Individual differences in Response Times

Adriana Felisa Chávez De la Peña

Chapter 1:

The study of individual differences.

In this chapter, we talk about the differences between the experimental and differential approaches in psychological research (Cronbach, 1957). We emphasize how these two disciplines diverge in the way they treat between-subject variability, with experimental researchers trying to reduce it through experimental control and differential researchers depending on it to characterize individual differences. Finally, we introduce the concept of reliability and how these contrasting ways of treating between-subject variability have caused a “reliability paradox” where reliable experimental effects fail to produce reliable individual differences.

1.1 On the contrast between experimental and differential Psychology

Psychology is an empirical science mostly founded on evidence collected through a wide range of experimental paradigms and assessment tools. As such, the main goal of cognitive psychology is to build and test cognitive models that can account for the latent cognitive processes generating the data observed.

Depending on the type of questions and methods used to approach them, two main psychological disciplines have been identified (Cronbach, 1957). The first is experimental psychology, where meticulous variable manipulations are implemented in search of systematic effects. The second is correlational psychology –also referred to as differential psychology (Chamorro-Premuzic et al., 2015)–, whose main goal is to characterize the individual variability observed in the data.

> “The well-known virtue of the experimental method is that it brings situational variables under tight control (...) The correlational method, for its part, can study what man has not learned to control or can never hope to control” (Cronbach, 1957).

A key component of any experimental design involves the implementation of experimental controls that minimize the influence of variables other than the independent variable(s) of interest over the dependent variable(s) observed. Experimental research aims to minimize the between-subjects variance in order to ensure that the results observed are a product only of the experimental manipulations imposed by the experimenter. The goal is to identify systematic differences in the data collected between different experimental conditions (Borsboom et al., 2009).

Experimental research is primarily ANOVA-based. The focus is on the comparison between groups or conditions, under the assumption that between-subject variability has been reduced by the experimental design (Draheim et al., 2019). Participants are treated as exchangeable and their aggregate data is used to draw substantial conclusions. In these cases, the mean performance is commonly reported, as it is assumed to remove the effect of any form of measurement error (Lee & Webb, 2018).

Differential psychology –defined by its interest in characterizing differences across individuals–, has mostly been built upon correlational studies. The correlations between measurements taken across subjects are used as an indicator of the existence of consistent individual differences that could be the reflection of individual-level traits (Revelle et al., 2011). While the experimental researcher tries to reduce the between-subject variance, the study of individual differences only makes sense in settings where there is high variability across subjects (Chamorro-Premuzic, et al., 2015).

The study of systematic individual differences is of great interest to many different areas within cognitive science (Cronbach, 1956; Chamorro-Premuzic et al., 2015). For example, Kamphaus et al (1997) found high correlations across different intelligence measures, suggesting that people who score high on one intelligence task will likely also get a high score on different intelligence tasks. Similarly, Cretenoud et al. (2021) report stable individual differences in the perception of optical illusions that remain consistent regardless of exposure time, visual acuity, and whether the illusion is presented in a monocular vs binocular setting (Cappe et al., 2014; Mollon et al., 2017).

Despite the differences between the goals and foci of the experimental and correlational approaches, there is a growing consensus pointing toward the value of reconciling these disciplines by considering individual differences as part of the cognitive models used to account for the data observed in experimental tasks. Aggregate measures are effective in settings where there are small individual differences in performance (i.e., small between-subject variance) because it removes the effect of noise. However, in cases where there's high variability between participants, aggregate measures can lead to faulty conclusions (Estes, 1956). The use of models that incorporate some form of individual-level random effects is necessary to portray the notion that participants are indeed non-interchangeable (Batchelder, 2007). Many cognitive modelers advocate for the implementation of hierarchical models that capture individual differences by incorporating individual-level parameters that can also provide information about the group performance (Lee & Webb, 2005; Rouder & Lu, 2005).

1.2. Reliability

We talk about reliability to refer to the degree to which we can *reliably* obtain the same results under similar conditions, giving us a sense of how much we can trust our findings. In Classical Test Theory, reliability is approximated by computing the correlation between parallel measures (i.e., independent measures related to the same latent construct with equal variance; Lord et al., 1968).

In the context of experimental psychology, we say that an experimental effect is reliable to the extent that similar findings are reported across studies and labs when the experimental design is replicated. In differential psychology, measurements are considered reliable when they produce consistent rank-orderings of individuals across applications (Hedge et al., 2018**;** Draheim et al., 2019).

A counterintuitive but widely reported finding is that robust experimental effects reliably associated with specific experimental paradigms can fail to produce reliable individual differences. There are multiple examples in the literature of experimental effects that appear reliably at the aggregate level, but not at the individual level. As an illustrative example, Draheim et al. (2019) present the results Logie et al. (1996) obtained while exploring the reliability of two empirical effects commonly reported in serial order recall tasks. The authors report robust evidence for the effects of interest at the aggregate level on the test and retest, but the individual performance observed across applications showed weak correlations ($\rho = 0.1$ to 0.3), indicating that performance in the test was not predictive of the performance in the retest.

To make sense of this so-called “reliability paradox”, Hedge et al. (2018) draw our attention back to the distinction between the experimental and differential approaches. Having participants perform similarly is highly desirable for the experimental researcher, but detrimental to the differential researcher interested in characterizing individual differences. Experimental tasks are designed to minimize between-subject variability, which ultimately restrains the correlations one can compute from these measures (Rouder & Haaf, 2019).

The low correlations found between task-derived measures assumed to capture similar underlying processes present a challenge for differential researchers. The following section focuses on the particular case of the study of individual differences in cognitive control (Friedman & Miyake, 2004).

Chapter 2:

Individual differences in cognitive control.

In this chapter, we discuss the study of individual differences in cognitive control. We start by describing what cognitive control is and its role in the regulation of goal-oriented behavior. We present an overview of a few of the best-known tasks used to study cognitive control, and the way in which they are suited to use difference scores –computed from Donders' subtraction method– as dependent measure. We then summarize the evidence reported in the literature for the lack of correlation between difference scores computed across different cognitive control tasks, and discuss the three main ways in which these low correlations have been interpreted: 1) as an indicator of cognitive control not being a psychological construct; 2) as suggesting the need to design different dependent measures to study cognitive control; and 3) as evidence of the limitations of using single point estimates that neglect the information contained in the variability of the data.

2.1. What is cognitive control?

Initially defined as the ability to ignore irrelevant information while solving a task (Logan & Cowan, 1984; Logan, 1985), we also refer to cognitive control as the suppression of prepotent responses (Friedman & Miyake., 2004). Filtering the inputs and responses available when solving a task is crucial to the adaptability of our behavior. This cognitive process has been widely studied across many different contexts of interest within psychology (Rozensky & Bellack, 1974; Stahl et al., 2014; Abrahamse et al., 2016), which has lead to it also been known as inhibition control (Miyake et al., 2000), attention control, cognitive inhibition (Rey-Mermet et al., 2018), interference resolution (Pettigrew & Martin, 2014), and self-regulation (Enkavi et al., 2019).

