stat Methodol* 34: 1–41.

Lipnicki DM, Gunga HC, Belavy DL, Felsenberg D (2009) Decision making after 50 days of simulated weightlessness. *Brain Res* 1280: 84-89.

Llewellyn DJ, Sanchez X (2008) Individual differences and risk taking in rock climbing. *Psychol Sport Exerc* 9: 413–426.

Logan GD (1981) Attention, automaticity, and the ability to stop a speeded choice response. In *Attention and performance IX*, Long J, Baddeley AD (ed) pp. 205–222. Hillsdale, NJ: Erlbaum.

Logan GD (1982) On the ability to inhibit complex movements: A stop-signal study of typewriting. *J Exp Psychol Hum Percept Perform* 8(6): 778.

Logan GD (1994) On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In *Inhibitory processes in attention, memory, and language*, Dagenbach D, Carr TH (ed) pp. 189–239. Cambridge: Academic Press.

Logan GD, Burkell J (1986) Dependence and independence in responding to double stimulation: A comparison of stop, change, and dual-task paradigms. *J Exp Psychol Hum Percept Perform* 12: 549–563.

Logan GD, Cowan WB (1984) On the ability to inhibit thought and action: A theory of an act of control. *Psychol Rev* 91(3): 295.

Logan GD, Cowan WB (1984) On the ability to inhibit thought and action: A theory of an act of control. *Psychol Rev* 91: 295–327.

Logan GD, Schachar RJ, Tannock R (1997) Impulsivity and inhibitory control. *Psychol Sci* 8(1): 60-64.

Logan GD, Van Zandt T, Verbruggen F, Wagenmakers EJ (2014) On the ability to inhibit thought and action: general and special theories of an act of control. *Psychol Rev* 121(1): 66.

López-Caneda E, Rodríguez Holguín S, Cadaveira F, Corral M, Doallo S (2014) Impact of alcohol use on inhibitory control (and vice versa) during adolescence and young adulthood: a review. *Alcohol Alcohol* 49(2): 173-181.

Luce R (1959) *Individual choice behavior*. Wiley, New York, USA.

Luengo MA, Carrillo-de-la-Pena MT, Otero JM, Romero E (1994) A short-term longitudinal study of impulsivity and antisocial behavior. *J Pers Soc Psychol* 66: 542–548.

Lynam DR, Smith GT, Whiteside SP, Cyders MA (2006) *The UPPS-P: Assessing five personality pathways to impulsive behavior*. Purdue University, West Lafayette, USA.

Macmillan M (1992) Inhibition and the control of behavior. From Gall to Freud via Phineas Gage and the frontal lobes. *Brain Cogn* 19: 72–104.

MacPherson SE, Phillips LH, Della Sala S (2002) Age, executive function, and social decision making: A dorsolateral prefrontal theory of cognitive aging. *Psychol Aging* 17(4): 598-609.

Magar EC, Phillips LH, Hosie JA (2008) Self-regulation and risk-taking. *Pers Individ Differ* 45(2): 153-159.

Maia TV, McClelland JL (2004) A reexamination of the evidence for the somatic marker hypothesis: What participants really know in the Iowa gambling task. *Proc Natl Acad Sci USA* 101: 16075–16080.

Maples JL, Carter NT, Few LR, Crego C, Gore WL, Samuel DB, Williamson RL, Lynam DR, Widiger TA, Markon KE *et al* (2015) Testing whether the *DSM-5* personality disorder trait model can be measured with a reduced set of items: An item response theory investigation of the Personality Inventory for *DSM-5*. *Psychol Assess* 27: 1195-1210.

Marks ADG, Hine DW, Blore RL, Phillips WJ (2008) Assessing individual differences in adolescents' preference for rational and experiential cognition. *Pers Individ Differ* 44: 42–52.

Matzke D (2014) Bayesian explorations in mathematical psychology. *J Exp Psychol Gen* 142: 1047-1073.

Matzke D (2019) Dora Matzke. Available at: https://osf.io/kk287

Matzke D, Curley S, Gong CQ, Heathcote A (2019) Inhibiting responses to difficult choices. *J Exp Psychol Gen* 148(1): 124.

Matzke D, Dolan CV, Logan GD, Brown SD, Wagenmakers EJ (2013a) Bayesian parametric estimation of stop-signal reaction time distributions. *J Exp Psychol Gen* 142: 1047–1073.

Matzke D, Hughes M, Badcock JC, Michie P, Heathcote A (2017a) Failures of cognitive control or attention? The case of stop-signal deficits in schizophrenia. *Atten Percept* 79*(*4): 1078-1086.

Matzke D, Love J, Heathcote A (2017b) A Bayesian approach for estimating the probability of trigger failures in the stop-signal paradigm. *Behav Res Methods* 49(1): 267-281.

Matzke D, Love J, Wiecki TV, Brown S D, Logan GD, Wagenmakers EJ (2013b) Release the BEESTS: Bayesian estimation of ex-Gaussian stop-signal reaction time distributions. *Front Psychol* 4: 918.

Matzke D, Verbruggen F, Logan G (2018) The stop-signal paradigm. In *Stevens' handbook of experimental psychology and cognitive neuroscience,* Wagenmakers EJ, Wixted JT (ed) pp 383-427. Hoboken, USA: Wiley-Blackwell.

Matzke D, Wagenmakers EJ (2009) Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychon Bull Rev* 16: 798–817.

Mazas CA, Finn PR, Steinmetz JE (2000) Decision making biases, antisocial personality, and early-onset alcoholism. *Alcohol Clin Exp Res* 24: 1036–1040.

McGarry T, Chua R, Franks IM (2003) Stopping and restarting an unfolding action at various times. *Q J Exp Psychol* 56(4): 1-20.

Meltzer SJ (1899) Inhibition. *N Y State J Med* 69: 661–666, 699–703, 739–743.

Metcalfe J, Mischel W (1999) A hot/cool-system analysis of delay of gratification: dynamics of willpower. *Psychol Rev* 106: 3–19.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24: 167–202.

Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD (2000) The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cog Psychol* 41(1): 49-100.

Moeller FG, Barratt ES, Dougherty DM, Schmitz JM, Swann AC (2001) Psychiatric aspects of impulsivity. *Am J Psychiatry* 158(11): 1783-1793.

Moffitt TE (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. *Psychol Rev* 100: 674–701.

Monterosso JR, Aron AR, Cordova X, Xu J, London ED (2005) Deficits in response inhibition associated with chronic methamphetamine abuse. *Drug Alcohol Depend* 79(2): 273-277.

Morein-Zamir S, Nagelkerke P, Chua R, Franks I, Kingstone A (2004) Inhibiting prepared and ongoing responses: Is there more than one kind of stopping? *Psychon Bull Rev* 11(6): 1034-1040.

Morris RG, Worsley C, Matthews D (2000) Neuropsychological assessment in older people: Old principles and new directions. *Adv Psychiatr Treat* 6: 362-370.

Moutoussis M, Garzón B, Neufeld S, Bach DR, Rigoli F, Goodyer I, Bullmore E; NSPN Consortium, Guitart-Masip M, Dolan RJ (2021) Decision-making ability, psychopathology, and brain connectivity. *Neuron* 109(12): 2025-2040.

Mueller ST (2010) A partial implementation of the BICA Cognitive Decathlon using the Psychology Experiment Building Language (PEBL). *Int J Mach Conscious* 2: 273–288.

Mueller ST (2012) Developing open source tests for psychology and neuroscience. Available at: http://opensource.com/life/12/12/developing-open-source-tests-psychology-and-neuroscience

Mueller ST (2013) The Psychology Experiment Building Language (Version 0.13) [Software]. Available at: http://pebl.sourceforge.net

Mueller ST, Piper BJ (2014) The psychology experiment building language (PEBL) and PEBL test battery. *J Neurosci Methods* 222: 250–259.

