# THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 30 mars 2021 par

# Gabriel WEINDEL

# On the measurement and estimation of cognitive processes with electrophysiological recordings and reaction time modeling

**Disciplines**
Cognition, Langage et éducation
Sciences de la vie et de la santé

**Spécialités**
Psychologie
Neurosciences

**Écoles doctorales**
356
62

**Laboratoires de recherche**
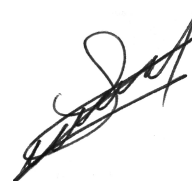Laboratoire de Psychologie Cognitive
Laboratoire de Neurosciences Cognitives

**Composition du jury**

| | |
|---|---|
| Scott BROWN | Rapporteur |
| Professeur des Universités | University of Newcastle |
| | |
| Valentin WYART | Rapporteur |
| Chargé de Recherche (HDR) | INSERM |
| | |
| Laurent MADELAIN | Président |
| Professeur des Universités | Université de Lille |
| | |
| Anna MONTAGNINI | Examinatrice |
| Chargée de Recherche (HDR) | CNRS |
| | |
| F.-Xavier ALARIO | co-Directeur de thèse |
| Directeur de Recherche | CNRS |
| | |
| Boris BURLE | co-Directeur de thèse |
| Directeur de Recherche | CNRS |

Je soussigné, Gabriel Weindel, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de Boris Burle et F.-Xavier Alario, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisées dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Marseille le 04 février 2021

*À Léa et à Margot et ses futurs cousin.e.s*

# Remerciements

First, I would like to thank Scott Brown, Laurent Madelain, Anna Montagnini and Valentin Wyart for accepting to read and to examine this thesis. Second, science is a collaborative enterprise and the work carried out in this manuscript is largely due to the people I met during these years. But given that most of these people are French the following acknowledgments are in French.

Je tiens tout particulièrement à remercier mes directeurs, Boris Burle et Xavier Alario.

Boris, je me souviens de cette première entrevue avec toi, dans ton bureau "à la Piaget" alors que je n'étais encore qu'un étudiant en licence. Tu as su m'inspirer l'envie d'une psychologie cognitive rigoureuse dans laquelle la neurophysiologie est une clé de compréhension de l'organisation du système cognitif. Presque 7 ans après, je suis allé bien plus loin que j'aurais imaginé dans cette envie. Je pense que c'est à toi en grande partie que je dois ce parcours, par ton esprit scientifique et par l'autonomie que tu as su me faire prendre.

Xavier, je te suis très reconnaissant pour les discussions que nous avons pu avoir, les connaissances et savoir-faire que tu m'as transmis. Ce sont des outils indispensables pour un scientifique mais pas seulement. Je te remercie aussi pour la patience et la gentillesse dont tu as fait preuve quand tu te trouvais face à un doctorant sur des montagnes russes bredouillant des explications pas toujours claires. Tu as su me transmettre un peu de ta sérénité, je crois, enfin ma sérénité est difficile à juger alors que je finis de rédiger cette thèse en pleine pandémie de COVID-19.

Je souhaite remercier les enseignants-chercheur.se.s qui tout au long de ma formation ont directement ou indirectement influencé mon parcours en particulier, Abdessadek El-Ahmadi, Jean-Marc Aimonetti, Eddy Cavalli, Pascale Colé, Fabrice Guillaume, Vincent Hok, Patrick Lemaire, Yousri Marzouki, Thierry Ripoll, Bruno Truchet, et tous les autres sauf Franck Chaillan :-). Tout particulièrement je veux remercier Mireille Bastien pour son engagement auprès des étudiant.e.s de sa formation dont j'ai eu la chance de faire partie.

Je tiens aussi à remercier les enseignant-chercheur.se.s et chercheur.se.s avec qui j'ai pu discuter de mes projets de thèse ou autre, en particulier, Laurence Casini, Chris Chambers, Frédéric Chavane, Jenny Coull, Karen Davranche, Steve Fleming, Thierry Hasbroucq, Andrew Heathcote, Anna Montagnini, Michael Nunez et Franck Vidal. Je remercie tout particulièrement Thibault Gajdos qui a su être un mentor, ton aide a été précieuse et indispensable à cette thèse autant sur le plan scientifique que sur le plan humain.

Je remercie aussi les personnels du LNC et du LPC qui rendent tout ça possible

Ce qui est simple est toujours faux. Ce qui ne l'est pas est inutilisable.
— Valéry, 1942


L'esprit n'use de sa faculté créatrice que quand l'expérience lui en impose la nécessité
— Poincaré, 1908

The following Ph.D. thesis has been made possible thanks to an inter-disciplinary grant by the Collège Doctorale of Aix-Marseille University. The work in this thesis was therefore conducted simultaneously in the laboratoire de psychologie cognitive and laboratoire de neurosciences cognitives in Marseille. The thesis is composed of one introductory chapter, four chapters combining experiments and cognitive mathematical modeling and a final discussion chapter. An extended abstract in French can be found at the end of the manuscript. All code for the analysis performed in the thesis can be fount at https://github.com/GWeindel/PhD_thesis.

# Contents

# Abstract

A primary goal in cognitive psychology is to describe the latent information processing units that operate between the onset of a stimulus and a measured behavior. Mathematical models of cognition aim at decomposing behavior into such processing units by formalizing an assumed generative model. Unfortunately, a generative model may explain the behavioral data while not necessarily reflecting the underlying processes. Obtaining measurements between the stimulus and the responses could provide additional information that fruitfully constrains the processing assumptions.

The present thesis explores this issue by focusing on models of perceptual decision making, a field with a long tradition of cognitive modeling. These models are constructed to account for decision choices and their durations (reaction time in the range of a second) on the basis of a decomposition into encoding, decision and response execution stages. We used electrophysiological measures (electromyography and electroencephalography) to decompose each reaction time into different intervals, presumed to contain these stages. Simultaneously, we manipulated time-honored experimental factors to compare the cognitive locus of experimental effects inferred from both electrophysiological recordings and from model fitting procedures.

Throughout four empirical chapters, we show that the inferences drawn from cognitive models conflict with the electrophysiological decomposition when: 1) the model's core assumption of independence between decision and non-decision processes is proven to be false; 2) standard modeling strategies are inadequate to capture the locus of an experimental effect revealed by the electrophysiological decomposition; 3) opposite experimental effects are revealed in decision *vs.* encoding and response execution processes.

This thorough assessment of a generative model of decision making delineates its validity, merits and limitations to account for the latent cognitive processes. New insights are thus provided on the information processes that allow humans to decide between alternatives.

Keywords: cognitive modeling, cognitive accuracy, decision making, electrophysiological chronometry, mathematical psychology.

# Résumé

L'un des principaux objectifs de la psychologie cognitive est de décrire les processus latents assurant le traitement de l'information entre le début d'un stimulus et un comportement mesuré. Les modèles mathématiques de la cognition rendent compte des mesures comportementales en les décomposant en unités de traitement sous la forme de modèles génératifs formalisés. Malheureusement les hypothèses de ces modèles mathématiques peuvent expliquer les données comportementales sans nécessairement refléter les processus sous-jacent. L'obtention de mesures complémentaires situées entre le stimulus et les réponses pourrait fournir des informations qui permettraient de contraindre de manière fructueuse les hypothèses de traitement de ces modèles.

La présente thèse explore cette question en se concentrant sur les modèles de la prise de décision perceptive, un domaine ayant une longue tradition de modélisation cognitive mathématique. Ces modèles sont construits pour rendre compte des choix et des durées des décisions lorsque le temps de réaction est de l'ordre de la seconde. Ils se basent sur une décomposition du comportement en étapes d'encodage du stimulus, de décision proprement dite et d'exécution de la réponse. Nous avons utilisé des mesures électrophysiologiques, l'électromyographie et l'électroencéphalographie, pour décomposer chaque temps de réaction en différents intervalles, censés contenir ces étapes. Simultanément, nous avons manipulé des facteurs expérimentaux canoniques pouvant affecter les différentes étapes de traitement, et comparé les inférences dérivées des procédures d'ajustement des modèles à celles dérivées des enregistrements électrophysiologiques.

Dans quatre chapitres empiriques, nous montrons que les inférences dérivées des modèles cognitifs entrent en conflit avec la décomposition électrophysiologique lorsque : 1) une hypothèse centrale du modèle, à savoir l'indépendance entre les processus de décision et de non-décision, est mise en défaut ; 2) l'effet d'un facteur expérimental est associé à des étapes de traitement différentes par la modélisation et les mesures electrophysiologiques ; 3) des effets expérimentaux opposés sont observés dans les processus d'encodage et d'exécution des réponses.

Cette évaluation approfondie d'un modèle mathématique des processus cognitifs latents à la prise de décision permet d'en cerner la validité, les mérites et les limites. De nouvelles perspectives sont ainsi fournies sur les processus d'information qui permettent aux humains de décider face à une alternative.

Mots-clés: modélisation cognitive, prise de décision, interprétabilité, chronométrie électrophysiologique, psychologie mathématique.

# 1. Introduction

## 1.1. Mental Chronometry

> If the processing of information by the mind is highly structured [...] then different paths through that structure will entail different time courses, and those differences will be reflected in the response times.
>
> — J. Jastrow, (1890) as cited by Luce, 1986

Mental chronometry, described Posner in 1978 as "the study of the time course of information processing in the human nervous system" (Posner, 1978, p.7) appeared central to both psychology and neurophysiology as early as the mid XIXth century. Since then, mental chronometry has become one of the most influential fields of study in psychology, as measures of reaction time (RT), the time from the onset of a stimulus to the corresponding overt response, span through a lot of the work done in experimental psychology. The latter postulates that mental functions are underpinned by elementary information processing operations (cognitive processes). The execution of each cognitive process takes a given amount of time. Thus, the RT is the sum of the time required to execute the cognitive processes involved in a task. The study of this RT has, therefore, been a privileged object of investigation in cognitive psychology, which aims to describe the architecture of the cognitive system. Neurophysiology, for its part, offers various tools for measuring brain activity, some of them with excellent temporal precision. This suggests a promising interaction between psychology and neurophysiology to understand cognitive processes, and how the brain implements, and thus constrains, these processes.

### 1.1.1. History - Thoughts take time

In the year 1796, royal astronomer Nevil Maskelyne fired his research assistant David Kinnebrook for reporting 800 milliseconds (ms) later than him the time course of stellar object through the Greenwich telescope (Mollon & Perkins, 1996). From this unlawful dismissal the study of mental chronometry, and for some historians experimental psychology, was born. Had employment tribunal existed, the case of M. Kinnebrook would have been hard to defend as philosophers of that time, influenced by the work of René Descartes, considered thoughts to be instantaneous, or at least immeasurable. Nevertheless, astronomers need for precision in measurement led Friedrich Bessel in 1822 to suggest the "personal equation" as a way to correct for inter-individual variability in observed stellar object transit time (Meyer, Osman,

Irwin, & Yantis, 1988). This work is the first known evidence that thoughts take an amount of time that can be measured and is variable among individuals.

Later, physiologists interested in the speed of nerve conduction velocity also came to the same conclusion. Hermann von Helmholtz in 1850 reported that the propagation of an electrical current in the sciatic nerve of a frog was of 25.43 meters per second (Schmidgen, 2002). However von Helmholtz was interested in measuring the time of a sensation in Human beings but could not apply the same experiment as he did with the nerve of a frog. He therefore measured what we call today a simple reaction time (RT) task. He applied a slight electrical shock at some point on the body and asked the participant to produce a response as soon as they detected the stimulation. Von Helmholtz assumed that the time of the reaction was defined based on three sequential additive stages, the conduction of the stimulation among the nerve fibers to the brain, the "processes of perceiving and willing" (von Helmholtz as cited by Schmidgen, 2002) and the conduction of the formed motor command to the muscles. Assuming that these two last stages are constant wherever the site of stimulation was, he calculated the speed of sensation conduction by substraction, for example between the time from a stimulation of the sacrum and the one from a stimulation of the big toe. However, von Helmholtz noted that "if at the time of perceiving the signal the thoughts are occupied with something else, [...], it [the reaction] takes much more time"(von Helmholtz as cited by Schmidgen, 2002). To the notion that thoughts take a time which varies among Humans, von Helmholtz added that this time can be modified by changes in the attention of the participant.

Following von Helmholtz's work, a lot of psychologists and physiologists measured the reaction time under various conditions. Wilhelm Wundt, 1874 summarized some of these findings by stating that the speed of the RT is inversely proportional to the intensity of the stimulation formulating by this, although vaguely, the first law on RT. Henri Piéron (Piéron, 1913, 1914) took the investigation further. By systematically reporting and analyzing the work by previous researchers and his measures of RT with different stimulation intensities, Piéron suggested that the mean RT is a power function of the stimulus intensity $I$ :

$$RT = \alpha I^{-\beta} + T_R \tag{1.1}$$

Where $\alpha$ is a multiplicative constant and $\beta$ determines the rate of decrease of the function and where $T_R$ is the residual asymptotic RT. With this work, Henri Pièron suggests that one can formalize the relationship between a physical property of a stimulus and the RT, a law still highly influential in today's research (e.g. Palmer, Huk, & Shadlen, 2005a; Servant, Montagnini, & Burle, 2014; van Maanen, Grasman, Forstmann, & Wagenmakers, 2012).

All these works relied on the observation of the time elapsed between a stimulation and the response of a participant. But, as made explicit by the approach of von Helmholtz, psychologists and physiologists assumed that RT was a composite measure from multiple stages. This same assumption brought Franciscus Cornelius Donders,

1868 to formalize the subtractive method. His idea was that if one can find two tasks that differ only in the addition of a stage between the stimulation and the response of a participant, then one can measure the duration and properties of that additional stage by comparing both tasks. To that end he added two other tasks to the simple RT task like used by von Helmholtz. First he designed a decision task in which participants were presented random series of two auditive stimuli and were required to produce the specific response assigned to each stimulus. Designing this task he assumed that subtracting the simple RT to this choice RT gives the duration of the decision stage (which stimulus) and of the response selection stage (which response). In the same idea he designed a go/no-go task where participants had to respond to only one of the two sounds. He assumed that comparing this go/no-go RT to the choice RT would allow one to measure the duration of the response selection stage as response selection is supposedly absent from the go/no-go task. However, the subtractive method requires several strong assumption (Coles, 1988) such as the seriality between stages (one stage starts after completion of the previous one) and that an added process such as the response selection does not change the other processes. Among these assumptions, especially the last pure insertion assumption was criticized on the basis of introspective reports by Külpe (Vidal, Burle, Grapperon, & Hasbroucq, 2011, as cited by), therefore casting doubt on the reliable decomposition of RTs through the subtractive method (but see Vidal, Burle, Grapperon, & Hasbroucq, 2011, for electro-physiological evidence supporting the pure insertion assumption). Decomposing RT into different stages quickly seemed to be out of reach, leading to disappointments in the early XXth century as captured by Robert Sessions Woodworth's quote "[Since] we cannot break up the reaction into successive acts and obtain the time of each act, of what use is the reaction time?" (as cited by Schall, 2019). Since then, researches trying to account for how humans can decide between different alternatives using tools developed in statistics, physics, economics and neuroscience have led to new ways to achieve the decomposition of RTs.

## 1.1.2. Studying decision making through choice RT

Our daily lives are rythmed by decisions, from the most common like choosing a face mask against COVID-19 to the most engaging such as a career choice. From these choices we can measure two dimensions, the reaction time and the choice made. These two indices can serve as a way to trace back what operations were made by the cognitive system between stimulation and response. However both of these indices are stochastic by nature, if presented the same stimulus at multiple times a participant will display different RT and even choices. This implies that in order to compare participants or experimental conditions one needs multiple RT measurements and summaries of these multiple RT (e.g. mean and standard deviation). For example Donders, 1868, in order to apply his subtractive method, asked participants to perform each task multiple times so that the difference between the mean of the RTs of each task reflects the difference in processing time caused by the task difference rather

than the variability of the measurement. By conduction several trials of a same task, therefore measuring several decisions from a participant we obtain both a distribution of the multiple measured RT and a distribution of the choices made.

### 1.1.2.1. Shape of RT data

In the context of a two alternative forced choice task (2AFC), the amount of information one can obtain about the mechanism at play during a decision using only the binary choice (e.g. yes/no, word/non-word) made by the participant can be extremely low. RTs being a continuous variable provides us with more fine grained information on the underlying process. However RTs are also often more variable within a participant than mean RT between conditions. Additionally their distribution generally follows a shape that does not seem to obey any mathematical functions defining usual stochastic processes (Noorani & Carpenter, 2013). The distributions of RT usually found for a participant in a given condition is almost always skewed toward higher values of RT (see Figure 1.1a). Other distributional properties have been noticed about the RT. For example, almost a century ago, Fessard, Fessard, Kowarski, Laugier, and Monnin, 1936 reported that the spread of the distribution (e.g. standard deviation) increases with the mean of the RT (see Figure 1.1b and Luce, 1986, p.64 for a demonstration).