Miyake et al., (2000) define cognitive as one of three main executive functions that regulate working memory, along with the ability to shift between mental sets or tasks

(Kiesel et al., 2010; Vandierendonck et al., 2010) and the ability to monitor and update the contents of working memory (Ecker et al., 2010; Kessler & Oberauer, 2014). These executive control processes are regarded as pivotal for the regulation of goal-directed behavior (Paap & Sawi, 2016).

In a study with healthy young adults, two components of cognitive control were identified. First we have the response-distracter inhibition, which refers to the ability to suppress available responses and information that is not relevant for the task at hand. The second component is the resistance to proactive interference, which corresponds to the resistance to intrusions from information previously available (Friedman & Miyake, 2004).

Cognitive control has been studied in experimental tasks that incorporate a distinction between "incongruent" trials –where some of the information presented needs to be ignored or an automatic response needs to be suppressed–, and "congruent" trials. The study of cognitive control relies heavily on the differences in response times (RT) across experimental conditions, under the assumption that participants should take longer to respond to incongruent trials, since they require the execution of cognitive control.

2.2 Cognitive control tasks

The best-known task to study cognitive control is the *Stroop task* (Stroop, 1935). Participants are presented with colored color names and asked to indicate the color of the words by pressing distinctive response keys. Throughout the task, participants encounter trials where the meaning of the word and the color in which it is presented match (i.e., congruent trials), and trials where they don't (i.e., incongruent trials), such that only the latter require cognitive control. Variations of the Stroop task have been reported in the literature. For example, in the number-Stroop task participants are asked to identify the number of digits presented on screen, such that the right answer can be congruent or incongruent with respect to the numeric value of the digits (MacLeod, 1991).

Another well-known cognitive control task is the *Flanker task* (Unsworth & Spillers, 2010). Participants have to make a binary judgment about a central stimulus that is surrounded by flanker stimuli that are either congruent or incongruent with the correct response. Variations of this task include different types of stimuli. For example, participants may be asked to indicate the direction of a central arrow, or identify a central letter (Rey-Mermet et al., 2019; Friedman and Miyake, 2004). Rouder and King (2003) introduced an extension of the task where a morphed character between an A and H letter is presented as the central stimulus in a 3-by-3 grid, with the remaining flanker elements being either clear A or H letters.

Another cognitive control task that requires participants to allocate their attention selectively across the visual field is the *Global/local task* (Kinchla, 1974; Navon, 1977). Participants are presented with a large letter composed of small letters that can be congruent or incongruent with respect to the former. The data collected is then analyzed separately depending on whether participants are asked to identify the large letter (global task) or the small letters (local task), with evidence suggesting that larger cognitive control effects occur in the local task (Bruyer & Scailquin, 2000).

As stated at the beginning of this chapter, cognitive control has also been associated with the suppression of prepotent responses. The *Simon task* (Simon & Rudell, 1967) follows from this definition. In the original version of the task, participants heard the words “left” and “right” coming from a left or right speaker, and were instructed to press either a left or right key depending on the meaning of the word and not its source. Multiple variations of the Simon task have been developed using different stimuli modalities (Hommel, 2011; Rey-Mermet et al., 2019).

Another relevant experimental paradigm that revolves around the suppression of preponderant responses is the *Antisaccade task* (Roberts et al., 1994). At the beginning of every trial, participants are presented with a cue stimulus whose location on screen is either congruent (i.e., same) or incongruent (i.e., opposite side) with respect to the location of the target stimulus. The type of trial is explicitly announced to participants, who are then expected to perform a saccade or antisaccade movement towards the target. This is one of the few cognitive control tasks where accuracy measures play a pivotal role in assessing participants' performance. The accuracy of the task has traditionally been assessed using eye-tracking devices (Basanovic et al., 2020), but it has also been measured by requiring participants to make binary judgments about the target's identity. For example, Miyake et al. (2000) implemented a 2AFC paradigm in which participants had to indicate the left/right direction of an arrow shown at the target location.

Although the specific details of the cognitive demands, stimuli, and responses required may be different, all cognitive control tasks previously described share a similar core structure that distinguishes between congruent and incongruent trials, with only the latter enforcing cognitive control. Since cognitive processes are not assumed to occur simultaneously, we expect larger RTs on average in incongruent trials than in congruent trials. Cognitive control tasks use Donders' subtraction to compute difference scores that are used as a dependent variable. These scores are obtained by subtracting the mean RT observed in congruent trials from the mean RT observed in incongruent trials (Donders, 1969).

Common data cleaning steps include the removal of trials with either particularly short or long RTs (i.e., RTs shorter than a couple of hundred milliseconds or larger than a couple of thousand milliseconds), as well as trials where an error was made (Rey-Mermet et al., 2018; Enkavi et al., 2019).

2.3 Donders' subtraction method and difference scores

The subtraction method described by Donders (1969) is rooted in the assumption that effects derived from experimental manipulations can be isolated by computing the difference between scores obtained under said manipulations and scores obtained under identical conditions except for the inclusion of the experimental manipulation (i.e., a baseline score). This subtraction method is often implemented as a difference between means (Draheim et al., 2019). Depending on the context, these difference scores are usually known by domain-specific names (e.g., *change scores*, *gain scores*, *residualized scores*, *cost effects*, *congruence effects*, *discrepancy effects*, *interference effects*, etc.).

Cognitive control tasks use as a dependent measure the difference scores computed from the mean RT observed across the congruent and incongruent conditions. As mentioned before, RTs are expected to be larger on average in incongruent conditions because participants are forced to exert cognitive control. Thus, by subtracting the mean RT observed in the congruent condition from the mean RT observed in the incongruent condition, the cognitive control effect is isolated from any other process elicited by every other aspect of the task (e.g., encoding the sensory input, preparing the motor response, etc.).

2.4 Individual differences in cognitive control

The vast number of cognitive control tasks available in the literature provides an ideal scenario to study individual differences in cognitive control. What can we learn about a person's cognitive control from any of these tasks? Would we be able to predict their performance on other cognitive control tasks, based on the performance observed in a first task? Several studies have addressed these questions by exploring the correlations between the scores observed across different cognitive control tasks (Friedman & Miyake, 2004; Pettigrew & Martin, 2014; Rey-Mermet et al., 2018).