Munakata Y, Herd SA, Chatham CH, Depue BE, Banich MT, O'Reilly RC (2011) A unified framework for inhibitory control. *TiCS* 15(10): 453-459.

Munoz DP, Schall JD (2003) Concurrent, distributed control of saccade initiation in the frontal eye field and superior colliculus. In *The superior colliculus: New approaches for studying sen- sorimotor integration*, Hall WT, Moschovakis A (ed.) pp. 55–82. Boca Raton, FL: CRC Press.

Münsterberg H (1900) *Grundzüge der Psychologie*. Barth, Leipzig.

Myrseth H, Tverå R, Hagatun S, Lindgren C (2012) A comparison of impulsivity and sensation seeking in pathological gamblers and sky-divers. *Scand J Psychol* 53: 340–346.

Myung IJ, Pitt MA (1997) Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychon Bull Rev* 4: 79– 95.

National Advisory Mental Health Council Workgroup on Tasks and Measures for Research Domain Criteria (2016) Behavioral assessment methods for RDoC constructs. Available at: https://www.nimh.nih.gov/about/advisory-boards-and-groups/namhc/reports/behavioral-assessment-methods-for-rdoc-constructs

Nederkoorn C, Jansen E, Mulkens S, Jansen A (2007) Impulsivity predicts treatment outcome in obese children. *Behav Res Ther* 45(5): 1071-1075.

Nejtek VA, Kaiser KA, Zhang B, Djokovic M (2013) Iowa gambling task scores predict future drug use in bipolar disorder outpatients with stimulant dependence. *Psychiatry Res* 210(3): 871–879.

Ness V, Arning L, Niesert HE, Stüttgen MC, Epplen JT, Beste C (2011) Variations in the GRIN2B gene are associated with risky decision-making. *Neuropharmacology* 61(5-6): 950-956.

Nicholl CG, Lynch S, Kelly CA, White L, Simpson PM, Wesnes KA, Pitt BM (1995) The cognitive drug research computerized assessment system in the evaluation of early dementia-is speed of the essence? *Int J Geriatr Psychiatry* 10: 199-206.

Nicholson N, Soane E, Fenton-O'Creevy M, Willman P (2005) Personality and domain-specific risk taking. *J Risk Res* 8(2): 157-176.

Nigg JT (2001) Is ADHD a disinhibitory disorder? *Psychol Bull* 127(5): 571.

Nigg JT (2005) Neuropsychologic theory and findings in attention-deficit/hyperactivity disorder: the state of the field and salient challenges for the coming decade. *Biol Psychiatry* 57: 1424–1435.

Nilsson H, Rieskamp J, Wagenmakers EJ (2011) Hierarchical Bayesian parameter estimation for cumulative prospect theory. *J Math Psychol* 55: 84 –93.

Noël X, Bechara A, Dan B, Hanak C, Verbanck P (2007) Response inhibition deficit is involved in poor decision making under risk in nonamnesic individuals with alcoholism. *Neuropsychology* 21: 778–786.

Nosek BA, Hardwicke TE, Moshontz H, Allard A, Corker KS, Dreber A, Fidler F, Hilgard J, Struhl MK, Nuijten MB, *et al* (2022). Replicability, robustness, and reproducibility in psychological science. *Annu Rev Psychol* 73: 719-748.

Ollman RT (1973) Simple reactions with random countermanding of the “go”-signal. In *Attention and performance IV*, Kornblum S (ed) pp 571–581. New York: Academic Press.

Olmstead MC (2006) Animal models of drug addiction: where do we go from here? *Quarterly J Exp Psychol* 59: 625–653.

Online Survey Available at: https://www.onlinesurveys.ac.uk/

Osman A, Kornblum S, Meyer DE (1986) The point-of-no-return in choice reaction-time—Controlled and ballistic stages of response preparation. *J Exp Psychol Hum Percept Perform* 12: 243–258.

Ottaviani C, Vandone D (2011) Impulsivity and household indebtedness: Evidence from real life. *J Econ Psychol* 32: 754–761.

Overman WH, Pierce A (2013) Iowa Gambling Task with non-clinical participants: effects of using real+ virtual cards and additional trials. *Front Psychol* 4: 935.

Pacini R, Epstein S (1999) The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio- bias phenomenon. *J Pers Soc Psychol* 76*:* 972–987.

Passetti, F, Clark L, Mehta MA, Joyce E, King M (2008) Neuropsychological predictors of clinical outcome in opiate addiction. *Drug Alcohol Depend* 94(1–3): 82–91.

Patton JH, Stanford MS, Barratt ES (1995) Factor structure of the Barratt impulsiveness scale. *J Clin Psychol* 51(6): 768-774.

Patton JH, Stanford MS, Barratt ES (1995) Factor structure of the Barratt impulsiveness scale. *J Clin Psychol* 51(6): 768-774.

Patton JH, Stanford MS, Barratt ES, (1995) Factor structure of the Barratt impulsiveness scale. *J Clin Psychol* 51: 768–774.

Perry JL, Carroll ME (2008). The role of impulsive behavior in drug abuse. *Psychopharmacology* 200: 1–26.

Persaud N, McLeod P, Cowey A (2007) Post-decision wagering objectively measures awareness. *Nat Neurosci* 10: 257–261

Petry N, Bickel W, Amett M (1998) Shortened time horizons and insensitivity to future consequences in heroin addicts. *Addiction* 93: 729 –738.

Phillips WJ, Fletcher JM, Marks AD, Hine DW (2016) Thinking styles and decision making: A meta-analysis. *Psychol Bull* 142(3): 260.

Piper BJ (2011) Age, handedness, and sex contribute to fine motor behavior in children. *J Neurosci Methods* 195(1): 88-91.

Piper BJ (2012) Evaluation of the test-retest reliability of the PEBL continuous performance test in a normative sample. *PEBL Technical Report Series*

Poletti M, Cavedini P, Bonuccelli U (2011) Iowa gambling task in Parkinson's disease. *J Clin Exp Neuropsychol* 33(4): 395–409.

Polit DF (2014) Getting serious about test–retest reliability: A critique of retest research and some recommendations. *Qual Life Res* 23: 1713–1720.

Postman L, Bruner JS, McGinnies E (1948) Personal values as selective factors in perception. *J Abnorm Psychol* 43: 142–154.

Qiu J, Helbig R (2012) Body posture as an indicator of workload in mental work. *Hum Factors* 54(4): 626-635.

Radat F, Chanraud S, Di Scala G, Dousset V, Allard M (2013) Psychological and neuropsychological correlates of dependence-related behaviour in medication overuse headaches: a one year follow-up study. *J Headache Pain* 14(1): 59.

Rai AA, Stanton B, Wu Y, Li X, Galbraith J, Cottrell L, Pack R, Harris C, D'Alessandri D, Burns J (2003) Relative influences of perceived parental monitoring and perceived peer involvement on adolescent risk behaviors: an analysis of six cross-sectional data sets. *J Adolesc Health* 33: 108–118.

Ratcliff R (1978) A theory of memory retrieval. *Psychol Rev* 85: 59–108.

Ratcliff R (1993) Methods for dealing with reaction time outliers. *Psychol Bull* 114: 510–532.

Ratcliff R, Murdock BB (1976) Retrieval processes in recognition memory. *Psychol Rev* 83: 190–214.

Raud L, Westerhausen R, Dooley N, Huster RJ (2020) Differences in unity: The go/no-go and stop signal tasks rely on different mechanisms. *NeuroImage* 210: 116582.

Revelle W, Zinbarg RE (2009) Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika* 74: 145–154.

Ribot T (1889) *Psychologie de l'attention*. Alcan, Paris.