(a) Histogram of RT                    (b) Densities of RT
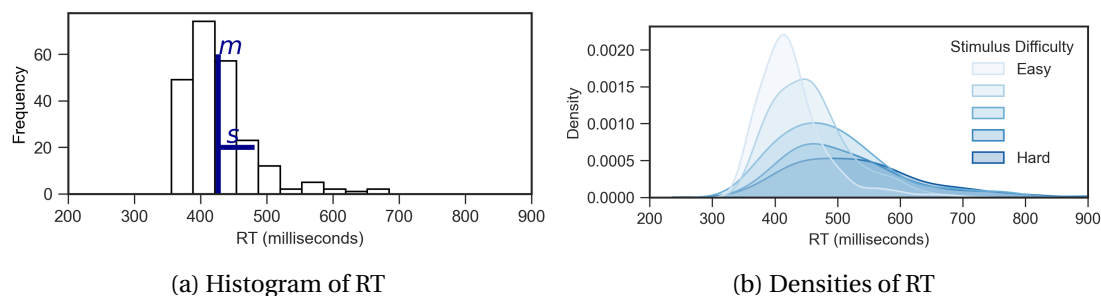
Figure 1.1.: a) Histogram of RTs for one participant (S1) from the experiment described in chapter 4 along with the corresponding mean (m) and standard deviation (s) in blue. b) Densities of RT of the same participant across five stimulus difficulty levels generated using a Gaussian kernel provided by the *seaborn* python package.

These characteristics have quickly appeared to mathematical psychologist as being relevant to the processes that shape the RT distribution. As summed up by Van Zandt, 2000 "The shape of the density often provides clues to the kind of random variable and, therefore, potentially to the candidate processes underlying the execution of a particular RT task". For example the property of the spread growing with the mean of the RT distributions (Wagenmakers-Brown law E.-J. Wagenmakers & Brown, 2007) illustrated in Figure 1.1b is so stable that some consider it as a "law" of RT . It has been shown to be caused by the manipulation of the decision difficulty and not by processes

outside of the decision (Sigman & Dehaene, 2005; E. J. Wagenmakers, Grasman, & Molenaar, 2005; E.-J. Wagenmakers & Brown, 2007).

The very notion that the shape of the density of RT contains information about psychological processes has led efforts to try to account for the properties of RT the distribution. For example Christie and Luce, 1956 suggested that the distribution of RTs, such as the one depicted in Figure 1.1 is the convolution of two distributions, one normally distributed described with a mean $\mu$ and a standard deviation $\sigma$, and one exponential described by $\tau$. The convoluted distribution has then the following probability density function (Matzke & Wagenmakers, 2009) :

$$f(x|\mu,\sigma,\tau) = \frac{1}{\tau\sqrt{2\pi}} \exp\left\{\frac{\sigma^2}{2\tau^2} - \frac{x-\mu}{\tau}\right\} \cdot \int_{-\infty}^{[(x-\mu)/\sigma]-(\sigma/\tau)} \exp\left\{-\frac{y^2}{2}dy\right\} \qquad (1.2)$$

Christie and Luce, 1956 assumed that the two distributions were generated from different processes, the normal distribution as the distribution linked to processes such as the execution of the response and the exponential distribution as the decision stage. Since then a lot of researchers have, using Equation 1.2, fitted the observed RTs to the ex-Gaussian to infer the effect of experimental factors on the assumed cognitive processes through the analysis of the ex-Gaussian parameters (for an overview see Matzke & Wagenmakers, 2009). But although the ex-Gaussian usually fits RT data extremely well (Luce, 1986) the proposed mapping of parameters of the ex-Gaussian with cognitive processes is far from being established (Matzke & Wagenmakers, 2009). From the start some researchers attributed the exponential component either to the decision time Christie and Luce, 1956 or to the "movement response" (Hohle, 1965, McGill, 1963 as cited by). This ambiguity emanates from the lack of a theoretical model therefore preventing from using the parameters as proxy for psychological processes (Heathcote, Popiel, & Mewhort, 1991). To the contrary, Matzke and Wagenmakers, 2009 failed to observe a consistency between the ex-Gaussian and the RT distributions generated with a plausible theoretical model. In the mean time other models such as the shifted Wald (Anders, Alario, Van Maanen, et al., 2016) or the LATER model by Carpenter, 1981 (Linear Approach to Threshold with Ergodic Rate, for a review see Noorani & Carpenter, 2013) have been proposed to describe RTs density shape based on a theoretical model. But these models cannot account for the choices made by the participant, in the context of choice RTs as used in the present thesis one needs a joint description of the distribution of RTs and these choices. First because information about psychological processes is contained in the choice made and second because participants can trade the speed of their response for their accuracy.

## 1.1.3. Speed accuracy trade-off

In most choice tasks used in psychological experiments participants are asked to respond as quickly and accurately as possible. However this simple instruction contains an unresolvable dilemma. As already observed more than 100 years ago by Henmon,

1911, the accuracy of a choice is function of its speed. Hence favoring the speed of a response comes at cost of an increased likelihood to make an error (Figure 1.2).
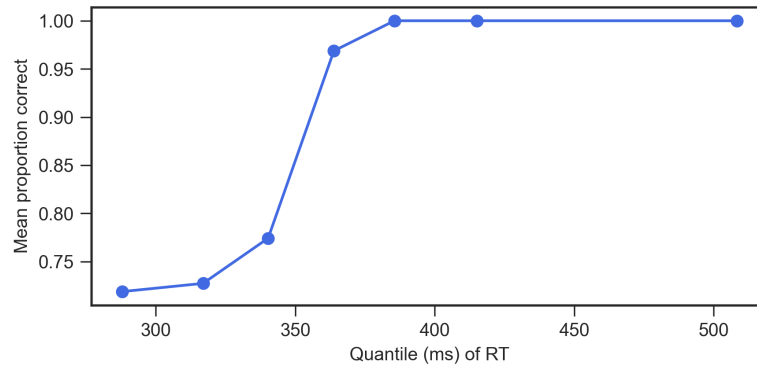


Figure 1.2.: Conditional accuracy function for one participant (S1) from the experiment described in chapter 4 for a mid-level stimulus difficulty and a condition emphasizing speed over accuracy. The distribution of RTs is binned in 7 equal-sized bins and for each bin the mean RT values on the x-axis and the mean proportion correct responses on the y-axis.

This speed accuracy trade-off (SAT) hinders comparison of participants or tasks on the RT alone. If for example, one group of participants is faster than an other but also makes more errors, it is hard to conclude which group performed the task better as the difference could simply be that one group put more emphasis on speed that the other. A part from the purpose of comparison, as stated by Heitz, 2014, SAT and more generally the link between RT and accuracy, is of central importance if one wants to describe cognitive processes operating during decision making. Therefore when studying cognitive processes present in choice tasks, we seek to explain both the shape of the distributions of RTs, the choices and how participants trade speed for accuracy.

## 1.2. Cognitive modeling of RT

> The essential purpose of cognitive modeling is to allow investigation of the implications of ideas, beyond the limits of human thinking. Models allow the exploration of the implications of ideas that cannot be fully explored by thought alone.
> — McClelland, 2009

Psychologists and neurophysiologists are interested in the way humans and other species process information through the analysis of RT. But observing slower RTs when participants are required to favor accuracy rather than speed does not necessarily show that participants were more cautious as the difference in RT could for example

be explained by a slower response execution. Stated differently, an observed performance does not necessarily translate in a difference in the, often implicitly, assumed psychological processes (Bastien & Bastien-Toniazzo, 2003).

Cognitive modeling is a way to translate a psychological theory of how a behavior is generated to a formal, mathematical, model (See Guest & Martin, 2020, for a recent review on the place of modeling in psychological science). A researcher can then directly test a research hypothesis on the latent processes hypothesized by the formal model rather than relying only on the observed dependent variable(s). The strength of cognitive modeling is, as one has to provide a working algorithm of the psychological theory, that all assumptions about how the mind works in the considered task have to be made explicit. This can lead to some surprise when it comes to formally specify a verbal theory. For example departing from the attempt of the theory of the working memory (Baddeley & Hitch, 1974) to explain why list of long words are harder to remember than a list of short words, Lewandowsky and Farrell, 2010 have derived a set of 144 formal models compatible with the original verbal theory. This gained transparency allows to put into test assumptions that would otherwise be overlooked and select among these assumptions which one is the most compatible with the data at hand. Therefore cognitive modeling also allows an easier falsification of a theory, a critical asset when most of modern science arguably rests on Popper's falsification principle. In echo to the quote of McClelland, 2009, simulating from a cognitive model allows to explore beyond the mere thoughts what the predictions from our theory implies on the data. In that idea, Palminteri, Wyart, and Koechlin, 2017 suggest that simulating a cognitive model in order to falsify a theory or claims about a cognitive function should be a standard practice when relying on cognitive models.

The field of mental chronometry is probably the one in psychology with the longest tradition of cognitive models and therefore builds on decade of research on the working algorithm allowing one to decide between alternatives in a perceptual decision making task. In the next sections we provide an overview of the dominant theories in decision making and some of the most used decision making models.

## 1.2.1. Deciding through accumulation of evidence

The basic principle of deciding through the accumulation of evidence is that samples from a signal (e.g. stimulus) is sequentially accumulated and stored in a decision variable until enough evidence has been accumulated to commit to a choice. The theoretical framework of sequential analysis can be traced back to the work of Alan Turing and his colleagues when they broke the falsely assumed unbreakable enigma code used by Nazi Germany during the $2^{nd}$ world war (see Gold & Shadlen, 2002, for a review of their work and how it can be translated to the neuroscience of decision making). Their work led to a statistical procedure allowing one to quantify evidence in favor or disfavor of an hypothesis based on the sequential analysis of incoming information and to stop the accumulation of evidence when an a priori fixed criterion is met. Apart from having been a decisive war effort, the work of Turing and colleagues,

along with the independent work of statistician Abraham Wald on the sequential probability ratio test (Wald, 1947), inspired mathematical psychologists on the nature of the mechanisms that could drive decision making in humans.

The first cognitive models of RT, such as the one described by Stone, 1960, assumed that a decision is performed based on a sequential analysis of evidence. In the context of a 2AFC between e.g. a response A (e.g. "it's a word") or a response B (e.g. "it's not word"), the model of Stone, 1960 assumes that participants sample evidence towards A or B from a continuous sensory signal either generated by stimulus associated with response A or response B (e.g. presentation of a word or a non-word such as "CRTAF"). Values of a random variable $X$ generated either by A or by B are sequentially recorded at every time sample $t$. Given that $p_A(x_t)$ and $p_B(x_t)$ represent respectively the probability that the sample $x$ of time $t$ is generated when the signal is A and B, the model of Stone, 1960 assumes that each sample $x_t$ is converted to a quantity $c(x_t)$, the log likelihood ratio, representing the evidence to be accumulated :

$$c(x_t) = \log p_A(x_t) - \log p_B(x_t) \qquad (1.3)$$

It is then assumed that participants make a running cumulative sum of $c(x_t)$ and stop deciding when this sum reaches a pre-defined level of evidence (Figure 1.3). When this threshold is reached, the decision time ($T_D$) is equal to the last time sample $t$ and the motor response corresponding to the threshold reached is executed.

It is then assumed that the RT observed for one participant is the sum of the $T_D$ and a time from the stimulus presentation to the start of the accumulation, encoding time ($T_E$), and a time for response execution response execution time ($T_R$) :

$$RT = T_E + T_D + T_R \qquad (1.4)$$

While early models of choice RT are useful to present the accumulation of evidence they suffer from unrealistic assumptions (e.g. that participants have access to the probability density function when stimulus A or B are presented) and they fail to account for properties of RT data such as the fact that under some conditions the RT of errors (e.g. most contrasted path in Figure 1.3) is faster than the RT for correct responses. Many propositions have since been made to try to account for these properties with more realistic assumptions.

## 1.2.2. Decision making models

Many decision making models have been introduced since Stone, 1960. Most of them assume that a decision is made through the sequential accumulation of evidence and are often referred to as evidence accumulation models (EAM) (for alternative models without accumulation see Stine, Zylberberg, Ditterich, & Shadlen, 2020). Figure 1.4 summarizes some of the most used models in the context of 2AFC tasks along with the difference between them. Importantly, the assumptionow evidence accumulation is achieved change the predictions on the distribution of RTs and choices.
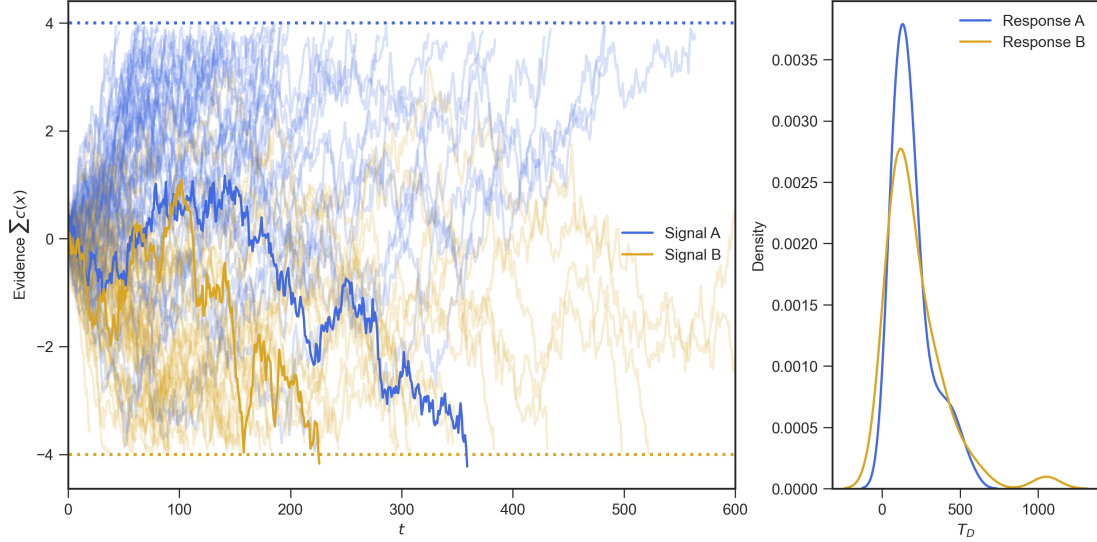
Figure 1.3.: Left plot represents 100 simulated trials of a personal adaptation of the
    sequential model of Stone, 1960. For the sake of simplicity signal A and signal B
    were given random normal distribution with mean 0.1 and -0.1 respectively and a
    common standard deviation of 1 and the threshold was fixed at a value of 4 for
    response A and -4 for response B. At Each trial, signal distribution is equal to A
    or B and at each time unit $t$ (ms, milliseconds) a sample from this distribution
    is added to the cumulative sum of the previous $t$. Accumulation stops once the
    cumulative sum of $c(x)$ reaches the value of one of the threshold. The yellow
    path with the most contrast represents a correct response, the blue path with the
    most contrast an error, signal A was given but because of the overlap between
    information coming from signal A and signal B, the decision variables happens to
    terminate at threshold B.
    Right plot represents the density of the finishing times, $T_D$, corresponding to when
    threshold A and threshold B were hit. Given that the accumulation is unbiased, as
    it starts at equidistance between response thresholds A and B, and that no other
    source of variability is present, we expect, on the long run, that both densities be
    the same. Note that $T_D$ is by construction positive but the kernel used to plot the
    densities forces the distribution to be continuous over 0.

23
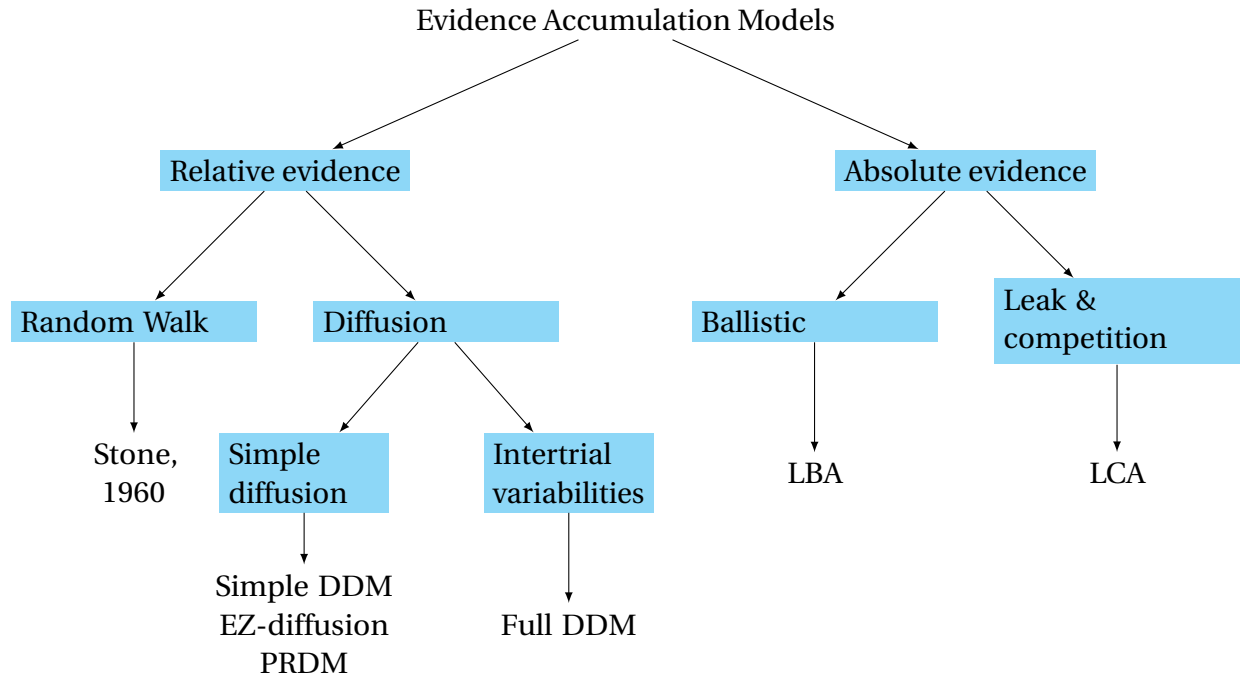
Figure 1.4.: Relations between a selection of evidence accumulation models. The first level splits the type of evidence being accumulated relative and absolute evidence. The Second level determines the dynamic of evidence accumulation.