Correlational studies that aim to capture individual differences in cognitive control consistently report low correlations among different cognitive control tasks, which often do not exceed a value of 0.2 (Rouder et al., 2019). The difference scores computed in a Stroop task don't seem to be predictive of the difference scores obtained in the Flanker task (Hedge et al., 2018; Rey-Mermet et al., 2018; Stahl et al., 2014). The correlations reported across cognitive control tasks are considered low in light of what's prescribed by Nunnally (1964), who suggested that correlations as low as 0.70 should only be acceptable in exploratory research, with higher-stake scenarios requiring higher values.

The low correlations reported across cognitive control difference scores have been interpreted in two main ways. First, from a substantive point of view, they have been taken as evidence that cognitive control is not a robust construct and that so-called cognitive control tasks are measuring multiple sources of variation. Second, from a statistical perspective, these correlations have motivated a revision of the reliability of cognitive control scores, ultimately casting doubt on the use of difference scores as a dependent measure.

In a large-scale study conducted by Rey-Mermet et al. (2018) to explore the role of age in cognitive control, a total of 289 participants went through a battery of 11 different cognitive control tasks. In this study, correlations across the difference scores computed for each task were found to range between -0.27 and 0.44. The authors interpreted these results as evidence of cognitive control not being a unified psychometric construct. Along these lines, other authors have suggested that cognitive control is more of a task-specific mechanism, rather than a robust individual trait (Xiao et al., 2022).

As mentioned above, other than casting doubt on the integrity of cognitive control as a psychological construct, the low correlations reported across cognitive control difference scores have also motivated a revision of the methods used to model individual differences in cognitive control, which we will discuss in the following section.

2.5 Correlations across difference scores

The correlation structures of sets of scores or related measures have been used as a tool to capture individual differences in latent traits of interest under two core assumptions. First, it is assumed that all scores included in the set are measuring the same process or a set of subprocesses that are related to the same process. Second, that each of these scores provides reliable information about each individual observed. As such, the low correlations reported across cognitive control difference scores has also been interpreted as suggesting a problem in reliability. Task-specific difference scores would need to be highly reliable in order for them to be able to correlate highly among each other (Rouder & Haaf, 2019).

Hedge et al. (2018) explored the reliabilities of seven different cognitive control tasks by conducting three studies, each of which consisted of a test-retest application of batteries containing between four and six cognitive control tasks. Across these three samples, the authors report that the task with the highest difference score reliability was the Stroop task, with values of 0.6 and 0.66 in the first and second study, respectively. In their literature review, Enkavi et al. (2019) show the distributions of the test-retest reliability indices reported across many published studies for 37 different self-regulation tasks. These distributions included the indices computed by the own authors for 150 participants who completed their test-retest of the same set of tasks. Only seventeen of these 37 tasks provide response time data, and just eight of them rely on difference scores as dependent measure. For the latter subset of tasks, we observe distributions of reliability indices that are highly variable. For example, for the Stroop task, the authors found reliability indices that vary between 0.27 and 0.8, and for the Simon task, reliability indices oscillate between 0 and 1.

In general, the results referred above seem to confirm that difference scores computed in cognitive control tasks are not reliable. A revision of the formula presented by Lord (1963) may serve to illustrate why this may be the case (Draheim et al., 2019; Chiou & Spreng, 1996):

$$\rho_{dd'} = \frac{\frac{(\rho_{xx'}+\rho_{yy'})}{2} - \rho_{xy}}{1-\rho_{xy}}$$

In this equation, $\rho_{dd'}$ refers to the reliability of the difference scores, $\rho_{xx'}$ and $\rho_{yy'}$ correspond to the reliabilities of observations collected across each condition (e.g., in congruent and incongruent trials), with ρ_{xy} capturing the correlation between them. This formulation can be further simplified if we assume that observations collected in both conditions are equally reliable (i.e., $\rho_{xx'} = \rho_{yy'}$ with $\frac{(\rho_{xx'}+\rho_{yy'})}{2} = \rho_{xx'}$). This assumption is supported by the notion that both conditions pertain to the same task.

$$\rho_{dd'} = \frac{\rho_{xx'} - \rho_{xy}}{1-\rho_{xy}}$$

To illustrate the implications of the formula presented above, Draheim et al. (2019) emphasize two scenarios. The first one is a case in which both conditions have perfect reliability ($\rho_{xx'} = \rho_{yy'} = 1$), in which case, the difference score reliability would also be perfect ($\rho_{dd'} = 1$). The second case presented is one in which the measures observed across the two conditions are independent ($\rho_{xy} = 0$), such that the reliability of the difference scores would be equal to the reliability of the scores observed on each condition. However, none of these two conditions seem plausible and thus, the reliability of the difference scores is expected to be lower than the reliability of the observations collected on each condition, and to decrease as the correlation between the scores observed across conditions increases. The latter is particularly problematic because the scores observed across different conditions in a task, are naturally assumed to be correlated.

In cognitive control tasks, a high correlation between the RTs registered across the congruent and incongruent conditions is both expected and observed. This high correlation is assumed to be indicative of a shared, systematic variance that follows from the fact that both conditions present participants with the same task, with the only difference being the additional cognitive control requirement enforced in the incongruent condition. Therefore, the subtraction of one of these measures from the other would leave us with a higher proportion of error variance (Hedge et al., 2018). This is in line with the notion that the difference between any two random variables has a larger variance than any of them individually (Enkavi et al., 2019).

$$Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$$

In the first hypothetical scenario described by Draheim et al. (2019), where observations collected for each condition are perfectly reliable ($\rho_{xx'} = \rho_{yy'} = 1$), we would be left with no error variance. In the second case where observations collected across conditions are independent ($\rho_{xy} = 0$), we would have no systematic variance. Altogether, this has led some authors into arguing against the use of difference scores as a dependent measure (Gollwitzer et al., 2014; see Drahem et al., 2019 for a collection of different quotes).

2.6 Alternatives

The discussion of the problems associated with the use of difference scores invites the question: *What alternatives are there to study cognitive control?* Perhaps the most intuitive answer to this question may be to use accuracy measures instead of RTs. However, this approach does not imply a smaller number of complications. On the one hand, accuracy rates are known to be susceptible to ceiling and floor effects and on the other hand, there is evidence of learning across trials that result in performance improvement during the execution of cognitive control tasks (Reisberg et al., 1980; Bustamante et al., 2021).

A second alternative would be to develop a dependent measure other than difference scores that ideally captures jointly participants’ accuracy and RTs. Along these lines, Draheim et al., 2016 propose a new type of scores computed with the following binning procedure: First, subtract the mean RT of every participant on only correct congruent trials from the RT observed on every correct incongruent trial. Then, the resulting differences get sorted across all participants and are transformed into a scale from one to ten based on their decile. The authors suggest that the sum of the decile points accumulated per participant and the number of incorrect incongruent trials times an arbitrary value of 20 can be used as a new dependent score. The proposed binning procedure has been applied to data collected across different studies, resulting in an improvement in the overall correlations found between these new scores and scores obtained reported on other working memory tasks. For a revision on different dependent scores proposed to integrate participants' accuracy and speed, see Vandierendonck (2017, 2018).