Ridderinkhof KR, Band GPH, Logan GD (1999) A study of adaptive behavior: Effects of age and irrelevant information on the ability to inhibit one's actions. *Acta Psychol* 101: 315–337.

Ritter LM, Meador-Woodruff JH, Dalack GW (2004) Neurocognitive measures of prefrontal cortical dysfunction in schizophrenia. *Schizophr res* 68: 65–73.

Robinson ES, Eagle DM, Economidou D, Theobald DE, Mar AC, Murphy ER, Robbins TW, Dalley JW (2009) Behavioural characterisation of high impulsivity on the 5-choice serial reaction time task: specific deficits in 'waiting' versus 'stopping'. *Behav Brain Res* 196: 310–316.

Rodríguez-Sánchez JM, Crespo-Facorro B, Perez Iglesias R, González-Blanch Bosch C, Álvarez M, Llorca J, Vázquez- Barquero JL (2005) Prefrontal cognitive functions in stabilized first- episode patients with schizophrenia spectrum disorders: A dissociation between dorsolateral and orbitofrontal functioning. *Schizophr res* 77: 279–288.

Roiser JP, Cannon DM, Gandhi SK, Taylor Tavares J, Erickson K, Wood S, Klaver JM, Clark L, Zarate Jr CA, Sahakian BJ, Drevets WC (2009) Hot and cold cognition in unmedicated depressed subjects with bipolar disorder. *Bipolar Disord* 11: 178–189.

Rosenbloom MH, Schmahmann JD, Price BH (2012) The functional neuroanatomy of decision-making. *J Neuropsychiatry Clin Neurosci* 24: 266–277.

Rouder JN, Lu J (2005) An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychon Bull Rev* 12: 573–604.

Rouder JN, Lu J, Speckman P, Sun D, Jiang Y (2005) A hierarchical model for estimating response time distributions. *Psychon Bull Rev* 12: 195–223.

Rouder JN, Sun D, Speckman PL, Lu J, Zhou D (2003) A hierarchical Bayesian statistical framework for response time distributions. *Psychometrika* 68: 589–606.

Salgado JV, Malloy-Diniz LF, Campos VR, Abrantes SS, Fuentes D, Bechara A, Correa H (2009) Neuropsychological assessment of impulsive behavior in abstinent alcohol-dependent subjects. *Rev Bras de Psiquiatr* 31(1): 4–9.

Schachar R, Logan GD (1990) Impulsivity and inhibitory control in normal development and childhood psychopathology. *Dev Psychol* 26(5): 710.

Schachar R, Logan GD, Robaey P, Chen S, Ickowicz A, Barr C (2007) Restraint and cancellation: multiple inhibition deficits in attention deficit hyperactivity disorder. *J Abnorm Child Psychol* 35(2): 229–38.

Schachar R, Logan GD, Robaey P, Chen S, Ickowicz A, Barr C (2007) Restraint and cancellation: multiple inhibition deficits in attention deficit hyperactivity disorder. *J Abnorm Child Psychol* 35(2): 229-238.

Schall JD (2002) The neural selection and control of saccades by the frontal eye field. *Philos Trans R Soc Lond B Biol Sci* 357(1424): 1073-1082.

Scheres A, Oosterlaan J, Sergeant JA (2001) Response execution and inhibition in children with AD/HD and other disruptive disorders: the role of behavioural activation. *J Child Psychol Psychiatry* 42: 347–357.

Schmitt N (1996) Uses and abuses of coefficient alpha. *Psychol Assess* 8(4): 350.

Schmitt WA, Brinkley CA, Newman JP (1999) Testing Damasio's somatic marker hypothesis with psychopathic individuals: Risk takers or risk averse? *J Abnorm Psychol* 108:538 –543.

Schmitz F, Kunina-Habenicht O, Hildebrandt A, Oberauer K, Wilhelm O (2020) Psychometrics of the Iowa and Berlin gambling tasks: Unresolved issues with reliability and validity for risk taking. *Assessment* 27(2): 232-245.

Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6: 461–464.

Schwebel DC, Severson J, Ball KK, Rizzo M (2006) Individual difference factors in risky driving: The roles of anger/hostility, conscientiousness, and sensation-seeking. *Accid Anal Prev* 38(4): 801-810.

Scott SG, Bruce RA (1995) Decision-making style: the development of a new measure. *Educ Psychol Meas* 55: 818-831.

Seguin JR, Arseneault L, Tremblay RE (2007) The contribution of “cool” and “hot” components of decision-making in adolescence: implications for developmental psychopathology. *Cogn Dev* 22: 530–543.

Shafir E, Simonson I, Tversky A (1993) Reason-based choice. *Cognition* 49: 11–36.

Sharma L, Markon KE, Clark LA (2014) Toward a theory of distinct types of “impulsive” behaviors: a meta-analysis of self-report and behavioral measures. *Psychol Bull* 140(2): 374.

Shiffrin RM, Lee MD, Kim W, Wagenmakers EJ (2008) A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cogn Sci,* 32: 1248–1284.

Shurman B, Horan WP, Nuechterlein KH (2005) Schizophrenia patients demonstrate a distinctive pattern of decision-making impairment on the Iowa Gambling Task. *Schizophr Res* 72: 215–224.

Shuster J, Toplak ME (2009) Executive and motivational inhibition: associations with self-report measures related to inhibition. *Conscious Cogn* 18: 471–480.

Sijtsma K (2009) On the use, the misuse, and the very limited usefulness of Cronbach’s alpha. *Psychometrika* 74: 107–120.

Simmonds DJ, Pekar JJ, Mostofsky SH (2008) Meta-analysis of Go/No-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia* 46(1): 224–32.

Simpson PM, Surmon DJ, Wesnes KA, Wilcock GK (1991) The cognitive drug research computerized assessment system for demented patients: a validation study. *Int J Geriatr Psychiatry* 6: 95-102.

Skaggs EB (1929) The major descriptive categories of inhibition in psychology. *J Abnorm Soc Psych* 24: 310–317.

Skeel RL, Neudecker J, Pilarski C, Pytlak K (2007) The utility of personality variables and behaviorally-based measures in the prediction of risk-taking behavior. *Pers Individ Differ* 43: 203–214.

Skippen P, Fulham WR, Michie PT, Matzke D, Heathcote A, Karayanidis F (2020) Reconsidering electrophysiological markers of response inhibition in light of trigger failures in the stop-signal task. *Psychophysiology* 57(10): e13619.

Skippen P, Matzke D, Heathcote A, Fulham WR, Michie P, Karayanidis F (2019) Reliability of triggering inhibitory process is a better predictor of impulsivity than SSRT. *Acta Psychol* 192: 104-117.

Slevc LR, Ferreira VS (2006) Halting in single word production: A test of the perceptual loop theory of speech monitoring. *J Mem Lang* 54(4): 515-540.

Sloman SA (2002) Two systems of reasoning. In *Heuristics and biases: The psychology of intuitive judgment*, Gilovich T, Griffin D, Kahneman D (ed) pp 379–398. New York: Cambridge University Press.

Smith L (1952) *A Dictionary of Psychiatry for the Layman*. Maxwell, London.

Smith R (1992) *Inhibition: History and Meaning in the Sciences of Mind and Brain*. University of California Press, Berkeley, CA.

Soltanifar M, Escobar M, Dupuis A, Schachar R (2021) A Bayesian mixture modelling of stop signal reaction time distributions: The second contextual solution for the problem of aftereffects of inhibition on SSRT estimations. *Brain Sci* 11(8): 1102.