### 1.2.2.1. Nature of the accumulated evidence

What can be considered as the first distinction among EAMs is the nature of the accumulated evidence. As the model of Stone, 1960, the evidence being accumulated can be relative, meaning that evidence (e.g. $c(x_t)$ in Equation 1.3) for one alternative is evidence against the other alternative. Alternatively, the evidence can be absolute. In that case each alternative has its own accumulator and the decision is then made when one of the accumulator reaches a decision threshold (usually shared across alternatives).

### 1.2.2.2. Dynamic of the accumulation

A second distinction can be made on the dynamic of the accumulation. For models with relative evidence, as seen previously the model of Stone, 1960, evidence is accumulated as a random walk with discrete time steps. The diffusion dynamic of evidence accumulation is the continuous counterpart of the random walk, evidence is integrated continuously. The first diffusion model was suggested by Ratcliff, 1978 as he considered more natural that the brain accumulates information continuously rather than at discrete time intervals (Smith & Ratcliff, 2015). On the side of the models assuming an integration of absolute evidence, the easiest dynamic is the ballistic. In

models such as the Linear Ballistic Accumulators (LBA) (S. D. Brown & Heathcote, 2008) evidence accumulation is deterministic rather than stochastic as no intra-trial source of variability is present in the accumulation process: as soon as the accumulation starts the evidence is integrated at a constant rate for each accumulator, usually with different rates, until one of the accumulator reaches a threshold (see also Heathcote & Love, 2012, for another model using a ballistic accumulator). The variability in the RT is assumed to be generated by two sources of between trial noise : a trial-to-trial variability in the rate of increase and a trial-to-trial variability in the starting point of the accumulation (accumulation can start closer or further away from the unique threshold). The Leaky Competing Accumulators (LCA) Usher and McClelland, 2001 is similar to the LBA except that, as models based on random walk and diffusion processes, it includes within trial noise. This noise is assumed to be generated first by a decay in the accumulated evidence and second by a response competition among accumulators: evidence in favor of one alternative decreases the evidence in favor of the other alternative. While this model is the most complex model from the presented one until now it is also thought of being more biologically plausible. Given that the accumulated evidence is absolute and that each alternative has its own accumulator both the LBA and LCA naturally extend to choices with more that two alternative which is not the case for the relative evidence accumulation models.

### 1.2.2.3. Simple Drift Diffusion Model

Among the EAMs presented, the diffusion model proposed by Ratcliff, 1978, often termed simple Drift Diffusion Model (DDM) is probably the most used model to make inferences on latent processes in 2AFC tasks. As for the model of Stone, 1960 the decision variable $x$ (each single path in Figure 1.5) presents variability during the trial so that it takes, under the assumption that the boundaries are at equidistance from the starting point, a probabilistic value given by :

$$dx = cdW + vdt, \quad x(0) = 0 \tag{1.5}$$

where $dx$ is the evolution of the decision variable over a small time interval $dt$. This evolution is a diffusion because of the first term $cdW$ is a Gaussian distributed white noise with mean 0 and standard deviation $c$ termed the *diffusion coefficient*. But this diffusion can be drifting towards an alternative depending on the value of $v$ termed the *drift rate*, which in the context of a choice task represents the average increase over $dt$ towards the alternative favored by the stimulus. This diffusion process stops once the absolute value of $x$ exceeds the *boundaries* parameter $a$ (see Figure 1.5). Bogacz, Brown, Moehlis, Holmes, and Cohen, 2006 provide two expressions to show how the mean proportion correct ($P_c$) and the mean $T_D$ ($\overline{T_D}$) depends on these 3 parameters :

$$P_c = 1 - \frac{1}{1 + \exp(2va/c^2)}, \quad \text{and} \tag{1.6}$$

$$\overline{T_D} = \frac{a}{v} \tanh\left(\frac{va}{c^2}\right) \tag{1.7}$$

To obtain the mean RT one additional parameter is needed, a constant to shift the RT distribution as in the model of Stone, 1960, the sum of $T_E$ and $T_R$ (see Figure 1.5). This parameter is then added to the $\overline{T_D}$ of Equation 1.7 (as in Equation 1.4) to compute the predicted mean RT.

Equations 1.6 and 1.7 assume that the starting point of the accumulation is located at 0, the mid distance of both boundaries, hence the choices are unbiased prior to the accumulation. However it is likely that participants or experimental conditions bias the choice towards one alternative. This is captured by the *starting point* parameter $z$, if $z$ takes positive values the accumulation is biased toward the alternative represented by the positive threshold ($a+$) and conversely to $a-$ if $z$ takes negative values ($z$ is equal to 0 in Figure 1.5).
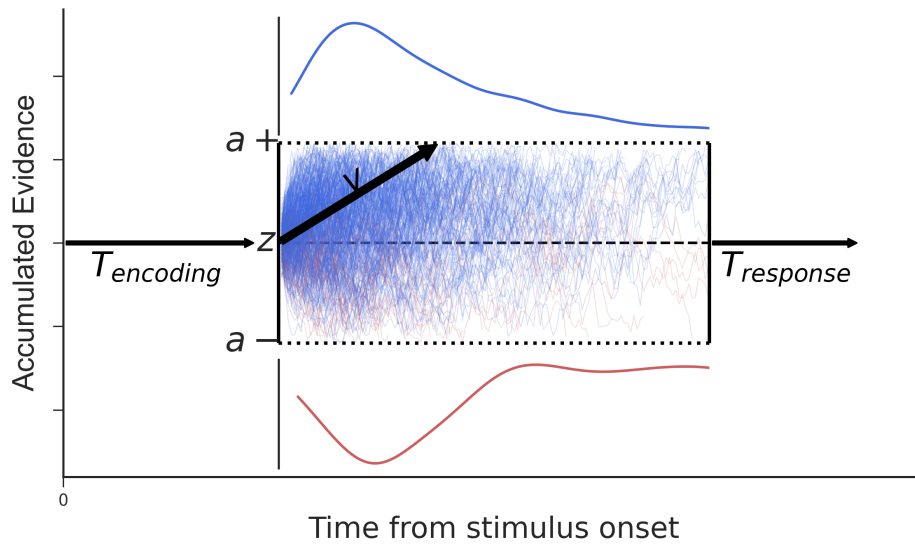


Figure 1.5.: Simulations of 500 trials of a DDM as written by Ratcliff, 1978 assuming the presentation of a single stimulus associated with the positive threshold ($a+$). Blue paths represent correct trials while red paths errors. Over each threshold is presented the distribution of $T_D$ associated with each first passage time of the threshold. In these paths, the *starting point z* is assumed to be 0, the *drift rate v* positive (0.4 units / second), the *thresholds a+* and *a−* 1/6th value of the drift rate and the diffusion coefficient $c$, not displayed in the figure, approximately 1/3rd of the value of the drift rate.

As suggested by the legend of Figure 1.5, where parameter values are given as fraction of the drift rate, the parameters are relative to each other, i.e. one can multiply

each parameter (except the constant added to the $T_D$) with the same value and get the exact same predictions. It is therefore necessary to fix one of these parameter to a given value in order to scale the parameters (i.e. a scaling parameter). Usually in the applications of the DDM the *diffusion coefficient c* is fixed at 1 (or 0.1 in certain applications)[1]. Constraining *c* therefore brings down the number of free parameters in the simple DDM to 4 parameters : the drift rate, the thresholds, the starting point and the additional residual time.

### 1.2.3. Speed of processing

The drift rate ($v$ in Figure 1.5) is conceptualized as the average increase in evidence inside the decision variable. In *relative* accumulation models, such as the DDM, the decision path is stochastic. Noise ($cdW$ in Equation 1.5) will randomly push the decision variable to either threshold at each step. When information about a stimulus is present, the decision variable will tend to drift towards the threshold that is associated with the stimulus. This drift can be null, whereby the decision variable will be driven solely by noise and will cross either thresholds with equal probability. If the drift rate is non-null then the decision variable will be biased toward the choice with the same sign as the drift rate (i.e. favored by the environment). The value of this drift will then influence the time to reach a threshold and the frequency at which each of the two thresholds will be hit.

Based on these properties, the drift rate is associated with the the speed at which information is extracted from a stimulus. When the extraction of the choice related dimension of a stimulus is difficult it is expected that this rate of evidence accumulation will be lower than for an easy stimulus (e.g. a word with a low *vs.* high frequency in the participants language). The drift rate has therefore been associated with the speed of processing of a stimulus and has proven to explain common patterns on RT data. For example, the Proportional Rate Diffusion Model (PRDM) proposed by Palmer, Huk, and Shadlen, 2005a, a low parameter version of the DDM, connects stimulus intensity and the distribution of mean RT (as the Pièron law in Equation 1.1) and response accuracy through the drift rate (see Chapter 3 for an application of the model to real data). Another example comes from the Wagenmaker-Brown law, described in Section 1.1.2.1, that has been shown to hold when the drift rate of a DDM is varied.

### 1.2.4. Decision thresholds

The decision threshold ($a$ in Figure 1.5) is formalized in EAMs as the amount on evidence, relative or absolute, that has to be accumulated before a choice is made. In the context of a DDM, for a given drift, changing the values of the threshold will

---

[1]This choice has however been criticised by Donkin, Brown, and Heathcote, 2009 as being unnecessarily over-constraining and making an untested psychological assumptions about the dynamic of decision making. This convention has nevertheless been adopted by the modelling community and will also be adopted in the present thesis.

change both the speed at which the decision is made and the frequency at which one of the two decision is made. Therefore the threshold parameter is associated with the response caution. In the standard account, when a participant is cautious, s/he will increase the height of their boundaries hence maximizing the likelihood that the decision variable will end on the threshold of the same sign as the drift rate $\nu$ (Ratcliff & McKoon, 2008). Conversely when a participant is more concerned about the speed of a decision than its accuracy, s/he will decrease the decision threshold in order to provide a quick decision. Therefore with the parameter of the thresholds the DDM provide a principled way to explain the SAT pointed out in Section 1.1.3.

Most of the time, the thresholds are assumed to be symmetrical around the starting point ($z$ in Figure 1.5)[2]. As a consequence, if the drift rate is null, the decision variable will land with equal probability at one of the threshold. However it is likely that decision can be biased towards one of the threshold, (e.g. if words are more frequent that non-words, the participant will a priori assume that word responses are likelier). The starting point parameter of the DDM captures this possibility, mathematically changing the starting point between the thresholds (changing the location of $z$ in Figure 1.5) is equivalent to moving one of the threshold closer to the starting point of the accumulation. Hence the starting point is assumed to capture the bias in the decision.

## 1.2.5. Non-decision times

### 1.2.5.1. Formalism

Stone, 1960 formalized a residual time that need to be added to the $T_D$ to account for the RT. He suggested Equation 1.4 where the RT is the sum of a pre-accumulation time $T_E$, he calls input time, a decision time $T_D$, and a motor time $T_R$ ($T_{encoding}$ and $T_{response}$ in Figure 1.5) :

The goal of the model of Stone, 1960 presented in Section 1.2.1 is to account for $T_D$, the time usually relevant to the researcher working on RT (Luce, 1986). The $T_E$ and $T_R$ are then lumped into a residual time we refer to as Encoding and Response execution Time ($T_{er}$).

$$T_{er} = T_E + T_R \tag{1.8}$$

Interestingly, while labels and definitions changes slightly, all models presented in Figure 1.4 share the goal of estimating the $T_D$ along with a residual time.

Based on Equation 1.4, mathematically the residual time is estimated as a constant and therefore acts like a location parameter that shifts the distribution of RTs according to its value (e.g. the value to generate the data in Figure 1.6 was equal to 350ms while it would be of 0 in Figure 1.3). This parameter is then considered as being independent of the decision as it doesn't impact the choice made. This residual time is, as a result, called non-decision time. This mathematical formalism also implies that the

---

[2]For practical numerical purposes in many applications including in the present thesis the lower threshold is often 0 and $z$ is equal to $a/2$ if unbiased

processing components contained in the non-decision time do not contribute to the variance of RT but solely on the mean of the RT. A latter addition to the DDM (Ratcliff & Tuerlinckx, 2002) but also other EAMs such as the LBA (Heathcote & Love, 2012) added inter-trial variability to the $T_{er}$ as a uniform distribution in order to account for the possibility that processes inside the non-decision time might display some variability.

### 1.2.5.2. Interpretation of the non-decision time

Like Stone, 1960, the original formulation of the DDM Ratcliff, 1978 stated that the non-decision time is composed of a time for the encoding of a probe (in a memory task) or stimulus, preparation of the decision and the execution of the response. This interpretation is shared by most modelers and experimentalist to the point that a high number of applications of the DDM and other EAMs directly interpret this constant as reflecting these processes. As a consequence a change in this non-decision time is interpreted as a difference in response execution time, (e.g. Gomez, Ratcliff, & Childers, 2015; Ho, Brown, & Serences, 2009; A. Voss, Rothermund, & Voss, 2004), input time (e.g. Gomez & Perea, 2014; Gomez, Perea, & Ratcliff, 2013; Oganian, Froehlich, Schlickeiser, et al., 2016) or even thought to reflect an added non-decision process between experimental conditions (Ratcliff & McKoon, 2008).

Based on previous estimations, Ratcliff, 1978 suggested that this $T_{er}$ falls into the range of 250ms to 350ms. About the variability of the non-decision processes, Ratcliff and Tuerlinckx, 2002 estimated that its standard deviation is less than 1/4 the standard deviation of the RT. However to date not much data is available to assess the plausibility of these estimates.

### 1.2.5.3. Limitations of the simple DDM

The simple DDM has already been applied in a variety of tasks and has also been declined in alternative models to ease the use by other researchers. For example E.-J. Wagenmakers, van der Maas, and Grasman, 2007 propose the EZ-diffusion model that uses Equations 1.6 and 1.7 to provide a way to estimate (assuming that the starting point is at equidistance from the boundaries), $v$, $a$ and the $T_{er}$, using only the observed mean RT, the variance of RT and the observed proportion correct for one participant in a given condition.

But the simple DDM and its variants suffer from the problem that they impose that the distribution of RT for correct responses is identical to the distribution of error responses (as in Figure 1.5 and as is also the case for the model of Stone, 1960). However this prediction is regularly falsified by the observed data, for example in difficult tasks the mean RT of error responses is often found to be slower that the mean RT for correct responses and this pattern can be reversed when the experiment emphasizes speed over accuracy (e.g. E.-J. Wagenmakers, Ratcliff, Gomez, & McKoon, 2008). To account for these phenomenons refinement of the simple DDM have been suggested.

### 1.2.5.4. Full Drift Diffusion Model

Two main changes have been brought to the simple DDM in order to account for the failure to capture the relative speed of errors *vs.*correct responses (Ratcliff & Rouder, 1998). These changes boil down to the idea that there can be trial to trial variability in the starting point of the accumulation and in the value of the drift rate. It is thought that these inter-trial variabilities can reflect some genuine effects such as sequential effects changing the value of the starting point from trial to trial or fluctuating attention generating different values of drift rate. As a consequence using a full DDM, one assumes that the drift rate and the starting point of each trial are two draws from the respective distribution of drift rates and starting points. These distributions are classically assumed to be normal with a mean and a standard deviation for the drift rate (Ratcliff & Tuerlinckx, 2002) and a uniform distribution for the starting point (Ratcliff, 2013). As a consequence using the full DDM implies to estimate two additional parameters, the standard deviations of drift rates and the bounds of the starting point. Figure 1.6 shows how this addition changes the predictions of a DDM.