A third alternative to the problem presented by the use of difference scores to study cognitive control involves moving away from the use of point statistics to the analysis of the response time distributions observed across all congruent and incongruent trials. This alternative would be the focus of the following section.

Chapter 3:

Response time modeling

In this chapter, we present an overview of different models commonly used in psychological science when working with response time (RT) data. In doing so, we distinguish between two types of RT models. First, we discuss some simple RT models used to capture the quantitative properties of empirical RT distributions. Then, we describe a few choice RT models that jointly describe the RTs and choices observed. For this latter category, we focus our discussion on sequential sampling models, a family of procedural cognitive models that follow from the core assumption that participants accumulate evidence in favor of each of the response alternatives as time goes by, until a response threshold is reached and a choice is made, thus explaining the variability in the RTs observed across trials.

3.1 Simple response time models

One fact about RTs that seems to hold true across the animal kingdom is that they are highly variable (Noorani & Carpenter, 2016), even in the context of very simple RT tasks where participants are instructed to press a button every time a stimulus is presented. This high variability presents a major advantage when placed in contrast with how limited is the information one can gain from the choices observed in an experimental set up, which may be reduced to frequencies, proportions or binary accuracy measures. The analysis of empirical RT distributions has played a major role in cognitive science (Hohle, 1965; Luce, 1986; Ratcliff, 1978; Dutilh et al., 2019), with many authors suggesting that characteristics of these distributions beyond measures of central tendency can shed light on the decision and cognitive processes that underlie the choices observed (Heathcote et al., 1991; Spieler et al., 2000).

The use of point estimates such as the mean or median RT to describe participants' performance in experimental tasks neglects a great deal of the information conveyed by the variability of RTs observed throughout the task. As an example, we will briefly discuss the *worst performance rule* (Larson & Alderton, 1990), which states that the slower RTs observed in a cognitive task are more informative of underlying cognitive processes than faster RTs. This claim was based on the fact that when we sort and assign the RTs observed across all individuals to different bands, the correlation between the median RT computed across every band and measures of general intelligence, working memory and processing speed increases as we move from the fastest to the shortest RT band (Coyle, 2003). While some authors have interpreted this empirical finding as indicative of how mechanisms of working memory and attention operate, some other authors have emphasized how taking the median across all RT bands neglects the fact that we expect much greater variance for slower RTs, since empirical RT distributions are known to be positively skewed (Ratcliff, 1979; Schmiedek et al., 2007).

The use of mathematical models suited to describe empirical RT distributions has proven useful to map quantitative properties such as their shift, scale and shape onto specific experimental manipulations (Andrews & Heathcote, 2001; Heathcote et al., 1991; Hockley, 1984; Hohle, 1965; Spieler et al., 2000). Next we present a general overview of some of the best-known mathematical and statistical models used to describe RT distributions (Luce, 1986).

3.1.1 exGaussian distribution

The exGaussian distribution (also known as the *exponentially modified Gaussian distribution*) is a statistical model that describes the probability density function of the sum of a normal random variable and an exponential random variable. The ex-Gaussian distribution is characterized by three parameters: the mean (μ) and standard deviation (σ) of the normal component, and the rate (τ) of the exponential component. This distribution is known to be positively skewed with a slow decay, which makes it suitable to model empirical RT distributions (Heathcote et al., 1991; Hockley, 1984; Mewhort et al., 1992; Ratcliff, 1979).

In terms of the substantive interpretation of the parameters defining the exGaussian distribution, we note that parameters related to the normal component have been associated with peripheral sensory–motor and automatic processes, while the rate parameter associated with the exponential component has been found to be related to decision-related processes (Hohle, 1965; Gordon & Carson, 1990; Rohrer & Wixted, 1994). The relationship between the exponential component and attention-demanding processes has been supported by studies on individual differences that compare groups of participants who are assumed to differ in their attentional capacities, for example, based on their age (Spieler et al., 1996; West et al., 2002).

The exGaussian parameters have also been interpreted in relation to decision models that also take into account RTs. We briefly discuss the case of the drift diffusion model that will be covered in greater detail later in our revision, (Ratcliff, 1978; Ratcliff & Rouder, 1998). Simulation studies have shown that faster drift rates can be associated with smaller values of τ, while more conservative response criteria seem to be linked with larger values of μ, (Spieler, 2001; Schmiedek et al., 2007).

3.1.2. logNormal distribution

The logNormal distribution is a statistical model that describes the probability density function associated with random variables whose logarithm is known to be normally distributed. From this definition follows that the exponential transformation of any normally distributed random variable has a logNormal distribution. The logNormal distribution is positively skewed with a slow decay and it has been widely used to model RT data and to make inferences about the underlying decision mechanisms of several different behavioral tasks (van der Linden, 2006; Rouder et al., 2015; Becker et al., 2021).

3.2 Choice response time models

The notion that RTs can be regarded as the stopping points of an internal process producing the choices we observe, plays a pivotal role in decision making (Noorani & Carpenter, 2016). As phrased by Forstmann et al. (2016) "...most real-life decisions are composed of two separate decisions: first the decision to stop deliberating and act, and then the decision or act itself", with the former being heavily determined by the time constraints imposed by the task. To account for the speed-accuracy tradeoff associated with these two decision components, several procedural models known as accumulator or sequential sampling models have been developed in mathematical psychology (Stone, 1960; Ratcliff & Smith, 2004).

Accumulator models are developed from the core intuition that evidence for any of the response alternatives is accumulated over time until a response threshold is reached, accounting jointly for the responses and RTs observed. This type of model is useful both in the context of decisions where information is made available gradually (e.g., in a court trial) or simultaneously (e.g., in a match of chess). In the present section, we introduce some of the best-known sequential sampling models.

3.2.1 Linear approach to threshold with ergodic rate model

The linear approach to threshold with ergodic rate (LATER) model (Carpenter, 1981; Noorani & Carpenter, 2016) is a model with only two free parameters that was originally conceived as a descriptive model of observed RT distributions which eventually started being regarded as a normative decision model that can predict and account for the underlying mechanisms of the data collected (i.e., in terms of decision-making and cognitive processes).

The LATER model finds its roots on the empirical finding that the inverse of the RTs follow a normal distribution, suggesting that RT data can be described using only two parameters (i.e., μ and σ). RTs are then conceptualized as the finishing point of an internal process that occurs at a given rate until completion.

In the LATER model, we assume that every stimuli S is associated with an initial value S_0 that works as the starting point of an internal process that progresses linearly with constant rate r until a threshold $S_T = S_0 + \theta$ is met, signaling the moment in which the response is initiated. According to this account, the RTs observed are given by $T = \frac{\theta}{r}$, and since r is assumed to vary across trials following a normal distribution with mean μ and standard deviation σ, it follows that $\frac{1}{T}$ would also be normally distributed, thus replicating a well-known property of empirically observed RTs distributions (Noorani & Carpenter, 2016).