Somma A, Krueger RF, Markon KE, Borroni S, Fossati A (2019) Item response theory analyses, factor structure, and external correlates of the Italian translation of the personality inventory for DSM-5 short form in community-dwelling adults and clinical adults. *Assessment* 26(5): 839-852

Somma A, Marelli S, Barranca M, Gialdi G, Lucini C, Castelnuovo A, Mombelli S, Ferini-Strambi L, Fossati A (2020) Executive functioning and personality traits in insomnia disorder: a preliminary report on the clinical importance of objective and subjective reduction of total sleep time. *Mediterr J Clin Psychol* 8(1).

Soreni N, Crosbie J, Ickowicz A, Schachar R (2009) Stop signal and conners' continuous performance tasks: Test—retest reliability of two inhibition measures in adhd children. *J Atten Disord* 13(2): 137-143.

Stahl C, Voss A, Schmitz F, Nuszbaum M, Tüscher O, Lieb K, Klauer KC (2014) Behavioral components of impulsivity. *J Exp Psychol* 143(2): 850.

Stanovich KE, West RF, Toplak ME (2011) The complexity of developmental predictions from dual process models. *Dev Rev* 31: 103–118.

Starcevic V (2016) Behavioural addictions: A challenge for psychopathology and psychiatric nosology. *Aust N Z J Psychiatry* 50(8): 721-725.

Starcke K, Wolf OT, Markowitsch HJ, Brand M (2008) Anticipatory stress influences decision making under explicit risk conditions. *Behav Neurosci* 122(6): 1352.

Stein DJ (2000) Neurobiology of the obsessive–compulsive spectrum disorders. *Biol Psychiatry* 47(4): 296-304.

Stein DJ, Hollander E, Simeon D, Cohen L (1994) Impulsivity scores in patients with obsessive-compulsive disorder. *J Nerv Ment Dis* 182: 240-241

Stein DJ, Trestman RL, Mitropoulou V, Coccaro EF, Hollander E, Siever LJ (1996) Impulsivity and serotonergic function in compulsive personality disorder. *J Neuropsychiatry Clin Neurosci* 8: 393-398

Steingroever H, Pachur T, Šmíra M, Lee MD (2018) Bayesian techniques for analyzing group differences in the Iowa Gambling Task: A case study of intuitive and deliberate decision-makers. *Psychon Bull Rev* 25(3): 951-970.

Steingroever H, Wetzels R, Horstmann A, Neumann J, Wagenmakers EJ (2013) Performance of healthy participants on the Iowa Gambling Task. *Psychol Assess* 25(1): 180-193.

Steingroever H, Wetzels R, Wagenmakers EJ (2013a) A comparison of reinforcement-learning models for the Iowa gambling task using parameter space partitioning. *J Probl Solving* 5: 2.

Steingroever H, Wetzels R, Wagenmakers EJ (2013b) Validating the PVL-Delta model for the Iowa gambling task. *Front Psychol* 4.

Steingroever H, Wetzels R, Wagenmakers EJ (2014) Absolute performance of reinforcement- learning models for the Iowa gambling task. *Decis* 1: 161–183.

Steingroever H, Wetzels R, Wagenmakers EJ (2015) $\hat{w}$ = .2, $\hat{a}$ = .8, $\hat{c}$ = .6: So what? On the meaning of parameter estimates from reinforcement-learning models. *Decis* 2: 228–235.

Steingroever H, Wetzels R, Wagenmakers EJ (2016) Bayes factors for reinforcement-learning models of the Iowa gambling task. *Decis* 3(2): 115.

Stout J, Busemeyer J, Lin A, Grant S, Bonson K (2004) Cognitive modeling analysis of decision-making processes in cocaine abusers. *Psychon Bull Rev* 11: 742–747.

Stout JC, Busemeyer JR, Lin A, Grant SJ, Bonson KR (2004) Cognitive modeling analysis of decision-making processes in cocaine abusers. *Psychon Bull Rev* 11(4): 742–747.

Stout JC, Rodawalt WC, Siemers ER (2001) Risky decision making in Huntington's disease. *J Int Neuropsychol Soc* 7(1): 92-101.

Strack F, Deutsch R (2004) Reflective and impulsive determinants of social behavior. *Pers Soc Psychol Rev* 8(3): 220–247.

Stroop JR (1935) Studies of interference in serial verbal reactions. *J Exp Psychol* 18(6): 643.

Swann AC, Bjork JM, Moeller FG, Dougherty DM (2002) Two models of impulsivity: relationship to personality traits and psychopathology. *Biol Psychiatry* 51: 988–994.

Tannock R, Schachar RJ, Carr RP, Chajczyk D, Logan GD (1989) Effects of Methylphenidate on Inhibitory Control in Hyperactive Children. *J Abnorm Child Psychol* 17(5): 473–91.

Teasdale JD, Lloyd CA, Hutton JM (1998) Depressive thinking and dysfunctional schematic mental models. *Br J Clin Psychol* 37: 247–257.

Teese R, Bradley G (2008) Predicting recklessness in emerging adults: A test of a psychosocial model. *J Soc Psychol* 148: 105–128.

Thompson VA (2009) Dual-process theories: A meta-cognitive perspective. In *In two minds: Dual processes and beyond*, Evans StBT, Frankish K (ed) pp. 171–196). New York: Oxford University Press.

Thorndike EL (1898) Animal intelligence: An experimental study of the associative processes in animals. *Psychol Rev* 2(4): i–109.

Tremblay RE, Pihl RO, Vitaro F, Dobkin PL (1994) Predicting early onset of male antisocial behavior from preschool behavior. *Arch Gen Psychiatry* 51: 732–739.

Tversky A, Kahneman D (1992) Advances in prospect theory: Cumulative representation of uncertainty. *J Risk Uncertain* 5: 297–323.

Ulrich R, Miller J (1994) Effects of truncation on reaction time analysis. *J Exp Psychol* 123: 34–80.

Upton DJ, Bishara AJ, Ahn WY, Stout JC (2011) Propensity for risk taking and trait impulsivity in the Iowa Gambling Task. *Pers Individ Differ* 50(4): 492-495.

Urcelay GP, Dalley JW (2012) Linking ADHD, impulsivity, and drug abuse: a neuropsychological perspective. *Curr Top Behav Neurosci* 9: 173–197.

van de Laar MC, van den Wildenberg WP, van Boxtel GJ, van der Molen MW (2010) Processing of Global and Selective Stop Signals. *Exp Psychol* 57(2): 149–159.

Van den Bos R, De Visser L, Van de Loo AJAE, Mets MAJ, Van Willigenburg GM, Homberg JR, Verster JC (2012) *Sex differences in decision-making in adult normal volunteers are related to differences in the interaction of emotion and cognitive control*. Handbook on Psychology of Decision-making. Nova Science Publisher.

Van den Bos R, den Heijer E, Vlaar S, Houx B (2007) *Exploring gender differences in decision-making using the Iowa Gambling Task*. Encyclopedia of psychology of decision making. Nova Science Publishers.

van den Wildenberg WPM, van der Molen MW (2004) Developmental trends in simple and selective inhibition of compatible and incompatible responses. *J Exp Child Psychol* 87(3): 201-220.

van den Wildenberg WPM, van der Molen MW, Logan GD (2002) Reduced response readiness delays stop signal inhibition. *Acta Psychol* 111: 155– 169.

van Gaal S, Lamme VAF, Fahrenfort JJ, Ridderinkhof KR (2010) Dissociable brain mechanisms underlying the conscious and unconscious control of behavior. *J Cognit Neurosci* 23: 91–105.

van Gaal S, Ridderinkhof KR, Fahrenfort JJ, Scholte HS, Lamme VAF (2008) Frontal cortex mediates unconsciously triggered inhibitory control. *J Neurosci* 28: 8053–8062.

van Gaal S, Ridderinkhof KR, van den Wildenberg WPM, Lamme VAF (2009) Dissociating consciousness from inhibitory control: evidence for unconsciously triggered response inhibition in the stop-signal task. *J Exp Psychol Hum Percept Perform* 35: 1129–1139.