(a) Simple DDM (Ratcliff, 1978)  (b) Full DDM (Ratcliff & Rouder, 1998)
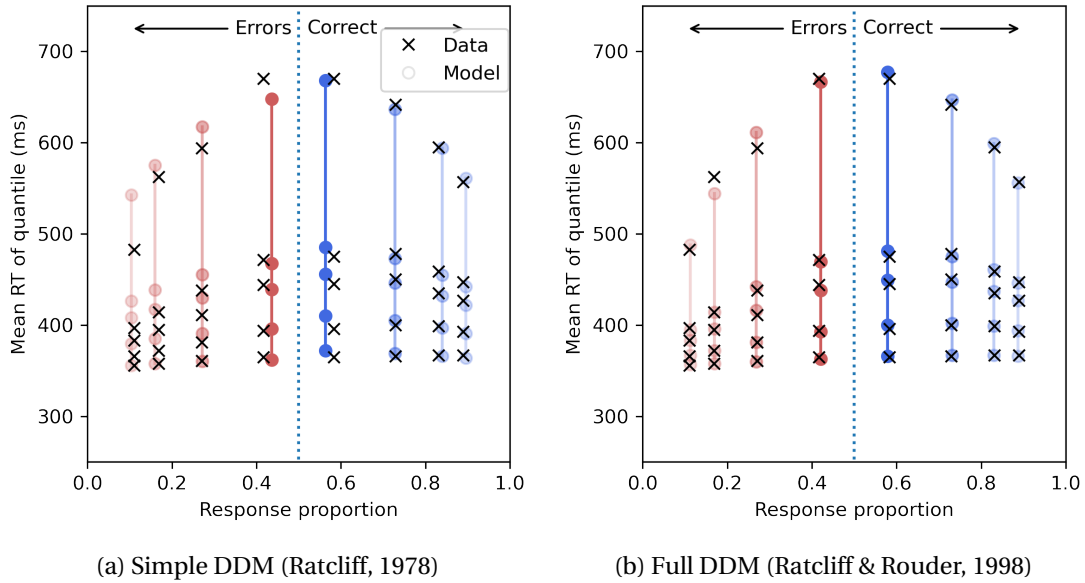
Figure 1.6.: Quantile-probability plots (see Ratcliff & McKoon, 2008, for a complete description of these types of representation) for simple DDM (a) and the full DDM (b). The X-axis displays obtained response proportion across different conditions, symmetrically for errors (red-ish colors) and correct responses (blue-ish colors). The Y-axis displays the *RT* distribution binned in 5 even sized quantiles (.1, .3, .5, .7 and .9 quantiles, from bottom to top). Simulated response proportion and *RT* quantiles (crosses) were computed for one fictive participant whose performance was generated with a full DDM using the *fast-dm* program (A. Voss & Voss, 2007). The recovered or estimated response proportions and *RT* quantiles (dots) were computed from the parameters estimated by a fit of the simple (a) or full DDM (b) to the simulated data.

From the figure a) we see that the simple DDM fails to capture some characteristics of the data such as the faster *RT* for errors and the spread of the last quantile. Figure b) The full DDM describes the quantiles well.

## 1.2.6. State of the art

Since their first introduction, EAMs have become an intensively used tool to study diverse aspect of cognition and inter-individual differences.

### 1.2.6.1. Decomposing psychological effects on cognitive processes

Apart from explaining SAT and stimulus intensity manipulations as a change in respectively the threshold and drift rate parameter, EAMs have proven to be useful in decomposing cognitive effects in terms of cognitive processes. For example us-

ing an EAM, van Maanen, van der Mijn, van Beurden, et al., 2019 have shown that participants placed in a hot-tub will have decreased decision thresholds, coherent with previous studies suggesting that increasing body temperature speeds-up time perception and in return adding on the speed stress of the participants response. EAM are also useful for revealing different strategies used by participants. Anders, Hinault, and Lemaire, 2018 have for example shown, using an EAM, that arithmetic problems resolved with heuristics yielded lower thresholds than items resolved by calculation. This ability to decompose psychological effects into measured and identified cognitive processes has extended the use of EAMs from economical decisions (Towal, Mormann, & Koch, 2013) to sociological and political questions. For example, Todd, Johnson, Lassetter, et al., 2020 have suggested that priming participants with faces of black *vs.*white male in classifying object as a tool or a gun lead to a starting point change. Presenting black faces when race was salient biased participants to the gun response threshold. These successes in terms of explanatory power have also encouraged the use of EAMs to characterize group and individual differences in terms of cognitive processes rather than the mere performance in decision making.

### 1.2.6.2. Measuring group and individual differences at the process level

The slower RT in 2AFC of elderly compared to young participants has for example usually been interpreted as a general slow down linked to age. But an impressive number of studies has in the mean time shown, using decision making models, that this slow down is linked to an increased response caution, a slower non-decision times but not a lower speed of processing (e.g. see Dirk, Kratzsch, Prindle, et al., 2017; Ratcliff, Thapar, & McKoon, 2003; Ratcliff, Thapar, Gomez, & McKoon, 2004; Ratcliff, Thapar, & McKoon, 2001). Another study by White, Ratcliff, Vasey, and McKoon, 2010 showed that participants with a high anxiety trait had a higher drift rate towards threatening stimuli. EAMs have been used to study pathologies such as depression (Pe, Vandekerckhove, & Kuppens, 2013), aphasia (Ratcliff, Perea, Colangelo, & Buchanan, 2004), schizophrenia (Moustafa, Kéri, Somlai, et al., 2015), Parkinson's disease (Herz, Bogacz, & Brown, 2016), language impairments (Anders, Riès, van Maanen, & Alario, 2017), and so on (see Ratcliff, Smith, Brown, & Mckoon, 2016, for a review).

This overview illustrates the different applications for which the EAMs have notably intervened. It reflects the endorsement of models of decision making by the scientific community, in a wide range of scientific disciplines, as a tool for characterizing the latent processing dynamics of decision-making. But all of these applications rest on the interpretative (cognitive) validity of the estimated parameters. This indispensable condition depends on a number of key criteria being met, as follows.

# 1.3. Cognitive accuracy of decision making models

> All models are wrong but some are more wrong than others.
> — Guest and Martin, 2020

Cognitive accuracy can be defined as the ability of a model to accurately reflect the latent cognitive processes. This cognitive accuracy can be expected from a model that is good enough. In this section we therefore start to define what a good model is and whether EAMs are good model regarding these criteria.

## 1.3.1. What is a good model ?

Many suggestion have been made about how to determine a good cognitive model. We outline five criteria from Myung and Pitt, 2016 as they are of particular relevance for the present thesis. A cognitive model is considered as a good model notably if it is *descriptively adequate, faithfull, confirmable, interpretable* and *plausible.*

## 1.3.2. Goodness-of-fit

Descriptive accuracy or goodness-of-fit defines whether a model does describe the data well. In the example of Figure 1.6 we provided two fits on simulated RT and choice data, a fit using the simple DDM and a fit using a full DDM. From this figure it appears clearly that the full DDM fits the data well while the simple DDM missed some important features in the data (see legend of Figure 1.6). The test is obviously easy because the data to which the model was fitted to was simulated from a full DDM. But even on observed data from decision making tasks the observed goodness-of-fit is generally highly satisfactory for current EAMs (e.g. see goodness of fits reported in the following chapters of the thesis). Nevertheless a good fit, while being a necessary criterion, does not guarantee that the model is actually a good representation of the generative model of the behavior. Roberts and Pashler, 2000 argue that a good fit on data does not reveal whether there is any pattern of data that cannot be accounted for by the model. And in fact, because EAMs often come with a high number of parameters, researchers are more concerned about whether the data is over-fitted rather than under-fitted. On that matter, Ratcliff, 2002 showed that the constraints placed on the parameters of a model such as the full DDM were strong enough to prevent it from fitting plausible but fake data. But even once a model has been defined as both descriptively adequate and inflexible enough that it can account to a limited range of data, some additional criteria are needed to validate a cognitive model.

## 1.3.3. Confirmability

Confirmability is invoked when multiple models are under consideration for a same task. It refers to the existence of a specific pattern of data that can be accounted only by the tested model and not the alternative models. Among EAMs they exist a certain model mimicry. Donkin, Brown, Heathcote, and Wagenmakers, 2011 have for example shown that the LBA and the DDM recovered each others core parameters (i.e. non-decision time and drift rate), except for the threshold parameter. The authors argue that inferences on real data are equivalent for both models. A confirmability

evidence has however been provided when comparing the DDM to the LCA. Ratcliff, Voskuilen, and Teodorescu, 2018 have shown that a task of comparative judgment generated data patterns that only the DDM could fit as the LCA performed worse. However not enough confirmability prone effects have been found to decipher among decision making models.

## 1.3.4. Faithfullness

The faithfullness of a model is the observation that the successes of a model comes from the theory it implements rather than the ancillary assumptions used to implement it. According to Perfors, 2020 not much, if any, models of cognition are good enough, one of the main reason being that "the precision of our computational theories often comes from ancillary elements, not core ones". Jones and Dzhafarov, 2014 pointed out a related problems in EAMs such as the DDM and the LBA. Based on their analysis the authors argue that the inter-trial variabilities parameters of the DDM and the LBA renders the models unfalsifiable as they can fit any pattern of data when the distributions are unconstrained. Their claim has been rejected mainly based on the fact that the authors are using a model that does not correspond to actual models as used in practice (Heathcote, Wagenmakers, & Brown, 2014; Smith, Ratcliff, & McKoon, 2014). However they do raise the fact that some ancillary assumptions are dictating the successes of EAMs such as for example the choice of the distribution for the inter-trial variability parameters. Verdonck and Tuerlinckx, 2016 also showed that using other distributions for the non-decision times did change the inferences made with a DDM from real data set. However using simulations, Ratcliff, 2013 showed that even moderately different inter-trial variability distributions can be recovered by the canonical setting.

## 1.3.5. Plausibility

The plausibility of a model refers to the adequacy of the models assumptions with what has been established by previous studies about the cognitive processes at play. EAMs share a lot of assumptions. For example all models presented in Figure 1.4 assume that there exist a bounded accumulation process (see Stine, Zylberberg, Ditterich, & Shadlen, 2020, for an example of alternative models) or that the non-decision time is additive to the decision time. The former assumption has received a surprising source of plausibility in neural data of awake monkeys performing a decision making task. Shadlen and Newsome, 2001 have for example observed that the pattern of firing rates of some neurons in the lateral intra-parietal area behaved as a bounded accumulator (see Gold & Shadlen, 2007, for a review). Purcell, Heitz, Cohen, et al., 2010 have shown that bounded-accumulator models are able to predict the RT directly using neurons activity better than models that didn't postulate an accumulation mechanism. In humans also physiological measurement have provided evidence for a bounded-accumulation mechanism of decision making as postulated by EAMs. O'Connell,

Dockree, and Kelly, 2012 have shown that an electrical potential recorded using electro-encephalography progressively increases until it reaches a fixed threshold prior to the response.

These lines of results are important, first because they allegedly renewed the interest in decision making models and because they showed that one can determine the plausibility of some assumptions and even select among models using physiological data. While plausibility is obviously a necessary criterion, cognitive models are used because we expect that their estimation are interpretable.

### 1.3.6. Interpretability

The interpretability criterion is defined by Myung and Pitt, 2016 as the link between estimated parameters and the activity of the underlying processes. This criterion is not easily verified because the putative studied cognitive processes are not directly measurable and comparable to the estimated parameters. The major diagnostic to ascertain this criterion is known as the "Selective Influence Test" (Heathcote, Brown, & Wagenmakers, 2015). This approach assesses if an experimental manipulation assumed to selectively-affect a given psychological process, only change the value of the corresponding estimated parameter of the model. In the description of the DDM for example we linked drift rate and stimulus difficulty as usually interpreted. This description predicts that manipulating stimulus easiness, e.g. the strength of the evidence, should selectively affect the drift rate. Other selective influence test can be derived from the description usually made of EAMs parameters. Manipulating response caution, e.g. by changing instructions from emphasizing the speed of a response to emphasizing its accuracy is expected to affect selectively the thresholds as they are supposed to implement SAT. Manipulating the delay to extract relevant information for the accumulation or the delay to produce a response is expected to change the value of the non-decision time if those latency are indeed contained in this parameter. Selective influence tests have given some support to the interpretation of EAMs parameters (see for example Lerche & Voss, 2019; A. Voss, Rothermund, & Voss, 2004). But the above listed manipulations have not always produced selective changes in the expected parameters. For example the canonical manipulation of changing SAT emphasis has translated to other parameters. A. Voss, Rothermund, and Voss, 2004 observed that manipulating SAT could also affect the non-decision time parameter (see also Palmer, Huk, & Shadlen, 2005b; Ratcliff, 2006). Another study by Rae, Heathcote, Donkin, Averell, and Brown, 2014 showed an effect of SAT manipulation on the estimated drift rate. Starns and Ratcliff, 2014 did not find evidence for this effect on drift rate in multiple data sets and have argued that such results are to be expected when a too high speed emphasis is given to the participants. This does raise the problem that defining an experimental manipulation as selective is in itself a strong assumption and by no means mandated.

In a collaborative project involving many experts in EAMs, Dutilh, Annis, Brown, et al., 2019 asked the modelers to recover the cognitive processes behind experimental

manipulations. Modelers were unaware of the type of experimental manipulation occurring between conditions from the de-labeled datasets. Among the manipulations, the authors included changes in SAT instructions, stimulus difficulty and bias in the frequency of a response. About the SAT manipulation most experts mapped the effect to the boundary parameter, but also to the drift rate or the non-decision time. Manipulation of stimulus difficulty as well as the bias also resulted in non-coherent attributions to model parameters. Overall the authors noted general agreement, especially when inferences on non-decision times were ignored, across the EAMs and modeling strategies used. But this study does show that establishing the interpretability of an EAMs parameter is not unequivocally established through selective influence tests. This study also comforts the model mimicry outlined previously. Smith and Lilburn, 2020b have recently studied the task used by Dutilh, Annis, Brown, et al., 2019 and argued that the task was not appropriate to be modeled by the EAMs used. However the fact that a suited task would generate more coherent attribution of manipulations to parameters of a decision making model remains to be shown.

Based on these results the interpretability of a parameter through selective influence tests seems like an elusive goal. The interpretability of a parameter is however the building block of the applications outlined in the previous section. In the next one we provide an alternative way to test for the interpretability of a parameter and more generally the plausibility of EAMs assumptions.

## 1.4. Post-modern mental chronometry

> There's no free lunch. A strict empiricist orientation will not suffice in cognitive psychophysiology. The field needs precise quantitative models for evaluating chronometric and psychophysiological measures. One cannot interpret these measures fruitfully without making some specific theoretical assumptions about the temporal properties and products of mental processes. Such assumptions also have a cost associated with them; they must be tested as best possible along the way.
> — Meyer, Osman, Irwin, and Yantis, 1988

The two first sentences of the quote of Meyer, Osman, Irwin, and Yantis, 1988 from their articles titled *Modern mental chronometry* points to the need of models to understand cognitive psychophysiology. What the authors refer to as cognitive psychophysiology is the association between mental chronometry and measurement of physiological activities related to cognition. Using both sources allows to refine our understanding on the cognitive processes at play but cognitive modelling is necessary to bridge both measures. For example when trying to account how visual perception of a near perceptive threshold stimulus varies over multiple trials, Wyart and Tallon-Baudry, 2009 have shown that an EAM allowed to interpret the fluctuation of a pre-stimulus physiological activity in the visual cortex, recorded using magneto-encephalography, as a perceptual decision bias at stimulus onset. As another example of interaction between physiology and mental chronometry through the use of models,

Jepma, Wagenmakers, Band, and Nieuwenhuis, 2009 have shown that the speed up in RT associated with the presentation of accessory stimuli is linked to a facilitation of the pre-accumulation time $T_E$. They did so by reporting a variation in the estimated non-decision time of an EAM and deciphered the pre-accumulation location by using electro-encephalography (see also Van Den Brink, Murphy, Desender, De Ru, & Nieuwenhuis, 2020, for a similar methodology on temporal expectation).

In decision making the very notion that models can jointly account for behavior and physiology has known a rapid development in the past years. This development has led to the emergence of a whole discipline based on this joint account, *Model Based Cognitive neuroscience* (Forstmann & Wagenmakers, 2015). Using this framework Turner, Rodriguez, Norcia, McClure, and Steyvers, 2016 have for example shown that an EAM could jointly account for behavior, electrical activity from the brain recorded through electro-encephalography and metabolic activity of the brain as measured using functional magnetic resonance imaging. This *tour de force* however does not necessarily establish that the assumptions of a behavioral model are valid nor that its parameters are cognitively accurate in the mapping of the cognitive processes. In the present thesis we choose to use physiology to test for assumptions and interpretability of EAMs.

## 1.4.1. A physiological-based cognitive psychology

As shown by the section on the plausibility of EAMs, physiological data can serve the purpose to test the assumptions of cognitive models. Servant, White, Montagnini, and Burle, 2016 have shown that an EAM applied to a decision task with conflicting dimensions of a stimulus was congruent with the predictions from the electrophysiological data they recorded. While such tests can provide important information about how psychological theories on decision making can and cannot be implemented as cognitive models, they remains largely under-applied. One reason is probably that, in order to do so, one has to provide a physiological index of a cognitive process with both an appropriate linking proposition and a precise measurement.