It is worth noticing that unlike many other theoretical RT models that incorporate explicit assumptions about the internal processes associated with the RTs observed, the LATER model does not make a distinction between time spent deliberating in the task and nondecision time (Noorani & Carpenter, 2015). Nondecision time is gathered as a small but fixed amount of time (between 50 and 60 ms) that should not affect the linearity of the internal process described by the LATER model and that doesn't provide any meaningful or additional information about behavior.

The LATER model has proven useful in clinical settings where it has been found to produce precise predictions on quantitative measures of pathological progression for a wide range of neurological conditions (Antoniades et al., 2012; Burrell et al., 2013; Pearson et al., 2007; Smyrnis et al., 2009).

3.2.2 Linear ballistic accumulation model

The Linear Ballistic Accumulation (LBA) model is an extension of the LATER model (Noorani & Carpenter, 2016) with an additional parameter. The LBA model is a sequential sampling model that assumes that evidence is accumulated as a linear and continuous process that continues until a response threshold is met (Brown & Heathcote, 2008).

The LBA model attributes the variability in the responses and RTs observed to trial-by-trial variability of the accumulation rate (now referred to as *drift rate,* d), with a separate and independent accumulator for each response alternative, such that the response observed corresponds to the accumulator that reaches some threshold value (b) first. Each accumulator process starts with some initial amount of information, which is also assumed to vary across trials following a uniform distribution with bounds 0 and a (with $a < b$), while the drift rate is assumed to be normally distributed (Annis et al., 2017).

A key distinction between the LBA model and the LATER model is that the former incorporates a parameter t that captures the nondecision time that can be attributed to processes other than the accumulation of information (i.e., stimuli encoding and motor responses). This nondecision time is assumed to be constant across trials.

3.2.3 Drift diffusion model

The drift-diffusion model (DDM) is a decision-making model that accounts for the responses and RTs observed in two alternative forced choice tasks (2AFC) as the result of a stochastic sequential sampling process (Ratcliff, 1978; Ratcliff & Rouder, 1998). The DDM provides a theoretical framework on the mechanisms that underlie the data collected that has proven to be consistent with the distinctive characteristics of the RT distributions observed for correct and incorrect responses.

The DDM model is a random walk model that assumes that participants start accumulating information over time, from the moment a new stimulus is presented until one of two decision boundaries is reached and a response is produced. The DDM considers four parameters: 1) The starting point of the information accumulation process with respect to the two response boundaries; 2) The nondecision time; 3) The response criterion, defined as the distance between the two response boundaries around the starting point; 4) The drift rate, regarded as the average amount of evidence sampled per unit of time.

The parameters of the DDM capture relevant aspects of the decision process that underlies the choices and RTs observed, such that specific experimental manipulations have been associated with consistent changes in the parameters of the model. The response criterion parameter indicates how cautious participants are being while responding to the task and it has been found to be linked to the speed/accuracy tradeoff imposed by the task. The drift rate parameter corresponds to the speed of information accumulation and is known to be related to the quality of the information conveyed by the stimuli presented and to individual differences in processing efficiency. Finally, the starting point is assumed to be a reflection of the participants initial bias towards one of the two response choices.

Along with the substantially relevant parameters included in the model, another key aspect to the DDM is the notion of trial by trial variability that is assumed across all of them, and which allow the model to account for the interactions between accuracy and the RT distributions (Ratcliff & Rouder, 1998).

3.2.4 Circular drift diffusion model

The circular drift-diffusion model (CDDM) is an extension of the DDM where instead of a binary choice scenario, we deal with decisions that take place in a circular space (i.e., a bounded continuum; Smith, 2016). Real life examples of this kind of decisions can be found anywhere, from the spatial location of relevant auditory stimuli in our surroundings, to judgments about the occurrence of cyclic phenomena (e.g., day-night cycles or the seasons of the year; Chávez De la Peña et al., 2023). The CDDM is a bivariate model that describes both the response times and the choices made by participants, with the latter expressed in radians.

The CDDM is a stochastic sequential sampling model that assumes that information gets accumulated over time, moving from a starting point located at the origin of the circle used to represent the decision space, towards the circumference. The CDDM considers four parameters. First, we have two parameters commonly used in sequential sampling models: 1) The nondecision time and 2) the response criterion (i.e., the radius of the circle). Then, we have a pair of parameters related to the information provided by the stimulus and its effect on the decision process. These last two parameters describe the overall direction and speed of the random walk either in terms of cartesian or polar coordinates. If expressed as a cartesian coordinates, parameters μ_1 and μ_2 are used to denote the average step size that the random walk takes per unit of time along the X and Y axes, respectively. Whereas, if expressed as polar coordinates, parameters θ and δ are used to describe the angle that corresponds to the correct answer (i.e., drift angle) and the average speed at which the random walk approaches the circumference (i.e., drift magnitude), respectively (Smith, 2016; Kvam, 2019). Going back and forth between these two coordinate systems is as straightforward as indicated by the following system of equations:

$$\delta = \sqrt{\mu_1^2 + \mu_2^2}$$

$$\theta = arctan(\frac{\mu_2}{\mu_1})$$

$$\{\mu_1 , \mu_2\} = \{\delta \times cos(\theta) , \delta \times sin(\theta)\}$$

Chapter 4:

Cognitive latent variable modeling of response times

In this chapter, we talk about latent variable models (LVM) and their role in psychological science as tools to uncover the latent variables that account for the data observed, and their levels of variation (i.e., per subject, per condition, etc.). Then, we present cognitive psychometrics as an area of convergence between cognitive modeling and the application of LVMs, where the goal is to unveil the latent variable structure of the parameters recovered from a given cognitive model of interest. We briefly discuss examples of the application of cognitive psychometrics in the literature, which are usually performed in two steps (i.e., first, estimating the parameter values of the cognitive model and then fitting a LVM to unveil the underlying structure to these estimates), before introducing cognitive latent variable models (CLVM). This last type of model follows the principles of cognitive psychometrics as well, but it does it as part of a single, hierarchical Bayesian structure (i.e., parameter values and their latent variable structure are recovered as part of a single model). We emphasize the advantage presented by this distinctive quality, which allows for the uncertainty in the data to inform the inferences made at the level of both, the cognitive model parameters and their interpretation, and the latent variables identified as explaining the variability across subjects, conditions, tasks, and so on.

4.1 Latent variable models and psychometrics

The models described in previous chapters are identified as cognitive models because they account for the psychological processes that underlie the data observed. In contrast, psychometric models, or LVMs, constitute a type of model that assumes that the data observed results from the linear combination of a small number of latent variables that vary across different random levels (i.e., across participants, conditions, tasks, etc.) and some measurement error (Baribault, 2019).