Vasconcelos AG, Sergeant J, Corrêa H, Mattos P, Malloy-Diniz L (2014) When self-report diverges from performance: The usage of BIS-11 along with neuropsychological tests. *Psychiatry Res* 218(1-2): 236-243.

Verbruggen F (2019) Software for the stop-signal paradigm. Available at: osf.io/wuhpv

Verbruggen F, Adams RC, van‘t Wout F, Stevens T, McLaren IP, Chambers CD (2013) Are the effects of response inhibition on gambling long-lasting?. *PloSone* 8(7): e70155.

Verbruggen F, Aron AR, Band GP, Beste C, Bissett PG, Brockett AT, Brown JW, Chamberlain SR, Chambers CD, Colonius H *et al* (2019) A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. *eLife* 8: e46323.

Verbruggen F, Chambers CD, Logan GD (2013) Fictious inhibitory differences: How skewness and slowing distort the estimation of stopping latencies. *Psychol Sci* 24: 352–362.

Verbruggen F, Liefooghe B, Vandierendonck A (2004) The interaction between stop signal inhibition and distractor interference in the flanker and Stroop task. *Acta Psychol* 116: 21–37.

Verbruggen F, Logan GD (2008) Response inhibition in the stop-signal paradigm. *Trends Cogn Sci* 12(11): 418–24.

Verbruggen F, Logan GD (2008a) Long-term aftereffects of response inhibition: Memory retrieval task goals and cognitive control. *J Exp Psychol Hum Percept Perform* 34: 1229–1235.

Verbruggen F, Logan GD (2009) Models of response inhibition in the stop-signal and stop-change paradigms. *Neurosci Biobehav Rev* 33: 647–661.

Verbruggen F, Logan GD, Stevens MA (2008b) STOP-IT: Windows executable software for the stop-signal paradigm. *Behav Res Methods* 40: 479–483.

Verbruggen F, Stevens T, Chambers CD (2014) Proactive and reactive stopping when distracted: An attentional account. *J Exp Psychol Hum Percept Perform* 40: 1295–1300.

Verdejo-Garcia A, Lawrence AJ, Clark L (2008) Impulsivity as a vulnerability marker for substance-use disorders: review of findings from high-risk research, problem gamblers and genetic association studies. *Neurosci Biobehav Rev* 32: 777–810.

Verdejo-Garcia A, Perez-Garcia M (2007) Profile of executive deficits in cocaine and heroin polysubstance users: common and differential effects on separate executive components. *Psychopharmacology* 190: 517–530.

Vince MA (1948) The intermittency of control movements and the psychological refractory period. *Br J Psychol* 38(3): 149.

Volkow ND, Fowler JS, Wang GJ (2002) Role of dopamine in drug reinforcement and addiction in humans: results from imaging studies. *Behav Pharmacol* 13: 355–366.

Wagenmakers EJ, Lodewyckx T, Kuriyal H, Grasman R (2010) Bayesian hypothesis testing for psychologists: A tutorial on the Savage- Dickey method. *Cogn Psychol* 60: 158–189.

Wagenmakers EJ, van der Maas HL, Dolan CV, Grasman RP (2008) EZ does it! Extensions of the EZ-diffusion model. *Psychon Bull Rev* 15(6): 1229-1235.

Wang G, Shi J, Chen N, Xu L, Li J, Li P, Sun Y, Lu L (2013). Effects of length of abstinence on decision-making and craving in methamphetamine abusers. *PLoS One* 8(7): e68791.

Weigard A, Matzke D, Tanis C, Heathcote A (2021) Cognitive process modeling addresses context independence violations in the ABCD Study stop-signal task. *bioRxiv*.

Werner KM, Inzlicht M, Ford BQ (2022) Whither Inhibition? *Curr Dir Psychol Sci* 31(4): 333–339.

Wetzels R, Vandekerckhove J, Tuerlinckx F, Wagenmakers EJ (2010) Bayesian parameter estimation in the Expectancy Valence model of the Iowa gambling task. *J Math Psychol* 54: 14–27.

Whiteside SP, Lynam DR (2001) The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Pers Individ Differ* 30: 669–689.

Wilder KE, Weinberger DR, Goldberg TE (1998) Operant conditioning and the orbitofrontal cortex in schizophrenic patients: Unexpected evidence for intact functioning. *Schizophr Res* 30(2): 169-174.

Williams BR, Ponesse JS, Schachar RJ, Logan GD, Tannock R (1999) Development of inhibitory control across the life span. *Dev Psychol* 35(1): 205-213.

Williams BR, Ponesse JS, Schachar RJ, Logan GD, Tannock R (1999) Development of inhibitory control across the life span. *Dev Psychol* 35(1): 205.

Winstanley CA, Theobald DE, Cardinal RN, Robbins TW (2004) Contrasting roles of basolateral amygdala and orbitofrontal cortex in impulsive choice. *J Neurosci* 24: 4.718-4.722.

Wood S, Busemeyer J, Koling A, Cox CR, Davis H (2005) Older adults as adaptive decision makers: Evidence from the Iowa gambling task. *Psychol Aging* 20: 220–225.

Wood SMW, Bechara A (2014) The neuroscience of dual (and triple) systems in decision making. In *The neuroscience of risky decision making,* Reyna VF Zayas VF (ed) pp 177–202. Washington, DC: American Psychological Association Press.

Worthy DA, Hawthorne MJ, Otto AR (2013a) Heterogeneity of strategy use in the Iowa gambling task: A comparison of win-stay/lose-shift and reinforcement learning models. *Psychon Bull Rev* 20(2): 364-371.

Worthy DA, Pang B, Byrne KA (2013b) Decomposing the roles of perseveration and expected value representation in models of the Iowa gambling task. *Front Psychol* 4:640.

Wöstmann NM, Aichert DS, Costa A, Rubia K, Möller HJ, Ettinger U (2013) Reliability and plasticity of response inhibition and interference control. *Brain Cogn* 81(1): 82-94.

Xu S, Korczykowski M, Zhu S, Rao H (2013) Risk-taking and impulsive behaviors: A comparative assessment of three tasks. *Soc Behav Pers* 41(3): 477-486.

Xu S, Xiao Z, Rao H (2019) Hypothetical versus real monetary reward decrease the behavioral and affective effects in the Balloon Analogue Risk Task. *Exp Psychol* 66(3): 221.

Xue G, Aron AR, Poldrack RA (2008) Common neural substrates for inhibition of spoken and manual responses. *Cereb Cortex* 18(8): 1923-1932.

Yang H, Carmon Z, Kahn B, Malani A, Schwartz J, Volpp K, Wansink B (2012) The hot-cold decision triangle: A framework for healthier choices. *Mark Lett* 23: 457–472.

Yang Y, Green SB (2011) Coefficient alpha: A reliability coefficient for the 21st century? *J Psychoeduc Assess* 29: 377–392.

Yechiam E, Busemeyer J (2005) Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychon Bull Rev* 12: 387–402.

Yechiam E, Busemeyer JR (2008) Evaluating generalizability and parameter consistency in learning models. *Games Econ Behav* 63: 370–394.

Yechiam E, Busemeyer JR, Stout JC, Bechara A (2005) Using cognitive models to map relations between neuropsychological disorders and human decision-making deficits. *Psychol Sci* 16: 973–978.

Yechiam E, Ert E (2007) Evaluating the reliance on past choices in adaptive learning models. *J Math Psychol* 51: 75–84.

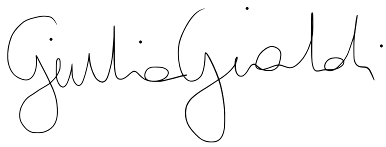
Yechiam E, Kanz JE, Bechara A, Stout JC, Busemeyer JR, Altmaier EM, Paulsen JS (2008) Neurocognitive deficits related to poor decision-making in people behind bars. *Psychon Bull Rev* 15: 44–51.