### 1.4.1.1. Linking propositions

The main difficulty of using a physiological correlate or measure of a candidate cognitive process is to reliably establish a link between a physiological measure and a cognitive process, a linking proposition. Schall, 2019 defines linking propositions as "a formal articulation of the relationship between a neural process and a behavior or cognitive process". But it is unclear how to appropriately link physiological activity metrics to psychological processes (see also Schall, 2004; Teller, 1984) and this precludes from falsifying a model on the basis of the physiological measure. For example, neuronal discharge frequency in certain brain areas increases as a ramping function during the decision, in a manner very similar to the postulated model dynamic. This has been interpreted as a clear indication that the brain might accumulates evidence

as suggested by EAMs. But how does one relate such recorded ramping activity to a parameter such as the drift rate ? E.g., Knowing that such activity has been observed in several brain areas (overview in Brody & Hanks, 2016; de Lafuente & Romo, 2006) what then should be the measure related to the putative single cognitive accumulator or parameter therein ? Therefore, if a physiological measure does not correspond to a parameter, one can hardly decipher what comes from the failure of a theory *vs.* the inappropriateness of the ancillary assumptions induced by the linking proposition. As a consequence, when suggesting a physiological measure of a cognitive process supposed to be captured by a parameter, one has to provide a strong linking proposition in which any discrepancy between estimation and measure of that cognitive process can only be attributed to an invalid specification of the model under consideration.

### 1.4.1.2. Precision

Another limitation to test assumptions of cognitive models and their interpretability from physiological measures comes from the usually low precision of the recordings implied by a low signal-to-noise ratio. This low precision forces the use of data processing method shipping new assumptions. For example brain activity related to the task recorded using electro-encephalography is mixed with artifacts and non-task related activity. Separating this task-related activity from the background noise is often achieved through the averaging of multiple trials so that only sources common to these trials are retained in the average signal. But this signal is overly distorted compared to the distributions afforded by single trial analyses (e.g. Burle, Roger, Vidal, & Hasbroucq, 2008; Callaway, Halliday, Naylor, & Thouvenin, 1984; Dubarry, Llorens, Trébuchon, et al., 2017; Latimer, Yates, Meister, Huk, & Pillow, 2015). This distortion again prevents from deciphering whether a discrepancy between a physiological measure and a parameter results from a problem in the measure or the theory tested.

Based on these two raised limitations, it appears clearly that, in order to use a physiological measure of a process to test the assumptions and interpretability of a cognitive model we need, 1) an unambiguous linking proposition and 2) a precise measure of that activity.

## 1.4.2. Measuring response execution time with electro-myography

For the present thesis we propose a physiological index that is thought to overrule the two outlined limitations. This physiological index is the activity of the effector muscles that perform the responses as measured using electro-myography (EMG). In contrast to electro-encephalography, the high signal-to-noise ratio of EMG allows for a reliable decomposition of the $RT$ of every single trial (Coles, Gratton, Bashore, Eriksen, & Donchin, 1985) into two sub-components: the time from stimulus onset to EMG onset (pre-motor time ($PMT$)), and the time from EMG onset to the behavioral response (motor time ($MT$); Botwinick & Thompson, 1966; Burle, Possamaï, Vidal,
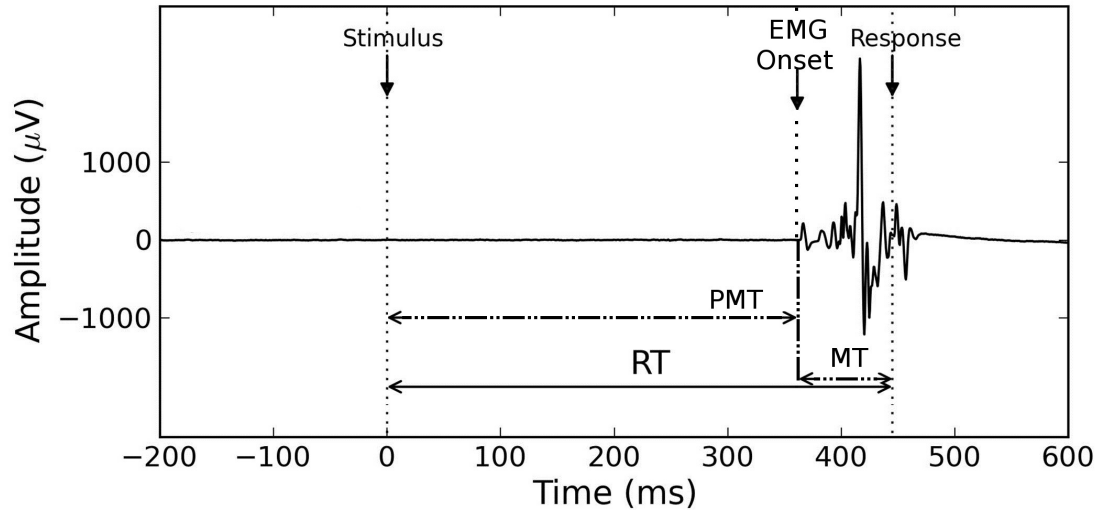
Bonnet, & Hasbroucq, 2002, Figure 1.7).



Figure 1.7.: Single-trial RT decomposition using EMG. The Pre-Motor Time $PMT$ is the time between stimulus and EMG onsets; the Motor Time $MT$ is the time between the EMG onset and the mechanical response. See Figure .1 in Appendix A for additional examples of EMG recordings.

Studies since the 1960s have contributed on building knowledge about this decomposition and shows that the measured $MT$ is undoubtedly linked to a voluntary motor command (Botwinick & Thompson, 1966; Burle, Possamaï, Vidal, Bonnet, & Hasbroucq, 2002; Hasbroucq, Mouret, Seal, & Akamatsu, 1995; Possamaı, Burle, Osman, & Hasbroucq, 2002; Tandonnet, Burle, Vidal, & Hasbroucq, 2003). Additionally these studies shows that the $MT$ is not a mere constant value added to the $PMT$ and its variability contributes to the overall performance one (*i.e* of the $RT$s). Furthermore, $MT$s have been found to be modulated by certain experimental manipulations. Early on, Grayson, 1983 suggested that stimulus intensity affects the estimated $MT$ in a simple reaction task, and Servant, White, Montagnini, and Burle, 2016 reports a recent confirmation in a conflict task. It has also been shown that $MT$s are shortened if a speed emphasis is given (Spieser, Servant, Hasbroucq, & Burle, 2017), advance information is provided about the timing (Tandonnet, Burle, Vidal, & Hasbroucq, 2003; Tandonnet, Burle, Vidal, & Hasbroucq, 2006) or about the nature (Possamaı, Burle, Osman, & Hasbroucq, 2002) of the forthcoming stimulus or response. Given the linking proposition developed below these effects questions the assumption made by modelers using EAMs as they often assume that non-decision processes contained in the $T_{er}$ are invariant to most experimental manipulations. Whether the variations in putative non-decision processes as indexed by MT can be ignored remains to be tested.

### 1.4.3. Linking $T_{er}$ and $MT$

About the linking proposition while the $PMT$ certainly contains many processes, linking the recorded $MT$ with a psychological process is more straightforward. Luce, 1986 (p. 97) states, "*the time from that event* [EMG onset] *to the response is a proportion of the entire motor time and that, in turn, is a proportion of the residual* **R** [in our notation, $T_{er}$]". In other words, the time postulated in the model to execute the response ($T_R$ in Equation 1.8) should be strongly related to the recorded $MT$. This linking proposition is complicated by the fact that all EAMs presented in the thesis estimate $T_{er}$ and not solely $T_R$. However, as proposed by Luce, 1986 (p. 97) :

> The assumption is that if **R'** [here, $MT$] exhibits a dependence upon signal intensity [or other factors], then the chances are that **R** [here, $T_{er}$] will be affected also, which it will be unless the effects on **R'** and **R** - **R'** are equal and opposite - an unlikely possibility.

Hence, using the linking proposition that a variation in $MT$ should be captured by a variation $T_{er}$ comes with the assumption that the non-EMG observed residual time from $T_{er}$ is independent from the variation in $MT$.

Previous application of the linking proposition between $MT$ and $T_R$ have already been made. Smith, 1995 while developing an EAM to account for simple RTs has for example used the EMG recordings of Smith, Anson, and Sant (unpublished, 1992), to determine that $T_R$ in the model take a mean value between 58 and 72 ms and accounts for a relative small part of the overall variability in RT. Servant, White, Montagnini, and Burle, 2015 also assumed a similar linking proposition when testing EAMs developed to account for tasks with conflicting dimensions. Using this linking proposition they showed that when a stimulus is raising a response conflict, decision and non-decision time are not serial. These applications points to the need of applying this linking proposition to EAMs used in decision making as they would allow to test for several common assumptions of EAMs applied in decision making.

Inasmuch as motor time can be mapped to the parameter capturing response time execution $T_{er}$ we believe that the measurement of motor processes *vs.* the estimation made by EAMs is a critical test for cognitive models of decision making.

## 1.5. Organization of the thesis

The developed linking proposition between $MT$ and $T_{er}$ is tested across four experimental chapters, each tailored to test different assumptions on the measured *vs.* estimated decomposition into cognitive processes. Chapter 2 assesses the sensitivity of $MT$ to well-studied effects in decision making such as stimulus strength, SAT, response force output and response correctness. We derive how these effects relate to the $T_R$ as estimated by the arguably most popular of the EAMs presented, the full DDM. We simultaneously evaluate how the by-trial measure of $MT$ covaries with the estimate of $T_{er}$. Chapter 3 tests the link between the starting point of accumulation as

estimated by a DDM and the $MT$. Chapter 4 reports how the overall decomposition of DDM relates to experimental and EMG based predictions. Chapter 5 refines the analysis in the previous chapter by adding a new linking proposition on the putative non-motor residual time of $T_{er}$.

# 2. Assessing model-based inferences in decision making with single-trial response time decomposition

## Sommaire

## 2.1. Note

## 2.2. The Present Study

The present study is based on two experiments, each involving a standard perceptual decision task. Our main hypothesis in this study is that decomposing $RT$s through EMG techniques into $PMT$ and $MT$ will help 1) better establish the locus of certain experimental manipulations, 2) test some central assumptions of the modeling framework regarding the postulated cognitive processes, and 3) assess to what extent EMG-based and model-based decompositions of $RT$s provide (di)similar information about the non-decision time parameter.

Visual contrast was manipulated to adjust task difficulty, and verbal instructions emphasizing speed or accuracy were used to implement an SAT setting. Across the two experiments, we also manipulated the force required for the response to be produced. We recorded the EMG activation and the latency of every manual response, hence deriving single trial distributions not only for $RT$, but also for $PMT$ and $MT$, and assessed the impact of the experimental factors on each of these variables. Based on previous studies (Spieser, Servant, Hasbroucq, and Burle, 2017; Steinemann, O'Connell, and Kelly, 2018, see also Osman, Lou, Muller-Gethmann, et al., 2000; Rinkenauer, Osman, Ulrich, Muller-Gethmann, and Mattes, 2004, who suggested the existence of a SAT effect on motor processes using LRP), we expected that the measured $MT$ would be affected by SAT, and possibly by stimulus strength (Grayson, 1983; Servant, White, Montagnini, & Burle, 2016). In addition, by having a reliable EMG measure for every single trial, we were able to assess the stochastic dependency between $PMT$ and $MT$. In other words, this provides a test for the subsidiary assumption of independence between decision and non-decision processes, a test that is not possible in most regular parameter estimation procedures.

Following this empirical exploration, we estimated the parameters of the Drift Diffusion Model (DDM; Ratcliff, 1978; Ratcliff and McKoon, 2008) from the data using a hierarchical Bayesian fitting method (Wiecki, Sofer, & Frank, 2013, HDDM package). First, we applied a model selection procedure to find the best fitting model. Secondly, we evaluated the correlation between the estimated $T_{er}$ parameter and the measured $MT$, across participants. Finally, we fitted a joint DDM that takes the $MT$ as a by-trial regressor, to assess the link between the $T_{er}$ parameter and the measured $MT$.

## 2.3. Experiment 1

### 2.3.1. Methods

#### 2.3.1.1. Participants

Sixteen participants (8 men and 8 women, mean age = 23.5, 2 left-handed) that were students at Aix-Marseille University, were recruited for this study. They were compensated at a rate of 15 € per hour. All participants reported having normal or corrected vision, and no neurological disorders. The experiment was approved by the ethical experimental committee of Aix-Marseille University, and by the "Comité de Protection des Personnes Sud Méditerrannée 1" (Approval n° 1041). Participants gave their informed written consent, according to the declaration of Helsinki.

#### 2.3.1.2. Apparatus

Participants performed the experiment in a dark and sound-shielded Faraday cage. They were seated in a comfortable chair in front of a 15 inch CRT monitor placed 100 cm away, that had a refresh rate of 75 Hz. Responses were given by pressing

either a left or a right button with the corresponding thumb. Buttons were fixed on the top of two cylinders (3 cm in diameter, 7.5 cm in height) separated by a distance of 20 cm. They were mounted on force sensors allowing to continuously measure the force produced (A/D rate 2048 Hz), and to set the force threshold needed for the response to be recorded. In Experiment 1, this response threshold was set to 6N (600g). Response signals (threshold crossing) were transmitted to the parallel port of the recording computer with high temporal accuracy (< 1ms). At button press, participants heard a 3ms sound feedback at 1000 Hz (resembling a small click). The forearms and hypothenar muscles of the participants rested comfortably on the table in order to minimize tonic muscular activity compromising the detection of voluntary EMG bursts. We measured the EMG activation of the flexor pollicis brevis of both hands with two electrodes placed 2 cm apart on the thenar eminences. This activity was recorded using a BioSemi Active II system (BioSemi Instrumentation, Amsterdam, the Netherlands). The sampling rate was 2048 Hz.

### 2.3.1.3. Stimuli

Stimuli presentation was controlled by the software PsychoPy (Peirce, 2007). Each stimulus was composed of two Gabor patches, presented to the left and the right of a fixation cross. The Gabor patches had a spatial frequency of 1.2 cycles/visual angle degree and had a size of 2.5 visual angle degrees each. The standard Gabor patch contrast was set to 0.5, on a scale between 1 (maximum contrast) and 0 (uniform gray). Five levels of stimulus contrast were used (0.01, 0.025, 0.07, 0.15, 0.30). These contrast values were added to the target Gabor patch and subtracted from the distractor Gabor patch. These contrast levels were decided based on performance from a pilot study where they were found to yield a full range of performance quality: from near perfect to almost chance level accuracy. The task of the participants was to press the button (left or right) ipsilateral to the patch with the highest contrast.

### 2.3.1.4. Procedure

All participants performed one single session with 24 blocks of 100 trials each. Session duration was approximately 1h30, including a training session of 15 minutes and self-paced breaks between each block. During the training session, participants were instructed that "Speed" instructions required a mean *RT* near 400 ms and that "Accuracy" instructions required a minimal response accuracy near 90% while maintaining *RT*s below 800 ms. Participants were also informed to keep their gaze on the central fixation cross during the blocks.

The beginning of each speed or accuracy block was preceded by the corresponding visual instruction (the French word "Vitesse" or "Précision" for Speed and Accuracy, respectively). The end of each block was followed by the presentation of the recorded mean *RT* and response accuracy performance of the block, along with oral feedbacks from the experimenter in those cases were the participant did not satisfy the condition goals. The training session included 40 trials without performance instructions

followed by 2 blocks of 10 trials in the Speed condition, followed by 2 blocks in the Accuracy condition, and ended with 4 blocks of 10 trials with alternating instructions. For the experimental session, speed instructions alternated every three consecutive blocks. The order of the instructions was counterbalanced across participants. The contrast conditions of the stimuli were fully randomized across the 5 levels within each block. No response deadline was applied, the stimulus disappeared when participant produced a button press, the response-stimulus interval was fixed to 1000ms.

### 2.3.1.5. EMG processing

The EMG recordings were read in Python using the MNE module (Gramfort, Luessi, Larson, et al., 2013), and filtered using a Butterworth 3rd order high pass filter at 10Hz from the scipy Python module (Oliphant, 2007). The by-trial EMG signal was then processed in a window between 150 ms before, and 1500 ms after stimulus onset. A variance-based method was used to detect whether EMG activation was significantly present in either hand's channel. The precise onset was then identified with an algorithm based on the "Integrated Profile" of the EMG burst. This method takes the cumulative sum of the rectified EMG signal on each epoch and subtracts it from the straight line joining the first and the last data-points (corresponding to the cumulative sum of an uniform distribution). The onset of the EMG burst corresponds to the minimum of this difference (see Liu and Liu, 2016; Santello and Mcdonagh, 1998 for more details[1]). The EMG onsets determined by the algorithm were examined by the experimenter who could perform manual corrections as needed (18.3% of the trials). For this processing stage, the experimenter was unaware of the trial type he was examining, to avoid any correction bias. Every muscular event (rapid change in the signal followed by a return to the baseline) in the trial was marked, thus quantifying how many times the muscle was triggered, even when the identified activation did not lead to an overt response (see Figure .1 for examples of recorded EMG activity).

Motor time ($MT$) was defined as the time between the onset of the last EMG activation preceding the responding hand button press. Pre-motor time ($PMT$) was defined as the time between stimulus onset and this last EMG onset. In this way, any intervening EMG activations were discarded. For the purpose of this study, we treated the trials with multiple activities (25.90% of the total number of trials, see for example participants S4 and S5 EMG plot in Figure .1) as trials with only the last EMG activation. Notice that if only $RT$ was measured these trials would not have had any special status. It has not escaped our attention that these trials can represent a challenge for evidence accumulation models (Servant, White, Montagnini, & Burle, 2015, 2016), and that they will have to be investigated more thoroughly in future research.