LVMs are used to identify the components that can account for the data observed and their different levels of variation. This type of model allows us to determine whether variability in the data can be attributed to differences between participants or to the effect of specific features of the experimental design (i.e., experimental conditions, stimuli characteristics, levels of an experimental manipulation, etc.).

An important difference between cognitive models and LVMs is that the latter do not assume that every latent variable associated with the data has a relevant psychological interpretation. LVMs are often implemented with the goal of identifying the smallest number of substantially relevant latent components that can be used to account for the majority of the variability contained in the data (Bollen, 2002).

4.2 Cognitive psychometrics and individual differences

We use the term cognitive psychometrics to refer to the psychometric treatment of cognitive models (Batchelder, 1998; Batchelder, 2010) with the intention of identifying whether the cognitive processes described by cognitive models vary across participants, items, conditions, tasks, and so on. For example, clinical researchers aim to identify how individuals from different populations may vary in terms of the parameter values estimated from cognitive models used to account for their performance in cognitive tasks, and how these changes may be elicited over time or across conditions.

Using LVM techniques onto parameters associated with a cognitive model presents a reconciliation point among experimental and differential research (Cronbach, 1957; Borsboom, 2006). The main advantage of cognitive psychometrics is that it preserves the best of both worlds: First, parameters derived from cognitive models are known to be informative of the cognitive processes underlying behavior and thus are highly interpretable. Second, we move away from the use of aggregated data to recognize the richness of the information provided by the data variability, considering differences in performance across many possible levels of random variation (Riefer et al., 2002).

As an example of the application of the cognitive psychometrics approach, we discuss the study conducted by Schmiedek et al. (2007). In this study, eight different choice response time tasks were applied along with six tasks used to measure working memory and intelligence. From the data collected, an exGaussian model was fit to the RT data and parameters of the DDM were estimated using both RTs and choice data. Then, the authors performed confirmatory factorial analysis techniques on all parameter values obtained, and report that individual differences in the rate parameter of the exGaussian distribution (i.e., τ) were found to be a unique predictor for the individual differences in the performance observed across the different working memory and reasoning tasks included in the study. Along these lines, we can also find studies on the latent variable structure of cognitive control from the perspective of cognitive psychometrics (Miyake et al., 2000; Friedman and Miyake, 2004).

4.3 Cognitive latent variable models

The joint approach postulated by cognitive psychometrics allows researchers to make conclusions far more informative than either a cognitive model or an LVM could afford separately. LVMs are regarded as informative of the internal structure of the data, but these types of models provide very limited insights into the underlying processes and mechanisms that could be generating the data, while cognitive models are limited in their ability to identify higher-order latent structures.

Despite the advantages presented by cognitive psychometrics in terms of the enhanced insights we can gain from the data collected about the underlying processes of interest, the applications described in the previous section present some major limitations. The studies conducted by Friedman and Miyake in 2004 and by Schmiedek et al. in 2007 consist of a two-step application of the joint modeling technique postulated by the cognitive psychometrics approach. Both studies perform the parameter estimation from the cognitive model of interest first, and then in a second step, they use latent variable modeling techniques to describe the latent variable structure of these parameter values. This two-step approach is limited because the variability contained in the data is only used to inform the parameter estimation, but it is neglected in the identification of their latent variable structure. Thus, the latent variable structure revealed in the second step considers only the variability in the parameter values recovered, while ignoring the uncertainty contained in the raw data.

As an alternative, Vandekerckhove (2014) introduced CLVMs as a class of models that explore the latent variable structure of the parameters of cognitive models, all as part of a hierarchical Bayesian model. As an illustrative example, the author presents a CLVM that uses the DDM as the main cognitive model, such that by imposing a latent variable model to describe the parameters of the DDM we are now able to draw substantive conclusions about the latent variables that could be explaining the differences observed in processing speed, response caution, and son on, that are made evident if we simply compare the parameter values recovered across different possible levels of variation.

CLVMs allow to make inferences about the unobserved cognitive processes generating the data collected in terms of the influence of many unobserved latent variables. A key feature of CLVM is that both the cognitive model parameters and their latent variable structure is being inferred as part of a single hierarchical Bayesian model. This feature comes with the huge advantage that the uncertainty contained in the data is captured and allocated across both model components (Vandekerckhove, 2014; Baribault, 2019).

Chapter 5:
Research proposals

In this chapter, we present a brief description of the main research projects that have been derived from the different topics discussed in the previous chapters. The first two projects are currently in the finishing stages and have already been presented at the 2022 Annual Meeting of the Society of Mathematical Psychology and the and the sixth European Summer School on Computational and Mathematical Modeling of Cognition, respectively. The last project here presented is still just a proposal.

5.1 Exploring the unidimensionality of processing speed

As mentioned in Chapter 2, a consistent finding in the literature on individual differences in cognitive control is that difference scores derived from cognitive control tasks are very weakly correlated. As part of a first study, we revisit nine publicly available data sets where batteries of cognitive control tasks are applied and find that the overall mean RTs (i.e., *processing speed*) do correlate across tasks at over .5 in value. This result implies that participants are consistently fast or slow to respond across tasks. The main source of individual variation is not cognitive control, but rather processing speed.

We explore the dimensionality and structure of processing speed across individuals and tasks in extended data sets. With several tasks per set, it is possible to ask whether there is a unified processing speed versus several varieties of processing speed. A principal component analysis (PCA) revealed a strong first factor in all sets, consistent with a unidimensional, unified construct of processing speed.

One way of contextualizing these results is to compare them to human anthropometrics. While human bodies are similar in many ways, they seemingly vary on a “size” factor. We analyze a publicly available set of 93 body measurements collected across 6,068 US military personnel. Indeed, a strong first factor of size emerges, but so does a second factor that captures how heavy people are for their height. Perhaps surprisingly, the first-factor solution for processing speed is comparable to, or even stronger than it is for anthropometrics. Moreover, we were unable to identify a coherent second factor for processing speed. We conclude that general speed is likely unidimensional.

5.2 exGaussian CLVM to explore individual differences in response times

In a second study, we explore the univariance found in processing speed across different cognitive control tasks through the lens of RT distribution analysis. The goal is to extend the PCA analysis to determine whether univariance is found in the shift, scale, or shape parameters of the RT distribution. In doing so, we hope to provide insight into whether the single underlying dimension behaves more like a decisional component versus a non-decisional component.

5.3 Thurstonian drift-diffusion model

As a third study, we propose to work on an extension of the CDDM where the bounded continuous decision space is segmented into bins (i.e., a Thurstonian Drift-Diffusion model). This project is motivated by the notion that although circular decisions live in a bounded continuous space, the space of possible responses that any individual would consider in a real life setting would very likely be discretized. For example, when trying to determine the location of an auditory stimuli, we may not have a need to express the location of the source in terms of radians, but rather in terms of discrete labels such as “left” or “right”.