Youzhi W, Jing L (2009) A research on impulsivity and delay discounting differences between high and low procrastinators. *Psychol Sci* 32: 371–374.

Zacks RT, Hasher L (1994) Direct ignoring: inhibitory regulation of working memory. In *Inhibitory Processes in Attention, Memory and Language*, Dagenbach D, Carr TH (ed) pp 241–264. San Diego, CA: Academic Press.

Zuckerman M, Kuhlman DM, 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. *J Pers Soc Psychol* 65(4): 757.

Zuckerman M, Kuhlman DM, Thornquist M, Kiers H (1991) Five (or three) robust questionnaire scale factors of personality without culture. *Pers Individ Differ* 12(9): 929-941.

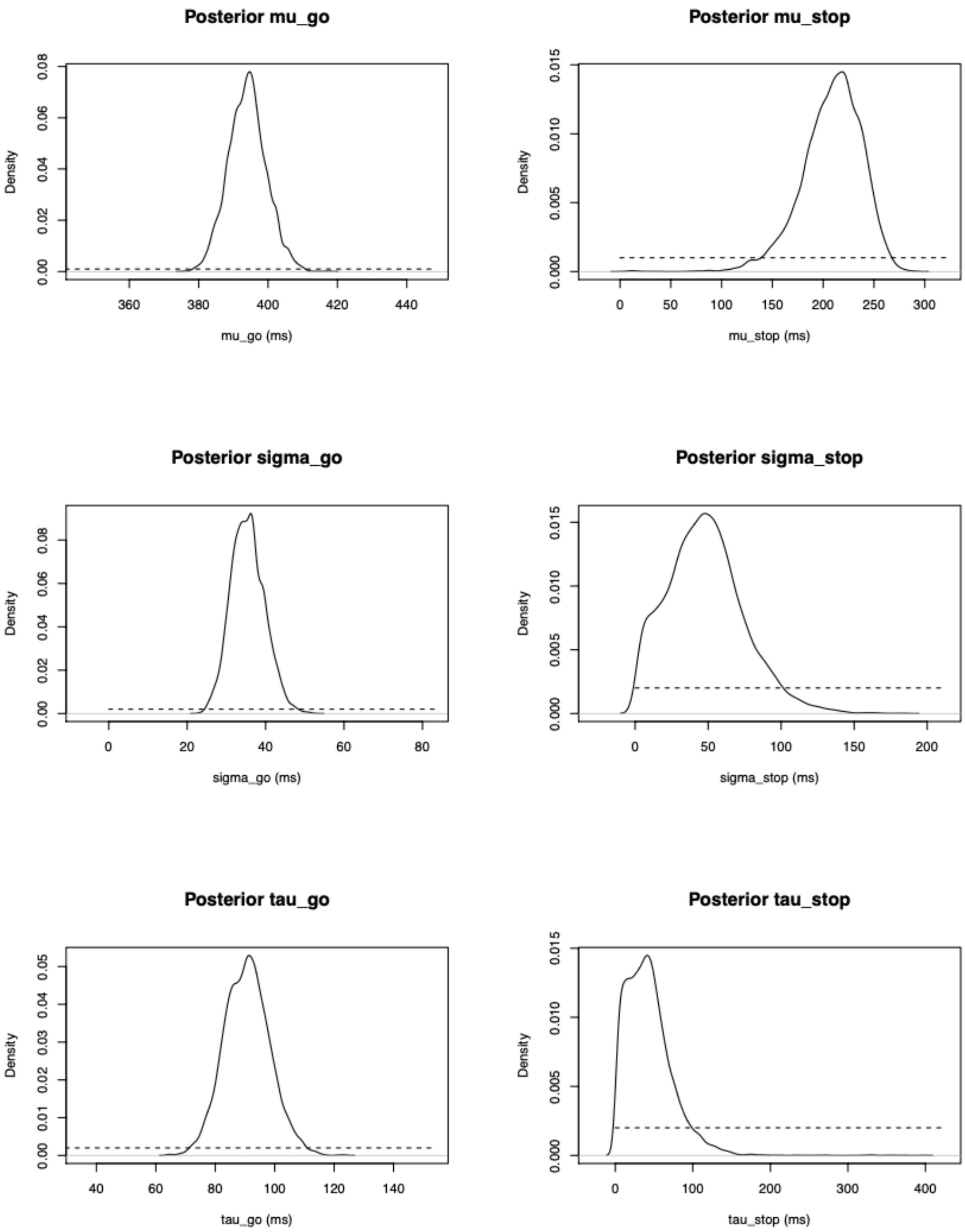


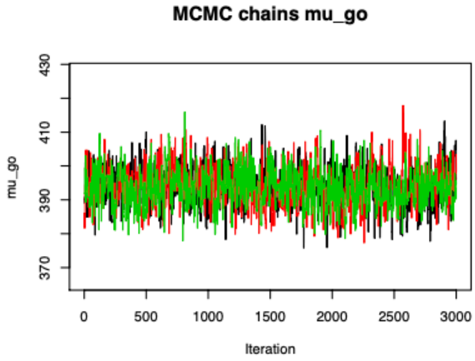
Appendices

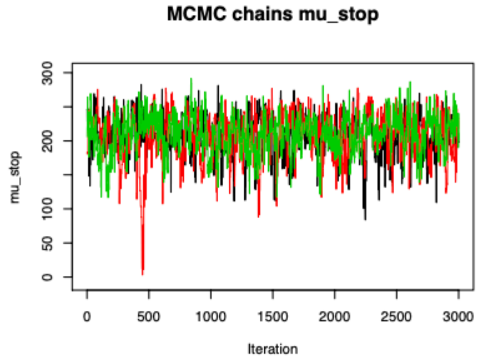
Appendix 1

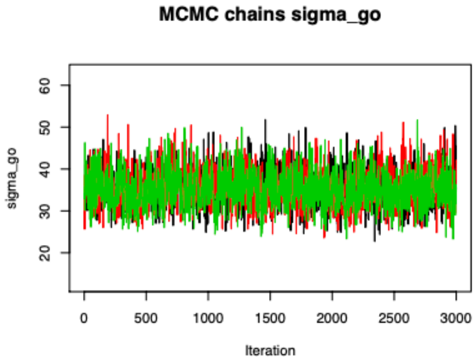
Summary statistics

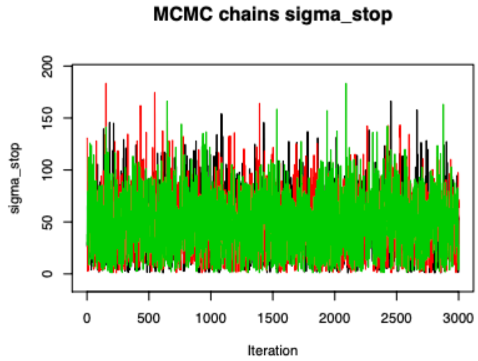
	Mean	Sd	2.5%	25%	50%	75%	97.5%
mu_go	394.012	5.4831	383.5732	390.324	393.9983	397.4258	405.3471
mu_stop	208.4333	29.7028	142.6693	191.3291	211.319	229.1015	257.0458
sigma_go	35.4639	4.3028	27.4835	32.4179	35.2437	38.2941	44.3939
sigma_stop	48.2104	26.3034	4.5062	29.6665	46.8205	63.824	106.3636
tau_go	90.7012	7.6708	76.0743	85.4144	90.6616	95.7104	106.4299
tau_stop	43.6953	31.4018	3.3051	21.2273	39.5123	58.7438	114.7254
mean go	484.7131	5.7413	465.0374	480.7677	484.5845	488.4963	508.1278
sd go	97.5682	6.4953	76.2388	93.0486	97.3997	101.7721	127.8674
mean SSRT	252.1286	15.3605	187.8794	242.6964	251.6564	261.1019	409.7568
sd SSRT	71.6649	27.848	15.4607	53.0538	66.5104	84.3075	411.8119