---

[1]A software implementing this two steps procedure will soon be released with an open-source license, and is already accessible upon request.

### 2.3.1.6. Statistical procedure

**Bayesian Statistics**   Apart from a few exceptions (see below), the analyses were performed within the Bayesian framework. Bayesian methods aim to estimate an unknown parameter (or set of parameters) and the uncertainty around it. More explicitly, Bayesian methods implement Bayes' rule to generate a posterior distribution for each parameter based on a combination of prior information and the likelihood of the data given the parameters. This posterior distribution can then be naturally interpreted as the probability of any given parameter value given the data, the priors and the tested model. In our study, we summarize the posterior distribution using the mean, standard deviation and the 95% Bayesian credible interval (CrI). Our criterion to assess the presence of an effect was that the null value lied outside the CrI. While this method does not quantify the evidence in favor of the null hypothesis, it does provide an estimation of the effect size and its uncertainty. All the priors used in the manuscript are detailed in Appendix A.2.

**Bayesian Mixed Models**   To test our hypotheses on the behavioral and EMG variables we used linear mixed models (LMMs). These models estimate fixed effects (*e.g.* the effect of SAT on $RT$) while accounting for random effects ((*e.g.* the inter-individual differences in the effect of SAT on RT), making them particularly useful in repeated measure design such as the one used in this study. Estimating inter-individual differences as random effects shares the information gathered from each participant while providing separate (but not independent) estimates for each one of them. Given our analysis approach, we derived one generic LMM fitted independently for all chronometric dependent variables: $RT$, $PMT$ and $MT$. In these LMMs, the log transformation of the chronometric variables on the $i$th trial for the $j$th participant ($y_{ij}$) was assumed to be drawn from a normal distribution with mean $\mu_j$ and residual standard deviation $\sigma_r$ :

$$y_{ji} \sim \mathcal{N}(\mu_j, \sigma_r) \tag{2.1}$$

Where ~ stands for "distributed as". The mean of each participant $\mu_j$ is then defined by an intercept ($\alpha_j$) and slope coefficient ($\beta_j$) for each experimental factor and their interactions.

$$\mu_j = \alpha_j + \beta_{1j} \, SAT + \beta_{2j} \, Cont. + \beta_{3j} \, Corr. + \\ \beta_{4j} \, RS + \beta_{12j} SAT \times Cont. + \beta_{13j} \, SAT \times Corr. + \\ \beta_{23j} \, Cont. \times Corr. + \beta_{123j} \, SAT \times Cont. \times Corr. \tag{2.2}$$

where $Cont.$ stands for "Contrast", $Corr.$ stands for "Correctness", and $RS$ stands for "Response side"[2]. The individual intercepts ($\alpha_j$) and slopes of each predictor $x$

---

[2]The common R syntax for these LMMs would be : y ~ SAT * Cont. * Corr. + RS + (SAT * Cont. * Corr. + RS | participant)

$(\beta_{xj})$ are modelled as drawn from a normal distribution :

$$\alpha_j \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha) \tag{2.3}$$

$$\beta_{xj} \sim \mathcal{N}(\mu_{\beta_x}, \sigma_{\beta_x}) \tag{2.4}$$

Where $\mu_\alpha$ and $\mu_{\beta_x}$ are the population estimated intercept and slope while $\sigma_\alpha$ and $\sigma_{\beta_x}$ the estimated population variance of the intercept and the slope (*i.e.* the random effect). In order to test our hypothesis, we report for each LMM the posterior distribution of the population-estimated intercept and regression coefficients.

In Equation 4.2, correctness of the response was included as a predictor because the relationship between the distributions of $RT$ on correct and incorrect trials is known to change under speed pressure (Grice & Spiker, 1979), and because $MT$ has been previously-reported to be affected by this factor (e.g. Allain, Carbonnell, Burle, Hasbroucq, & Vidal, 2004; Rochet, Spieser, Casini, Hasbroucq, & Burle, 2014; Śmigasiewicz, Ambrosi, Blaye, & Burle, 2020). Response side was included as an additive predictor because left and right $RT$s often differ. Two remarks are in order concerning this last point. First, we did not expect any interaction with the other predictors. Second, motoneurons synchronization has been shown to depend on handedness (Schmied, Vedel, & Pagni, 1994). As a consequence, we can expect Response side to affect $MT$ and the effect to be, at least substantially, of motor origin.

We also tested the effects of these factors on the proportion of correct responses using a generalized linear mixed model assuming that each response (correct or incorrect) was drawn from a Bernoulli distribution whose parameter depends on the same predictors as the LMM (except the correctness factor and its interactions).

For each LMM and generalized LMM, 6 Markov Chain Monte Carlo (MCMC) sampling processes were run in parallel, each composed of 2000 iterations among which the first 1000 samples were discarded as warm-up samples. We assessed convergence of the MCMC chains both by computing the potential scale reduction factor ($\hat{R}$, see Gelman and Rubin, 1992) and by means of visual inspection of the MCMC chains. We also visually checked the assumptions of the linear regression by inspecting the normality of the residuals through QQ-plots and assessment of homeoscedasticity. The LMM and generalized LMM were fitted with a custom Stan code, available in the online repository, inspired from the code provided by Nicenboim, Vasishth, Engelmann, and Suckow, 2018 and using the `pystan` package (Stan Development Team, 2018). Summary statistics and plots of the parameters were created with the `arviz` python package (version 0.4.1, Kumar, Carroll, Hartikainen, & Martin, 2019).

We also performed a frequentist replication of each G/LMM using the `lme4` R package as a means to check for prior sensitivity. Any discrepancy with the main (Bayesian) analysis is reported in the Results.

**Factor Coding and LMM parameter interpretation**   For all LMMs, sum-contrasts were used for SAT (-0.5 for speed and 0.5 for accuracy) and for response side (-0.5 for right responses and 0.5 for left responses). Treatment-contrast was used for correct-

ness (0 for correct and 1 for incorrect responses). The stimulus strength factor was centered on its middle value and transformed such that -.5 represented the lowest possible contrast and .5 the highest possible contrast. These coding features were chosen to ease the interpretation of the resulting coefficients. When the binary predictor is sum-contrasted (-0.5 and 0.5), the estimated $\beta$ value can be read as the difference between both conditions. When the binary predictor is treatment-contrasted (0 and 1), the estimated $\beta$ can be read as the difference to add to the intercept (predictor at 0) to obtain the mean of the condition where the predictor is at value 1. Hence, in our analysis, the intercept can be read as the predicted time for the reference condition where the response is correct, and at an intermediate value for the predictor SAT. The main effects and interactions can be read according to the coding scheme used, *e.g.* the correctness slope represents the benefit or cost of an error when contrast is at midlevel. The contrast slope represents the benefit or cost of a higher contrast when the response is correct. The interaction between contrast and correctness represents how the correctness/contrast effect changes when contrast is higher than mid-level/an error was made.

$RT$, $PMT$ **and** $MT$ **specific adjustments**    By-trial $RT$, $PMT$ and $MT$ were log transformed prior to the analysis because these variables have heavily skewed distributions that would violate the normality assumption of residuals of the LMM. To ease the interpretation of the estimated LMM parameters, we back-transformed the intercepts by taking their exponential, and each slope by subtracting the exponential of the intercept from the exponential of the sum of the intercept and slope. We applied this back-transformation at each iteration of the MCMC procedure, hence computing the uncertainty around the parameter values on their natural scale.

**Fast guess detection**    Before applying any analysis we performed the Exponentially Weighted Moving Average (EWMA) filter developed by Vandekerckhove and Tuerlinckx, 2007. This method iteratively computes a weighted accuracy measure (amount of correct responses relative to errors) on the sorted $RT$ distribution, from the fastest to the longest $RT$. Participants are considered as being in a fast guess state until the weighted accuracy is higher than a defined threshold. The $RT$ at which this change of state occurs is identified, and all trials faster than this $RT$ are censored. The user defined parameters in this method are the initial starting point of the weighted accuracy, the accuracy threshold for defining non-guess trials, and the amount of preceding trials (weight) retained in the accuracy computation. The starting point was defined at 0.50 based on the assumption that a guessing strategy yields a 50% chance of correct response. The threshold was fixed at 0.60 based on a reasonable assumption that participants were not guessing when accuracy was superior to 0.60. The weight was heuristically fixed at 0.02 (bounded from 0 to 1, with 0 being all preceding trials used), after visual inspection of the rejection plots with different weights. This method was applied for each participant's $RT$ distribution separately in the speed and accuracy conditions, as fast-guesses can have different latencies across

both conditions. The figures illustrating these rejection procedures can be found in the online repository. We thank Michael Nunez for kindly providing the code used for this method (https://github.com/mdnunez/bayesutils/blob/master/wienerutils.py).

## 2.3.2. Results

Two of the sixteen participants were excluded due to high tonic activity in the electromyogram, which otherwise would have made the detection of their EMG onsets too unreliable. This rejection was decided before performing any analyses. Due to technical constraints on the methods for detecting EMG onsets, we applied an upper limit of 1500 ms to the $RT$s, resulting in the loss of 0.50% of the trials. Trials with low signal-to-noise ratio, and trials with high spontaneous tonic activity, making appropriate EMG onset detection difficult, were also removed. In total, these acceptability conditions led to the exclusion of 4.08% trials for all analyses. Additionally, the EWMA method removed 5.80% trials of the data.

The descriptive statistics discussed for the chronometric variables are summarized in Figure 2.1. The parameters of the Bayesian LMM are represented on the same scale (in milliseconds) in Table 2.1 and in Figure 2.2. Additional analyses of response accuracy are provided in Appendix A.3.
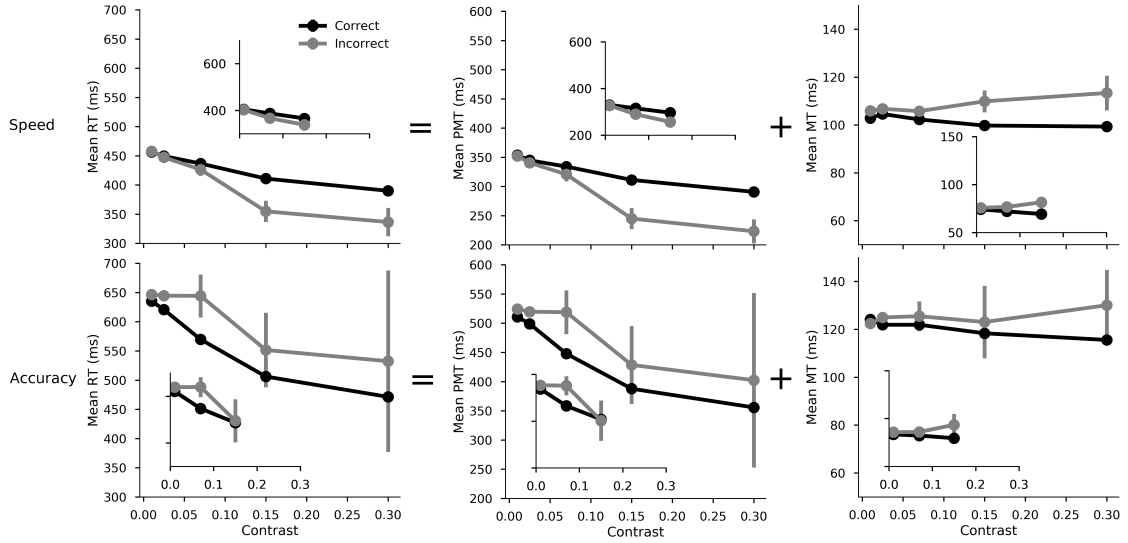
Figure 2.1.: Observed results in Experiment 1, effect of stimulus strength level (contrast) on the mean *RT* (left column), *PMT* (center column) and *MT* (right column). This plot illustrates the interaction between contrast (x-axis), SAT conditions (top vs. bottom rows), and correctness of the response (black - correct vs. grey - incorrect). Bars around the mean represent 95% confidence intervals corrected for within-subject design using the method developed in Cousineau, 2005. To assess replication, the small insets provide the results obtained in Experiment 2 for a subset of the contrast levels (0.01, 0.07, 0.15). Note: these figures analyze the means in millisecond units, while the LMM analysis presented in Table 2.1 model the means in log-transformed units.

### 2.3.2.1. *RT* and *PMT*

*PMT*s and *RT*s become shorter as contrast increases and when speed is stressed. Although the credible intervals contained the null value, we observe a weak positive main effect of correctness. However, an interaction with SAT instructions showed that when speed is emphasized, errors are faster, and when accuracy is emphasized, errors are slower. A strong three-way interaction furthermore specified that this correctness effect according to SAT conditions is even stronger for easier contrasts. Overall, the results for *PMT*s and *RT*s mirrored each other except for one difference: response side (laterality) had a significant effect on the *RT*s but not on the *PMT*s (not shown in Figure 2.1)[3].

---

[3]A frequentist replication of these tests provided the same results, except it included an additional significant interaction : between contrast and correctness for *PMT* ($log(\beta) = -0.07$, $t = 2.97$)

### 2.3.2.2. *MT*

Replicating previous results, *MT*s were also faster under speed emphasis. However, contrary to *RT* and *PMT*, the effect of correctness was in the same direction in both SAT conditions (slower *MT*s during errors). *MT*s were faster when participants responded with their right hand, which explains that the previous laterality effect observed on *RT*s (but not on *PMT*s) was due to *MT* differences. Finally, there was a small but consistent effect of contrast on *MT*. This effect interacted significantly with response correctness indicating that the lengthening of response execution for errors (see Allain, Carbonnell, Burle, Hasbroucq, & Vidal, 2004; Śmigasiewicz, Ambrosi, Blaye, & Burle, 2020, for interpretation) was larger in the easier contrast conditions. No additional interactions were observed (see Table 2.1).
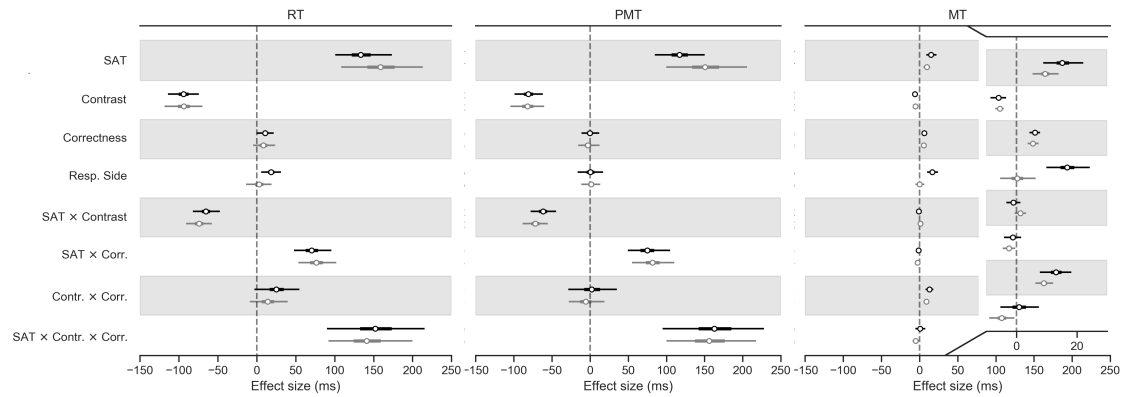


Figure 2.2.: Posterior distributions for the regression coefficients from the LMM model fitted on the *RT*, *PMT*, and *MT* of Experiment 1 (black) and Experiment 2 (grey). The segments around the estimates represent 95% CrIs. As the coefficients for MT are of smaller magnitude, we provide a zoomed inset on the far right of the Figure to ease their visualisation.

### 2.3.2.3. Correlations between *PMT* and *MT*

In most implementations of the evidence accumulation framework, the decision and non-decision stages are assumed to be independent from one another. We tested this assumption by examining, for each participant, the Spearman correlation between by-trial *PMT*s and *MT*s, in the speed and the accuracy conditions separately. It is important to note that certain trial features may bias the correlation estimates. For example, fast-guess trials defined on the basis of their *RT* value could bias the computed correlation coefficients towards negative values. Indeed, as the total *RT* is the sum of *PMT* and *MT*, fast *RT*s most likely result from both fast *MT* and *PMT*. Trimming lower *RT* values would remove trials in which *PMT* and *MT* likely show a positive co-variation, thus biasing the correlation estimates towards negative values. Additionally, as suggested by Stone, 1960, trials with long *PMT* might be subject to a trade-off between *PMT* and *MT* based on an implicit deadline (a process that is not

| Predictor | RT | | | | PMT | | | | MT | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coeff. | SE | 2.5% | 97.5% | Coeff. | SE | 2.5 | 97.5% | Coeff. | SE | 2.5% | 97.5% |
| Intercept | 469 | 12 | 446 | 494 | 354 | 8 | 338 | 370 | 105 | 6 | 93 | 118 |
| SAT | 135 | 19 | 101 | 173 | 118 | 17 | 85 | 150 | 15 | 3 | 9 | 22 |
| Contrast | -94 | 10 | -114 | -74 | -81 | 9 | -99 | -62 | -6 | 1 | -9 | -3 |
| Correctness | 11 | 6 | 0 | 22 | 0 | 6 | -11 | 12 | 6 | 1 | 4 | 8 |
| Resp. Side | 18 | 6 | 6 | 31 | 0 | 8 | -16 | 17 | 17 | 4 | 10 | 24 |
| SAT × Contrast | -65 | 9 | -82 | -47 | -62 | 8 | -78 | -44 | -1 | 1 | -3 | 1 |
| SAT × Correctness | 71 | 12 | 48 | 96 | 75 | 14 | 49 | 105 | -1 | 1 | -4 | 2 |
| Contrast × Correctness | 26 | 15 | -3 | 55 | 3 | 16 | -28 | 35 | 13 | 3 | 8 | 18 |
| SAT × Contr. × Corr. | 153 | 32 | 90 | 216 | 164 | 33 | 95 | 228 | 1 | 3 | -5 | 7 |

Table 2.1.: Results of the LMMs in Experiment 1 for each latency measure. Coeff. indicates estimated coefficients of the LMM fitted on the log scale and back-transformed to the millisecond-scale. SE indicates the standard error for the Coeff. 2.5 and 97.5% indicate the lower and upper CrI around the Coeff. For further details, see the Statistical Procedure section.

implemented in the DDM), hence generating a negative correlation.