References

1. Abrahamse, E., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. Psychological Bulletin, 142(7), 693.
2. Andrews, S., & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 514.
3. Annis, J., Miller, B. J., & Palmeri, T. J. (2017). Bayesian inference with Stan: A tutorial on adding custom distributions. *Behavior research methods*, *49*, 863-886.
4. Antoniades, C. A., Carpenter, R. H., & Temel, Y. (2012). Deep brain stimulation of the subthalamic nucleus in Parkinson's disease: similar improvements in saccadic and manual responses. *Neuroreport*, *23*(3), 179-183.
5. Baribault, B. (2019). *Using Hierarchical Bayesian Models to Test Complex Theories About the Nature of Latent Cognitive Processes* (Doctoral dissertation, UC Irvine).
6. Basanovic, J., Myles, O., & MacLeod, C. (2020). Do the eyes have it? A comparison of eye-movement and attentional-probe-based approaches to indexing attentional control within the antisaccade paradigm. Quarterly Journal of Experimental Psychology, 17470218221083556.
7. Batchelder, W. H. (2007). Cognitive psychometrics: Combining two psychological traditions. *CSCA Lecture, Amsterdam, The Netherlands*, *126*.
8. Batchelder, W. H. (2016). Cognitive Psychometrics. In *Mathematical Models of Perception and Cognition Volume I* (pp. 259-280). Psychology Press.
9. Becker, B., Debeer, D., Weirich, S., & Goldhammer, F. (2021). On the speed sensitivity parameter in the lognormal model for response times and implications for high-stakes measurement practice. *Applied Psychological Measurement*, *45*(6), 407-422.
10. Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual review of psychology*, *53*(1), 605-634.

11. Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*(3), 425-440.
12. Borsboom, D., Kievit, R. A., Cervone, D., & Hood, S. B. (2009). The two disciplines of scientific psychology, or: The disunity of psychology as a working hypothesis. In *Dynamic process methodology in the social and developmental sciences* (pp. 67-97). Springer, New York, NY.
13. Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive psychology*, *57*(3), 153-178.
14. Bruyer, R., & Scailquin, J. C. (2000). The fate of global precedence with age. *Experimental aging research*.
15. Burrell, J. R., Carpenter, R. H., Hodges, J. R., & Kiernan, M. C. (2013). Early saccades in amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *14*(4), 294-301.
16. Bustamante, L., Lieder, F., Musslick, S., Shenhav, A., & Cohen, J. (2021). Learning to overexert cognitive control in a Stroop task. *Cognitive, Affective, & Behavioral Neuroscience*, *21*(3), 453-471.
17. Cappe, C., Clarke, A., Mohr, C., & Herzog, M. H. (2014). Is there a common factor for vision?. *Journal of vision*, *14*(8), 4-4.
18. Carpenter, R. H. S. (1981). Oculomotor procrastination,‖ in Eye Movements: Cognition and Visual Perception, DF Fisher, RA Monty, and JW Senders Eds.
19. Chamorro-Premuzic, T., Von Stumm, S., & Furnham, A. (Eds.). (2015). *The Wiley-Blackwell handbook of individual differences*. John Wiley & Sons.
20. Chávez De la Peña, A.F., Villarreal, M., Vandekerckhove, J. & Lee, M. D. (2023, March 2–4). *Drifting Beyond the Bayesics* [Conference presentation]. The 58 Edwards Bayesian Research Conference, Fullerton, CA, United States.
21. Chiou, J. S., & Spreng, R. A. (1996). The reliability of difference scores: A re-examination. *The Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, *9*, 158-167.
22. Coyle, T. R. (2003). A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence*, *31*(6), 567-587.

23. Cretenoud, A. F., Grzeczkowski, L., Kunchulia, M., & Herzog, M. H. (2021). Individual differences in the perception of visual illusions are stable across eyes, time, and measurement methods. *Journal of vision*, *21*(5), 26-26.
24. Cronbach, L. J. (1956). Assessment of individual differences. *Annual review of psychology*, *7*(1), 173-196.
25. Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American psychologist*, *12*(11), 671.
26. Donders, F. C. (1969). Over de snelheid van psychische processen [On the speed of mental processes], Utrecht, 1868,(Translated by WG Koster). In *Attention and Performance II*. North-Holland Amsterdam.
27. Draheim, C., Hicks, K. L., & Engle, R. W. (2016). Combining reaction time and accuracy: The relationship between working memory capacity and task switching as a case example. *Perspectives on Psychological Science*, *11*(1), 133-155.
28. Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin*, *145*(5), 508.
29. Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P., ... & Donkin, C. (2019). The quality of response time data inference: A blinded, collaborative assessment of the validity of cognitive models. *Psychonomic bulletin & review*, *26*, 1051-1069.
30. Ecker, U. K., Lewandowsky, S., Oberauer, K., & Chee, A. E. (2010). The components of working memory updating: an experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 170.
31. Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences*, *116*(12), 5472-5477.
32. Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological bulletin*, *53*(2), 134.

33. Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of experimental psychology: Genera*l, 133(1), 101.
34. Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual review of psychology*, *67*, 641-666.
35. Gollwitzer, M., Christ, O., & Lemmer, G. (2014). Individual differences make a difference: On the use and the psychometric properties of difference scores in social psychology. European Journal of Social Psychology, 44(7), 673-682.
36. Gordon, B., & Carson, K. (1990). The basis for choice reaction time slowing in Alzheimer's disease. *Brain and Cognition*, *13*(2), 148-166.
37. Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior research methods*, *50*(3), 1166-1186.
38. Heathcote, A., Popiel, S. J., & Mewhort, D. J. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological bulletin*, *109*(2), 340.
39. Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 598.
40. Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of experimental psychology*, *69*(4), 382.
41. Hommel, B. (2011). The Simon effect as tool and heuristic. Acta psychologica, 136(2), 189-202.
42. Kamphaus, R. W., Winsor, A. P., Rowe, E. W., Kim, S., Flanagan, D. P., & Harrison, P. L. (1997). A history of intelligence test interpretation. *Contemporary intellectual assessment: Theories, tests, and issues*, 32-47.
43. Kessler, Y., & Oberauer, K. (2014). Working memory updating latency reflects the cost of switching between maintenance and updating modes of operation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 738.

44. Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological bulletin*, *136*(5), 849.
45. Kinchla, R. A. (1974). Detecting target elements in multielement arrays: A confusability model. *Perception & psychophysics*, *15*, 149-158.
46. Kvam, P. D. (2019). Modeling accuracy, response time, and bias in continuous orientation judgments. Journal of Experimental Psychology: Human Perception and Performance, 45(3), 301–318.
47. Larson, G. E., & Alderton, D. L. (1990). Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence*, *14*(3), 309-325.
48. Lee, M. D., & Webb, M. R. (2005). Modeling individual differences in cognition. *Psychonomic Bulletin & Review*, *12*(4), 605-621.
49. Logie, R. H., Sala, S. D., Laiacona, M., Chalmers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Mem*ory & Cognition, 24(3), 305-321.
50. Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. Psychological review, 91(3), 295.
51. Logan, G. D. (1985). Executive control of thought and action. Acta psychologica, 60(2-3), 193-210.
52. Lord, F. M. (1963). Elementary models for measuring change. *Problems in measuring change*, 21-38.
53. Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). Statistical theories of mental test scores.
54. Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization* (No. 8). Oxford University Press on Demand.
55. MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. Psychological bulletin, 109(2), 163.
56. Mewhort, D. J., Braun, J. G., & Heathcote, A. (1992). Response time distributions and the Stroop Task: a test of the Cohen, Dunbar, and McClelland (1990) model. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 872.

57. Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive psychology*, *41*(1), 49-100.
58. Mollon, J. D., Bosten, J. M., Peterzell, D. H., & Webster, M. A. (2017). Individual differences in visual science: What can be learned and what is good experimental practice?. *Vision Research*, *141*, 4-15.
59. Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive psychology*, *9*(3), 353-383.
60. Noorani, I., & Carpenter, R. H. S. (2015). Ultrafast initiation of a neural race by impending errors. *The Journal of Physiology*, *593*(19), 4471-4484.
61. Noorani, I., & Carpenter, R. H. S. (2016). The LATER model of reaction time and decision. *Neuroscience & Biobehavioral Reviews*, *64*, 229-251.
62. Nunnally, J. C. (1964). Educational Measurement and Evaluation. New York! McGraw-Hill Book Company.
63. Paap, K. R., & Sawi, O. (2016). The role of test-retest reliability in measuring individual and group differences in executive functioning. Journal of Neuroscience Methods, 274, 81-93.
64. Pearson, B. C., Armitage, K. R., Horner, C. W. M., & Carpenter, R. H. S. (2007). Saccadometry: the possible application of latency distribution measurement for monitoring concussion. *British journal of sports medicine*, *41*(9), 610-612.
65. Pettigrew, C., & Martin, R. C. (2014). Cognitive declines in healthy aging: evidence from multiple aspects of interference resolution. Psychology and aging, 29(2), 187.
66. Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, *85*(2), 59.
67. Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological bulletin*, *86*(3), 446.
68. Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological science*, *9*(5), 347-356.
69. Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological review*, *111*(2), 333.

70. Reisberg, D., Baron, J., & Kemler, D. G. (1980). Overcoming Stroop interference: the effects of practice on distractor potency. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 140.
71. Revelle, W., Wilt, J., & Condon, D. M. (2011). Individual differences and differential psychology: A brief history and prospect.
72. Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(4), 501.
73. Riefer, D. M., Knapp, B. R., Batchelder, W. H., Bamber, D., & Manifold, V. (2002). Cognitive psychometrics: assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological Assessment*, *14*(2), 184.
74. Roberts, R. J., Hager, L. D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General*, *123*(4), 374.
75. Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition*, *22*(5), 511-524.
76. Rouder, J. N., & King, J. W. (2003). Flanker and negative flanker effects in letter identification. *Perception & psychophysics*, *65*(2), 287-297.
77. Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic bulletin & review*, *12*(4), 573-604.
78. Rouder, J., Kumar, A., & Haaf, J. M. (2019). Why most studies of individual differences with inhibition tasks are bound to fail.
79. Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic bulletin & review*, *26*(2), 452-467.
80. Rouder, J. N., Province, J. M., Morey, R. D., Gomez, P., & Heathcote, A. (2015). The lognormal race: A cognitive-process model of choice and latency with desirable psychometric properties. *Psychometrika*, *80*, 491-513.

81. Rozensky, R. H., & Bellack, A. S. (1974). Behavior change and individual differences in self-control. Behaviour research and therapy, 12(3), 267-268.
82. Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H. M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of experimental psychology: General*, *136*(3), 414.
83. Simon, J. R., & Rudell, A. P. (1967). Auditory SR compatibility: the effect of an irrelevant cue on information processing. *Journal of applied psychology*, *51*(3), 300.
84. Smith, P. L. (2016). Diffusion theory of decision making in continuous report. *Psychological Review*, *123*(4), 425.
85. Smyrnis, N., Karantinos, T., Malogiannis, I., Theleritis, C., Mantas, A., Stefanis, N. C., ... & Evdokimidis, I. (2009). Larger variability of saccadic reaction times in schizophrenia patients. *Psychiatry research*, *168*(2), 129-136.
86. Spieler, D. H. (2001). Modelling age-related changes in information processing. *European Journal of Cognitive Psychology*, *13*(1-2), 217-234.
87. Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human perception and performance*, *22*(2), 461.
88. Spieler, D. H., Balota, D. A., & Faust, M. E. (2000). Levels of selective attention revealed through analyses of response time distributions. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 506.
89. Stahl, C., Voss, A., Schmitz, F., Nuszbaum, M., Tüscher, O., Lieb, K., & Klauer, K. C. (2014). Behavioral components of impulsivity. *Journal of Experimental Psychology: General*, *143*(2), 850.
90. Stone, M. (1960). Models for choice-reaction time. *Psychometrika*, *25*(3), 251-260.
91. Stroop, J. R. (1935). Studies of interference in serial verbal reactions. Journal of experimental psychology, 18(6), 643.

92. Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*(4), 392-406.
93. van der Linden, W. J. (2016). Lognormal response-time model. In *Handbook of Item Response Theory, Volume One* (pp. 289-310). Chapman and Hall/CRC.
94. Vandekerckhove, J. (2014). A cognitive latent variable model for the simultaneous analysis of behavioral and personality data. Journal of Mathematical Psychology, 60 , 58–71.
95. Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task switching: interplay of reconfiguration and interference control. *Psychological bulletin*, *136*(4), 601.
96. Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior research methods*, *49*(2), 653-673.
97. Vandierendonck, A. (2018). Further tests of the utility of integrated speed-accuracy measures in task switching. *Journal of cognition*, *1*(1).
98. West, R., Murphy, K. J., Armilio, M. L., Craik, F. I., & Stuss, D. T. (2002). Lapses of intention and performance variability reveal age-related increases in fluctuations of executive control. *Brain and cognition*, *49*(3), 402-419.
99. Xiao, Y., Chou, C. C., Cosgrove, G. R., Crone, N. E., Stone, S., Madsen, J. R., ... & Kreiman, G. (2022). Task-specific neural processes underlying conflict resolution during cognitive control. bioRxiv.