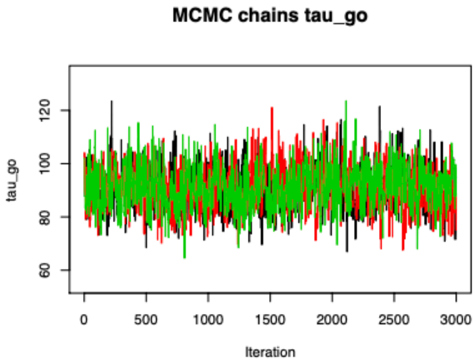

Posterior mu_go
Density
mu_go (ms)
Posterior mu_stop
Density
mu_stop (ms)
Posterior sigma_go
Density
sigma_go (ms)
Posterior sigma_stop
Density
sigma_stop (ms)
Posterior tau_go
Density
tau_go (ms)
Posterior tau_stop
Density
tau_stop (ms)

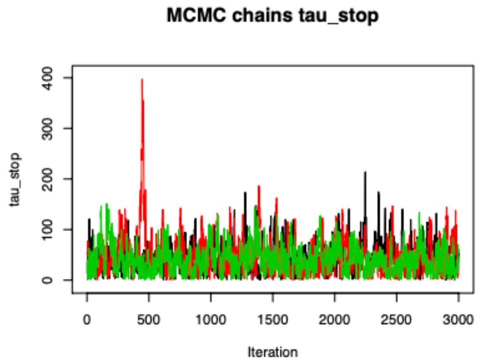

MCMC chains mu_go
mu_go
370
390
410
430
0
500
1000
1500
2000
2500
3000
Iteration


MCMC chains mu_stop
mu_stop
0
50
100
200
300
0
500
1000
1500
2000
2500
3000
Iteration

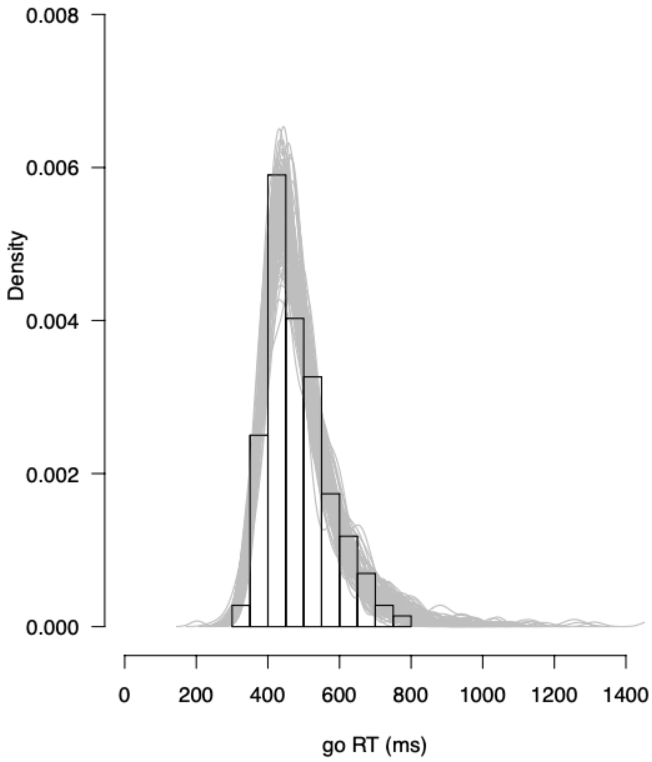

MCMC chains sigma_go
sigma_go
20
30
40
50
60
0
500
1000
1500
2000
2500
3000
Iteration


MCMC chains sigma_stop
sigma_stop
0
50
100
150
200
0
500
1000
1500
2000
2500
3000
Iteration


MCMC chains tau_go
tau_go
60
80
100
120
0
500
1000
1500
2000
2500
3000
Iteration


MCMC chains tau_stop
tau_stop
0
100
200
300
400
0
500
1000
1500
2000
2500
3000
Iteration

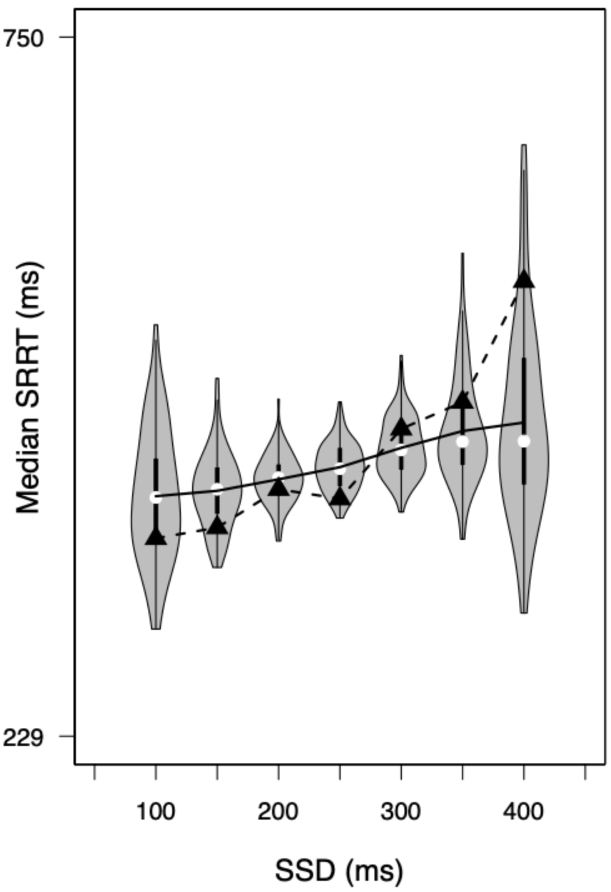
Posterior predictive model check for go RT distribution



Posterior predictive p values for median SRRT

	SSD=100	SSD=150	SSD=200	SSD=250	SSD=300	SSD=350	SSD=400
Number of observed SRRT	2	4	13	13	9	4	1
Observed median SRRT	376.5	384.5	413	406	458	478	568
Average predicted SRRT	407.87	411.87	420.44	429.49	443.91	456.42	462.77
One−sided p value	0.75	0.828	0.68	0.91	0.27	0.25	0.096
Two−sided p value	0.5	0.343	0.64	0.18	0.54	0.5	0.191

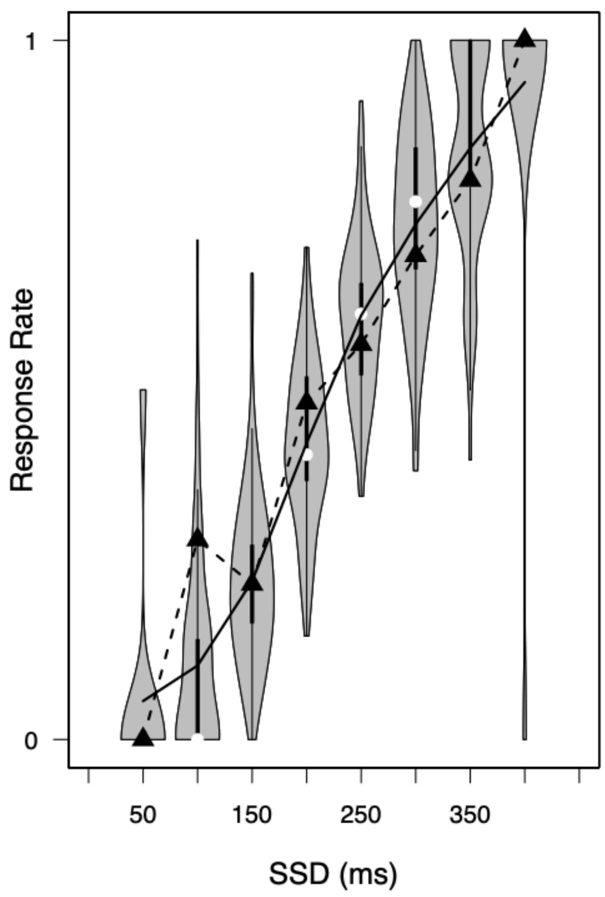
Posterior predictive model check for median SRRT