To address these concerns, the correlation between these $RT$ subcomponents was assessed separately for trials identified as fast-guesses and, the remaining trials, for which the correlation was computed across five different quantiles (namely, .1, .3, .5, .7, and .9). With respect to fast guesses, these were identified as before, with the exponentially weighted moving average (Vandekerckhove & Tuerlinckx, 2007) now based on the sorted distribution of $PMT$ rather than the distribution of $RT$ [4]. There was an average of 79.14 fast-guess trials per participant and SAT condition (range: 17-249). There was an average of 1065.93 remaining trials per participant and SAT condition (range: 786-1165), to be divided in 5 quantiles.

For the trials identified as fast guesses, we computed Spearman correlations between $PMT$ and $MT$ within each participant and SAT condition, and submitted these values to an LMM with SAT emphasis as a predictor[5]. The Accuracy condition was coded as 0 and Speed as 1, thus allowing us to interpret the intercept of the LMM as the mean correlation value when accuracy is emphasized and the slope (effect) as the change in this mean correlation when speed is emphasized. Fast guesses presented a negative mean correlation in the accuracy condition, as shown by the intercept of the LMM (m = -0.22, 2.5% = -0.33, 97.5% = 0.10). The slope of the LMM (in other words, the change in this mean correlation when speed is emphasized) did not suggest that the correlation differed between conditions (m = -0.03, 2.5% = -0.15, 97.5% = 0.09) (see Figure 2.3).

For the correlation analysis along quantiles, the likelihood of the mean correlation coefficient at each quantile was assessed with a Monte-Carlo procedure. We computed the Spearman correlation between draws from two random variables following a normal distribution (with mean = 0 and SD = 1) for 14 simulated participants divided into two conditions, and 5 bins of data with the same amount of trials as the real data; this procedure was repeated 1,000 times. This non-parametric analysis was motivated by the fact that we did not have a specific hypothesis (e.g. a linear trend) for the effect of the quantiles on the mean correlation value. The first quantile in the speed condition, was outside the range of expected values for the normal variables (Figure 2.3). Likewise, the first quantile in the accuracy condition was also outside the expected range, even if less negative than in the speed condition. The other correlation values do not differ from random levels. To illustrate the relationship between $PMT$ and $MT$ at the participant level, in Appendix A.4 we provide the scatter plots for the first 5 participants across the 5 quantiles.

---

[4]Both applications of the method revealed a strong but not perfect correlation on the amount of censored trials ($r(28) = 0.76$, $p < .001$).

[5]Note that, given low number of points (N = 14), this LMM only included the by-participant intercept as a random effect
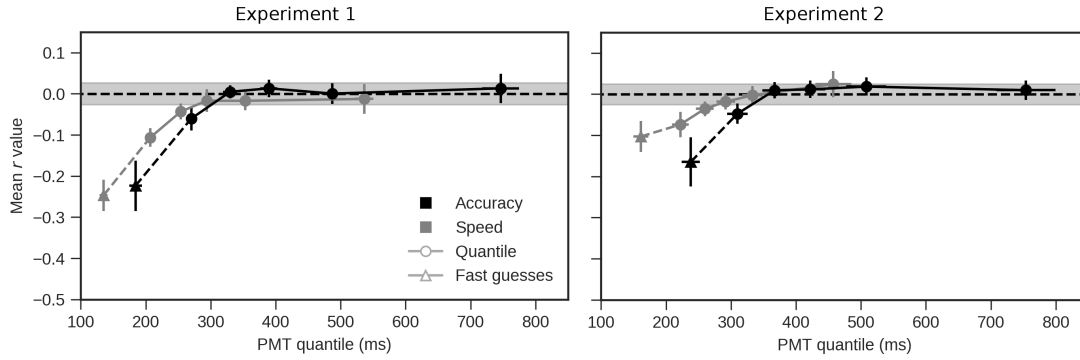
Figure 2.3.: Mean Spearman correlations between $PMT$ and $MT$ for the trials iden-
tified as fast-guesses (triangles), and the $PMT$ quantiles (circles) across
both speed (gray) and accuracy (black) conditions for Experiment 1 (left)
and Experiment 2 (right). Error bars represent one standard error of the
mean. Shaded intervals represent 95 % confidence intervals of the correla-
tion coefficient based on 1000 draws of the simulated random, normally
distributed variables.

### 2.3.3. Discussion

With respect to the $RT$s, the different manipulated factors led to clear results that
were congruent with previous works discussed in the Introduction. As theories for the
potential effects on $PMT$ and $MT$ are less developed, these results will be discussed
in more detail in this section.

Firstly, consistent with previous studies, the SAT manipulation significantly affected
$RT$ and response accuracy (the latter is reported in Appendix A.3). In agreement with
a decisional locus of SAT, this factor also had a large impact on pre-motor time ($PMT$).
However, the effect of SAT was not restricted to $PMT$, as this factor also affected
$MT$. This latter observation replicates recently reported results (Spieser, Servant,
Hasbroucq, & Burle, 2017; Steinemann, O'Connell, & Kelly, 2018) and is thus taken to
be robust. Furthermore, the effect of SAT on $MT$ is large, accounting for 11% of the
whole SAT effect measured on $RT$ (it was up to 20% in Spieser, Servant, Hasbroucq, &
Burle, 2017).

The issue of how to properly account for observed $RT$ variation between correct
and error responses, has a long history in the modelling of decision making. As early
decision-making models could not account for errors resulting in different latencies,
additional parameters were added (for example to the DDM: Ratcliff & Tuerlinckx,
2002). In our EMG decomposition, firstly we see that $PMT$ was shorter for errors
than correct responses when speed was stressed. This suggests that errors tend to be
made based on either shorter encoding times and/or shorter decision times. Second,
replicating previous reports in other task-settings (e.g. Allain, Carbonnell, Burle, Has-
broucq, & Vidal, 2004; Rochet, Spieser, Casini, Hasbroucq, & Burle, 2014; Roger, Núñez

Castellar, Pourtois, & Fias, 2014), $MT$ was longer for errors than for correct responses. Such longer $MT$s on errors are associated with a modulation of the EMG burst leading to the response (not explored here, but see: Allain, Carbonnell, Burle, Hasbroucq, and Vidal, 2004; Rochet, Spieser, Casini, Hasbroucq, and Burle, 2014; Śmigasiewicz, Ambrosi, Blaye, and Burle, 2020) which has prompted the interpretation that they could reflect a desperate attempt to stop the incorrect response, hence revealing an on-line process of cognitive control. To date, no decision making models incorporate such a process. The effect magnitude estimated for the reference condition (intermediate contrast) was a modest 6 ms, which may suggest that it is a reasonable simplification to not aim to account for this effect in decision making models. However for easy stimuli, the difference between correct and incorrect trials was threefold (mean $MT$ difference equal to 18ms in the condition emphasizing accuracy, see Figure 2.1), which highlights the importance of considering this effect, depending on the experimental conditions. Most importantly, this observation suggests that motor processes can indeed be affected by experimental manipulations that have been considered to be purely "decisional" in previous research.

Stimulus contrast has been classically considered to affect evidence accumulation processes (*e.g.* Palmer, Huk, and Shadlen, 2005b, experiment 5). In agreement with this view, its effect on $PMT$ was clear and very similar in magnitude to that observed on total $RT$. More surprisingly, a small but highly reliable effect of stimulus contrast was observed on $MT$, in the same direction as on $PMT$. The presence of perceptual effects on motor processes has been previously debated. For instance, in using a double response paradigm, Ulrich and Stapf, 1984 showed that increasing stimulus duration shortened $RT$ but also increased output response force. In three separate (unpublished) experiments, Grayson, 1983 also reported evidence that higher signal intensities shorten $MT$ in a simple reaction time task. More recently, Servant, White, Montagnini, and Burle, 2015 reported color saturation effects on $MT$ in a color discrimination conflict task. In contrast, other studies have reported that $MT$ is unaffected by stimulus intensity (Bartlett, 1963; Miller, Ulrich, & Rinkenauer, 1999; Smith, 1995). For example, Smith, Anson, and Sant (1992, unpublished manuscript, cited in Smith, 1995) reported the invariance of mean motor time across stimulus conditions. Miller, Ulrich, and Rinkenauer, 1999 also reported no effect of stimulus intensity on the lateralized readiness potential nor on the EMG-based $MT$ in two independent experiments using a forced choice task[6]. In the present study, the statistical robustness, along with the linear trend observed across contrast levels, leave no doubt that this effect exists in the data sets acquired. The question remains open, however, as to whether this effect reflects a "cognitive" or "energetic" process (see Sanders, 1983). For our current purpose, and irrespective of the origin of the effect, the important aspect is that the factor contrast affects $PMT$ and $MT$ in a similar direction.

---

[6]There is potentially a very serious flaw in the EMG recording of this study. The signal was low-pass filtered at 500 Hz before being sampled at 250 Hz. According to the Shannon-Nyquist theorem, the minimal sampling frequency given the filtering should have been at 1000 Hz. As a result, strong aliasing of the signal may have occurred, which could jeopardize the validity of its conclusions.

Response side affected the $RT$s, with longer $RT$s for the left than the right hand. This effect was selectively localized to $MT$ rather than $PMT$. The effect is likely due to a left-right difference in the innervation of the motor units (Schmied, Vedel, & Pagni, 1994)[7]. Such a pure motor laterality effect (no effect on the $PMT$) indicates that, contrary to what the $RT$ data may have suggested, decision latencies are independent of the hand with which the response is given.

Finally, $PMT$ and $MT$ were not significantly correlated in the vast majority of $PMT$ quantiles. It is important to dissociate the correlations between-trials, reported in the results, and the correlations between-participants, which would be computed on average measures per participant. The former were close to zero, which is a necessary, if not sufficient, condition for stochastic independence between the two measures. The latter, between participants, appears to be positive, indicating that slow participants tend to be slow on both decisional and motor components. To the best of our knowledge, the absence of a significant between-trial correlation is formally reported here for the first time, but it was already observed in previous datasets (unpublished observations made on published data, *e.g.* Burle, Possamaï, Vidal, Bonnet, and Hasbroucq, 2002) instilling confidence on its reliability. A more detailed analysis showed that there is a negative correlation between $PMT$ and $MT$ on the early quantile of the $PMT$ distribution, irrespective of the SAT instructions (Figure 2.3). Such modulation of the correlation pattern could indicate that the trials are not generated from the same architecture across the quantiles of either condition. This interpretation is in line with the observation that trials identified as fast-guesses also present a negative correlation. We come back to this issue in the General Discussion.

In summary, $MT$ appears to be substantially affected by various experimental factors. These results were robust, and most of them were consistent with previously reported, or unreported, findings. Before interpreting these observations any further, however, we take up the issue that the force threshold for triggering a response was rather high in this experiment, which motivated Experiment 2. A high force might lengthen $MT$ in such a way that it becomes modulated by parameters that do not affect it in more canonical decision settings. The other important limitation of the experiment, potentially connected to the high force setting, was there being a high rate of trials exhibiting multiple EMG activations. Repeated muscle triggering during very short intervals could modify the activation dynamics of the cortical and spinal neurons, as well as the excitability of the neuromuscular junction. This could result in a mis-estimation of the motor time for these trials, compared with trials showing a single EMG activation. To address these concerns, we hence sought to replicate our findings in a second experiment with lower response force requirements. By reducing the force required, we expected to, hopefully selectively, affect the motor components ($MT$s and $T_r$), and reduce the rate to which trials with multiple EMG activations occur (Burle, Possamaï, Vidal, Bonnet, & Hasbroucq, 2002).

---

[7]A later re-analysis after acceptance of the manuscript showed that this left/right difference was linked to a bias in the force button used

## 2.4. Experiment 2

The main goal of Experiment 2 was to replicate and extend the previous results by refining the design. Specifically, this experiment differs from the first one based on the following two adjustments. First, the force threshold needed to respond was divided by 3 (from 6 to 2 N). Second, in order to increase the total number of trials per design cell (participant × SAT × contrast level) from 240 to 432, the number of contrast levels was lowered (from 5 to 3). Higher trial counts would allow us to reject trials with multiple EMG activity while keeping a sizeable amount of trials.

### 2.4.1. Methods

#### 2.4.1.1. Participants

Sixteen participants (8 men and 8 women, mean age = 23.6, 1 left-handed) were recruited. None had participated in Experiment 1. They were all students from Aix-Marseille University. All reported having normal or corrected vision and no neurological disorder. Participants gave their informed written consent according to the declaration of Helsinki and were compensated at a rate of 15 € per hour.

#### 2.4.1.2. Procedure

All participants completed a single experimental session comprised of 24 blocks with 108 trials each (2592 trials per participant). Session duration was similar to Experiment 1 (~ 1h30), including an initial training component of 15 minutes and self-paced breaks between each block. The duration of the training session was shortened compared to Experiment 1, because asymptotic performance was reached quickly. Contrast levels were chosen from Experiment 1, targeting a full range of performance from almost-chance level to near-perfect (i.e. 0.01, 0.07, 0.15). The statistical procedures, EMG recordings, and processing techniques were the same as in Experiment 1.

#### 2.4.1.3. Statistical analysis

As a part of data analysis, no participants presented conditions for exclusion. However, during data collection, two participants were stopped for excessively high tonic activity in the EMG that could not be reduced. As for the identification of fast guesses and value transformations (log transform for $RT$, $PMT$ and $MT$), these were performed as in Experiment 1. The same factor coding features and priors were applied in the LMM models.

## 2.4.2. Results

The upper limit of 1500 ms for the $RT$s resulted in the removal of less than 1% of trials. Next, trials with low signal-to-noise ratio, high spontaneous tonic activity, or multiple activities led to the exclusion of 14.18% of trials. With regard to the effectiveness of lowering the force needed to respond, the rate of occurrence of trials with multiple EMG activations was successfully reduced in this experiment to 12.73% (from 25.90% in Experiment 1). After exclusion of the multiple activity trials, EMG onsets detected by the algorithm had to be visually corrected on 5.9% of the remaining trials. Additionally, the EWMA method removed 7% of trials taken to be fast guesses. As was done for Experiment 1, the analysis of the error rates is provided in Appendix A.3, and the overall pattern of variation on RT, PMT, and MT is represented in Figure 2.1 (insets). All discussed LMM parameters can be found in Table 2.2 and are represented in Figure 2.2.

### 2.4.2.1. $RT$ and $PMT$

As in Experiment 1, $PMT$ and $RT$ followed the same trends, as reflected by effect estimates with the same sign and comparable magnitudes (Table 2.2, Figure 2.2). Both $PMT$ and $RT$ were faster with increases in contrast or when speed is stressed, and all CrIs of the interactions excluded 0 as a plausible effect, with the exception of the interaction between correctness and contrast. The only notable difference with Experiment 1 was the result of response side not significantly affecting $RT$, which was localized to an $MT$ effect [8].

### 2.4.2.2. $MT$

All of the effects observed in Experiment 1 were replicated, except the effect of response side (Table 2.2). This is shown As Figure 2.2, where all estimates are of close magnitude across experiments, and the corresponding CrIs largely overlap, except for the response side factor. We hence successfully replicated the finding that $MT$ is sensitive to SAT, correctness, and contrast; and that the correctness and contrast factors interact.