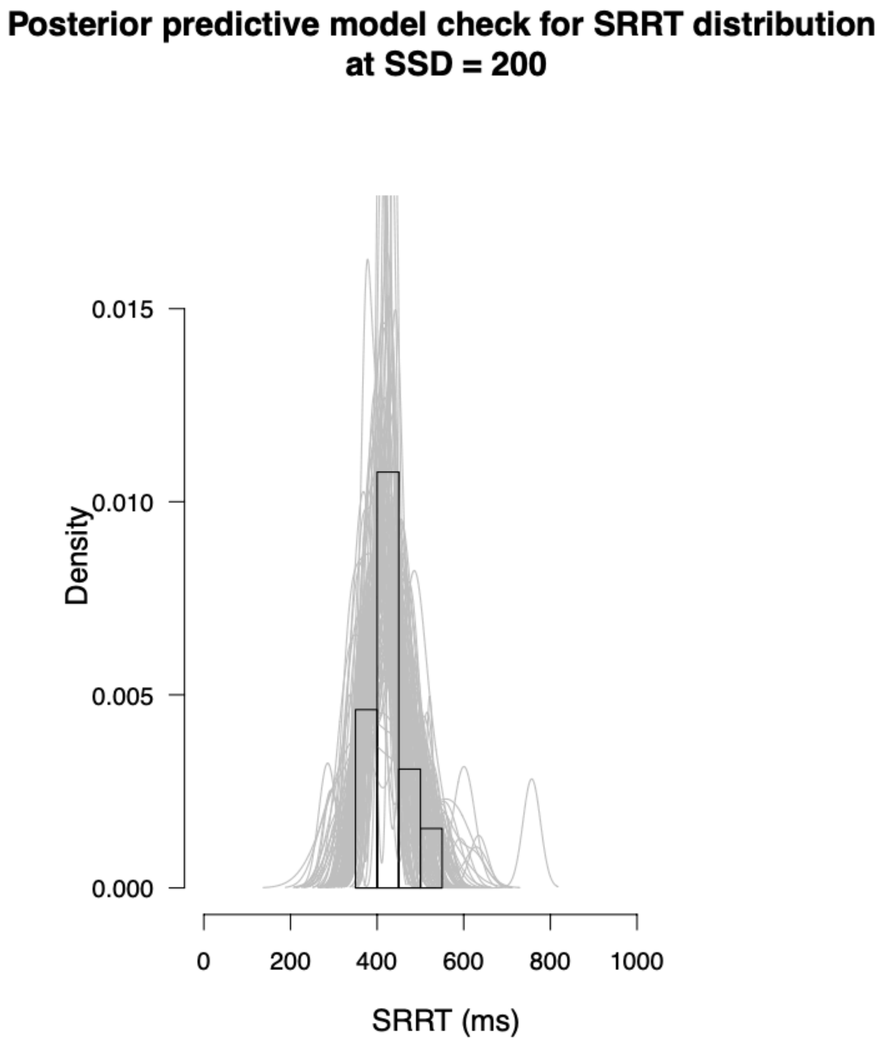


Posterior predictive p values for inhibition function

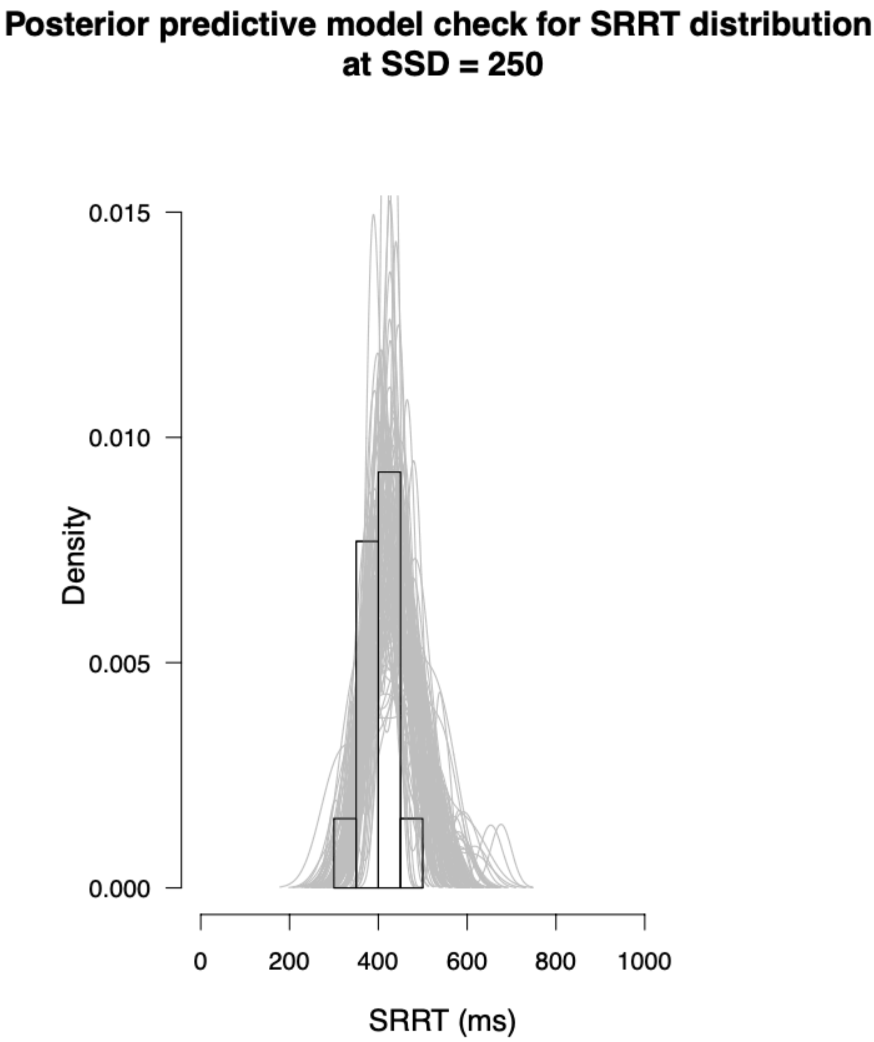
	SSD=50	SSD=100	SSD=150	SSD=200	SSD=250	SSD=300	SSD=350	SSD=40
Number of stop−signal trials	2	7	18	27	23	13	5	1
Observed response rate	0	0.29	0.22	0.48	0.57	0.69	0.8	1
verage predicted response rate	0.06	0.11	0.23	0.42	0.61	0.74	0.85	0.94
One−sided p value	0.11	0.06	0.39	0.27	0.59	0.52	0.4	0.94
Two−sided p value	0.22	0.12	0.78	0.54	0.82	0.96	0.8	0.12

Posterior predictive model check for inhibition function





Posterior predictive model check for SRRT distribution
at SSD = 200
Density
0.015
0.010
0.005
0.000
0
200
400
600
800
1000
SRRT (ms)


Posterior predictive model check for SRRT distribution
at SSD = 250
Density
0.000
0.005
0.010
0.015
0
200
400
600
800
1000
SRRT (ms)