### 2.4.2.3. Correlations between $PMT$ and $MT$

As in Experiment 1, we again computed the Spearman correlations on the fast-guess trials, as identified using the EWMA method on the $PMT$ distribution, as well as on the quantiles of the $PMT$ distribution that excludes fast guesses. As before, fast guess trials are associated with a negative correlation between $PMT$ and $MT$ in accuracy (m = -0.17, 2.5% = -0.26, 97.5% = -0.06); emphasizing speed over accuracy did not change the negative correlation (m = 0.07, 2.5% = -0.08, 97.5% = 0.20). The by-quantile

---

[8]As for Experiment 1, the frequentist replication shows the same results except the significant interaction between contrast and correctness for $PMT$ ($\beta$ = -0.06, $t$ = 2.45)

| Predictor | RT | | | | PMT | | | | MT | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coeff. | SE | 2.5% | 97.5% | Coeff. | SE | 2.5 | 97.5% | Coeff. | SE | 2.5% | 97.5% |
| Intercept | 449 | 29 | 393 | 511 | 368 | 26 | 319 | 417 | 72 | 5 | 63 | 83 |
| SAT | 160 | 27 | 108 | 213 | 152 | 27 | 100 | 206 | 10 | 2 | 5 | 14 |
| Contrast | -94 | 12 | -118 | -70 | -82 | 11 | -104 | -60 | -6 | 1 | -7 | -4 |
| Correctness | 9 | 7 | -5 | 24 | -3 | 7 | -16 | 12 | 5 | 1 | 4 | 7 |
| Resp. Side | 3 | 8 | -14 | 19 | 1 | 6 | -11 | 13 | 0 | 3 | -5 | 6 |
| SAT × Contrast | -74 | 8 | -91 | -58 | -72 | 8 | -88 | -55 | 1 | 1 | -1 | 3 |
| SAT × Correctness | 77 | 13 | 53 | 102 | 82 | 14 | 55 | 110 | -3 | 1 | -5 | 0 |
| Contrast × Correctness | 14 | 13 | -9 | 40 | -5 | 12 | -28 | 19 | 9 | 1 | 6 | 12 |
| SAT × Contr. × Corr. | 142 | 27 | 92 | 200 | 158 | 30 | 100 | 218 | -5 | 2 | -9 | -1 |

Table 2.2.: Results of the LMM models performed on each latency measure from Experiment 2 . Coeff. represents estimated coefficient of the LMM fitted on the log scale and back-transformed to the millisecond scale. SE represent standard error for the estimate; 2.5% and 97.5% represent, respectively, the lower and upper CrI around the estimate. The intercepts correspond to the mean value of the latency when all the predictors are kept at 0. Main effects can be interpreted as the change in the mean value when the other predictors are kept at null value. Interactions can be read as the change in the effect of the main effects when the other variable is added (cf. section on Statistical analysis of Experiment 1).

analysis revealed the same pattern as in Experiment 1, where early quantiles are associated with a negative correlation (Figure 2.3). Similarly as before, two quantiles also fell below the random simulation results in the speed condition, while only the first quantile did so in the accuracy condition.

## 2.5. Discussion

This experiment successfully replicated the principal results observed in Experiment 1. With the response force settings being lower, shorter $MT$ durations were obtained, and these force conditions are more resembling to the canonical publications in the field of decision making. Under these circumstances, the fact that we observed a sensitivity of motor processes (indexed on the basis EMG activations) to SAT instructions, stimulus contrast, and response correctness, indicates that these effects are not an artifact induced by the requirement of a large response force. Moreover, the exclusion of the multiple activity trials in Experiment 2 did not lead to a pattern of results that were different from those reported in Experiment 1. The only noticeable difference was the disappearance of the response-side effect on $MT$. This could indicate that, under low force requirements, the left-right difference in the innervation of motor units (Schmied, Vedel, & Pagni, 1994) can be functionally compensated for.

The negative correlation between $PMT$ and $MT$ on the early quantiles of the $PMT$ distribution was also replicated, confirming that the temporal relationship between "decisional" and "motor" processes changes with the duration of the processes contained in the $PMT$.

Overall, these results consolidate the interpretation that motor processes in decision making are not fixed ballistic processes, and that the factors thought to affect decision processes can also impact motor-related components. We will come back to the size of these effects in the General Discussion.

Having reliably established, from two experiments, how these experimental conditions affect decisional and motor components of $RT$s, it is worthwhile to explore the extent to which the parameters of a formal decision making model may covary with these EMG-based decompositions. We carry out this analysis with the DDM, which also aims to decompose $RT$s into decision and non-decision processes.

## 2.6. Modeling

### 2.6.1. Method

The DDM (thoroughly reviewed by Ratcliff & McKoon, 2008) was fitted to the $RT$s obtained in Experiments 1 and 2. We first determined the model that best fitted the behavioral data of both experiments. Then, to assess whether the EMG and model-based decompositions may lead to the same conclusions, we evaluated how the variation of $T_{er}$ across participants co-varied with the variation of $MT$ across participants. If such

co-variation is present, modelers can presumably get some information about the motor system in the absence of EMG activity recordings by comparing $T_{er}$ parameters between (groups of) participants. Finally, in order to formally probe the relationship between non-decision time (DDM) and motor processes (EMG), we estimated a linear dependency between $T_{er}$ and $MT$ within the best-fitting model.

As for the behavioral analysis in Experiment 1 and Experiment 2, the data used for the following modeling section was filtered for fast-guesses using the EWMA method and trials with RT longer than 1500 ms or with ambiguous EMG onset were discarded. No other filtering was applied.

### 2.6.1.1. Model estimation

We used a hierarchical Bayesian estimation method for the model fit. As discussed in the Methods of Experiment 1 (section on the behavioral analysis), a hierarchical Bayesian estimation of a cognitive model assumes that each parameter of a participant is drawn from a population distribution described by hyper-parameters, often the mean and variance of a normal distribution. This method preserves the uncertainties associated with the parameter values (due to its Bayesian procedure) while sharing the information between participants to estimate individual parameters (due to its hierarchical nature). As was done for the LME in the behavioral analysis, we only report the hyper-parameter of the estimated population mean for each parameter.

We used the implementation of a hierarchical Bayesian DDM provided in the *HDDM* python package (Wiecki, Sofer, & Frank, 2013). For each model, both in the "Model selection" section below and the model including $MT$ as a covariate, we ran 18500 burn-in samples and 1500 actual recorded samples across four Markov chains Monte-Carlo (MCMC). We inspected each parameter of each chain visually to assess whether they reached their stationary distribution, and whether the $\hat{R}$ (Gelman & Rubin, 1992) was under the conventional threshold of 1.01. Additionally, we examined the autocorrelation of each chain to ensure that samples were drawn independently. For the priors, because our design is canonical and in order to ease convergence, we used the default informative priors used in HDDM based on the work of Matzke and Wagenmakers, 2009.

### 2.6.1.2. Model selection

We designed a base model and added parameters according to our hypothesis. The base model was chosen based on previous studies and on the data reported in the previous section. For this base model, the boundary parameter was free to vary with SAT instructions, as it is thought to capture SAT changes. The drift rate was free to vary with the contrast, as this parameter has been shown to be associated with stimulus strength. The non-decision time was free to vary with SAT, as it has been observed (including in the present report) that this parameter also varies with SAT conditions (Palmer, Huk, & Shadlen, 2005b; Ratcliff, 2006; A. Voss, Rothermund, &

Voss, 2004). The accumulation starting point was assumed to be constant because the boundaries were accuracy-coded (correct and incorrect). We also added inter-trial variability of the drift rate and the non-decision time, because of their ability to reduce the influence of contaminant fast-trials (Lerche, Voss, & Nagler, 2017). Finally, we added the inter-trial variability of the starting point parameter which was free to vary with SAT instructions, because the data analysis clearly shows that the speed of errors compared to the speed of correct responses does change according to the SAT condition. Almost all parameters were estimated individually with the constrain of being drawn from a common normal distribution (or half-normal depending on the boundaries, *e.g.* variability parameters cannot have a negative value). Only the inter-trial variability parameters of the drift rate, of the starting point, and of the non-decision time were estimated at the group-level because they are notoriously difficult to estimate (Boehm, Annis, Frank, et al., 2018; Wiecki, Sofer, & Frank, 2016).

In addition to the base model, we also tested the following hypothesis, and combinations thereof: whether the drift rate also varies with SAT, as in Rae, Heathcote, Donkin, Averell, and Brown, 2014, whether the $T_{er}$ varies with the response side, as was observed for $MT$, and whether the model needs to account for potential bias in the starting point of accumulation [9]. The various possible combinations of these hypothesis is summarized in Table .4.

We used the deviance information criterion (DIC) to select among competing models. The DIC is an analog to the Akaiake information criterion (AIC) generalized to the hierarchical Bayesian estimation method, in which the improvement of the log-likelihood is weighted against the cost of additional parameters. Because, it has repeatedly been shown that DIC tends to select over-fitted models, we also report for each model the Bayesian predictive information criterion (Ando, 2007, BPIC). BPIC is intended to correct DIC's bias in favor of over-fitted models by increasing the penalty term for the number of parameters. For all these measures, a lower value of DIC or BPIC indicates a preferred model.

### 2.6.1.3. Assessing covariance

There are two reasons why we cannot use standard correlation coefficients to evaluate between-participant covariance of $T_{er}$ and $MT$. First, by taking a point estimate, a correlation coefficient ignores the uncertainty associated with the parameter estimation procedure. Second, point estimates are assumed to be independent from one another, which is not the case when using a hierarchical estimation method. To overcome these two issues, we used the plausible values method developed in Ly, Boehm, Heathcote,

---

[9]Estimating the starting point required a change in the coding of the boundaries, from "correct" and "incorrect" to "left" and "right" responses. This change in coding does also change the meaning of the drift rate as it will represent evidence in favor of left/right instead of evidence for correct/incorrect. In order to keep the same meaning, and to avoid estimating a drift rate for each side (times the number of stimulus strength levels), we simply took the positive or negative sign according to the side of the correct response. Note that this last modification does not allow to recover a left/right bias in the drift rate but still allows to estimate a starting point bias between left and right responses

et al., 2017. This method consists in drawing participants' parameters (*i.e.* plausible values) from the posterior distribution and correlating them with the variable of interest at each draw. In order to generalize this sample plausible correlation distribution to the population, we then used an analytic posterior method (see Ly, Boehm, Heathcote, et al., 2017) using an R code provided with the evidence accumulation model fitting package Dynamic Model of Choice (Heathcote, Lin, Reynolds, et al., 2019).

To evaluate the by-trial covariance between $T_{er}$ and $MT$ we use the implementation of a regression model provided by the HDDM package. This implementation allows to specify a by-trial linear relationship between a model parameter and an external covariate

$$T_{er} = \alpha + \beta MT \tag{2.5}$$

where $\alpha$ is the intercept of the linear model, *i.e.* the mean $T_{er}$ when the (centered) $MT$ is at 0, and $\beta$ is the slope of the predictor $MT$, hence the linear relationship between $T_{er}$ and $MT$. This by-trial covariance was estimated for each participant individually, while keeping the hierarchical constrain.

## 2.6.2. Results and discussion

### 2.6.2.1. Analysis of the model

The model selection analysis shows that including a modulation of the starting point always leads to lower criteria and that allowing the drift rate to vary with SAT conditions leads to lower DIC and lower BPIC. We therefore chose to use the model with starting point estimation and drift free to vary between stimulus strength level and SAT conditions for both experiments (see Appendix A.6 for the complete analysis of the model selection procedure). A summary of the estimated values of the parameters from the winning model is presented in Table 2.3, and an average representation of the goodness-of-fit for the selected model is provided in Figure 2.4. We provide a representation of the posterior distributions for each parameter that was free to vary in the different conditions in Figure 2.5. Table 2.3 and Figure 2.5 reveal a remarkable stability in the estimation of parameters across the two experiments, despite them involving two different samples of participants. We will not provide here a precise description of how the different manipulations map onto DDM parameters, because the manipulations and the results are congruent with those reported in the literature.
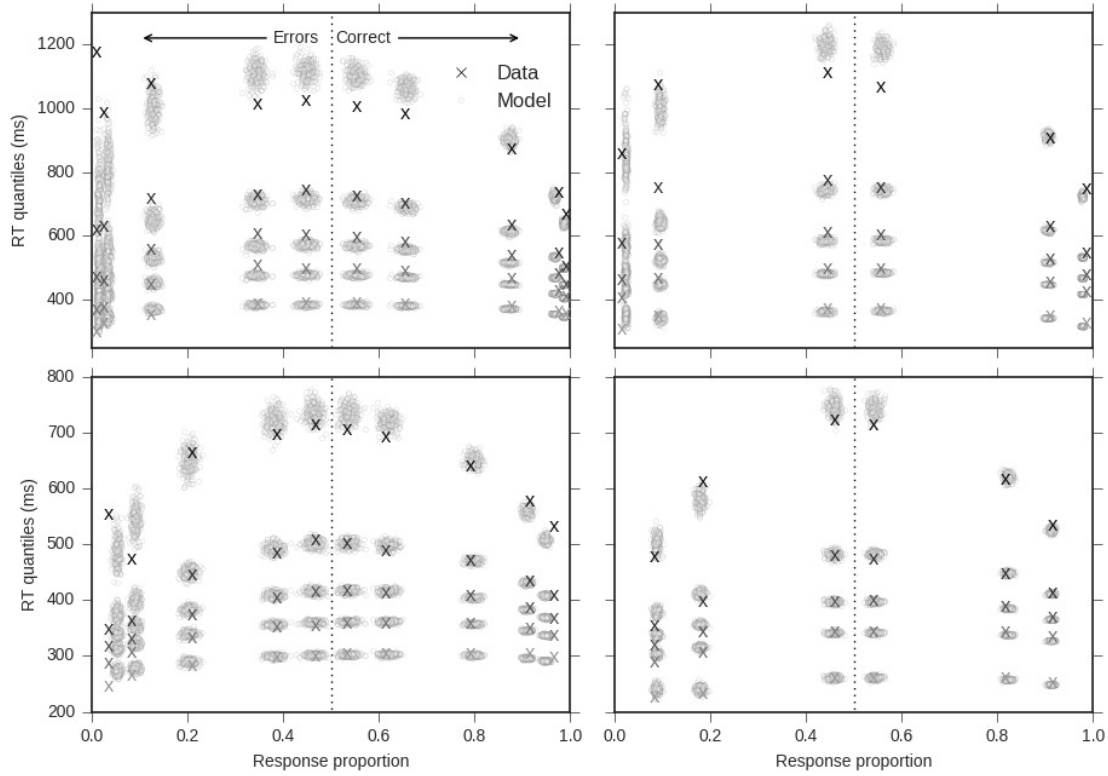
Figure 2.4.: Quantile-probability plots (Ratcliff & McKoon, 2008) for Experiment 1 (left column) and Experiment 2 (right column), in the accuracy (upper row) and speed (lower row) conditions, computed from the winning model. The X-axis displays obtained response proportion across contrast levels, symmetrically for errors (left side) and correct responses (right side). The Y-axis displays the fitted (dot) and observed (cross) $RT$ binned in 5 quantiles (.1, .3, .5, .7 and .9 quantiles, from bottom to top). Observed response proportion and $RT$ quantiles were computed from values pooled across participants. Model predictions were obtained by drawing 500 parameter values from the joint posterior distribution and computing their associated predicted performance. The substantially larger misfits observed on the far left of the QP-plot for Experiment 1 in the accuracy condition are likely due to the lower amount of errors in these conditions.

We started with a comparison between the SAT effect size quantified in the winning model and in the EMG results, bearing in mind that both estimates stem from different methods: subtraction of both SAT levels for $T_{er}$ vs. LMM estimate for $MT$. The mean estimated SAT effect on the $T_{er}$, computed on the subtraction of estimated $T_{er}$ in speed from the estimated $T_{er}$ in accuracy at each MCMC iteration, was of 38 ms (2.5% = 11 ms, 97.5% = 64 ms) in Experiment 1 and 42 ms in Experiment 2 (2.5% = 0 ms, 97.5% = 84 ms). For $MT$, the observed effects were 15 ms (2.5% = 9 ms, 97.5% = 22 ms)

| | Experiment 1 | | | Experiment 2 | | |
|---|---|---|---|---|---|---|
| | mean | 2.5% | 97.5% | mean | 2.5% | 97.5% |
| $a_{Acc}$ | 1.18 | 1.10 | 1.26 | 1.23 | 1.12 | 1.34 |
| $a_{Spd}$ | 0.93 | 0.86 | 1.01 | 0.88 | 0.78 | 1.00 |
| $v_{Acc1}$ | 0.22 | -0.09 | 0.55 | 0.24 | -0.16 | 0.64 |
| $v_{Acc2}$ | 0.70 | 0.37 | 1.02 | | | |
| $v_{Acc3}$ | 2.04 | 1.70 | 2.38 | 2.28 | 1.89 | 2.69 |
| $v_{Acc4}$ | 3.52 | 3.17 | 3.89 | 3.94 | 3.53 | 4.36 |
| $v_{Acc5}$ | 4.52 | 4.14 | 4.90 | | | |
| $v_{Spd1}$ | 0.21 | -0.12 | 0.54 | 0.27 | -0.13 | 0.68 |
| $v_{Spd2}$ | 0.70 | 0.34 | 1.03 | | | |
| $v_{Spd3}$ | 2.04 | 1.70 | 2.39 | 2.43 | 2.02 | 2.84 |
| $v_{Spd4}$ | 3.61 | 3.25 | 3.98 | 4.21 | 3.79 | 4.63 |
| $v_{Spd5}$ | 4.79 | 4.41 | 5.17 | | | |
| $T_{erAcc}$ | 0.33 | 0.31 | 0.35 | 0.33 | 0.30 | 0.36 |
| $T_{erSpd}$ | 0.29 | 0.28 | 0.31 | 0.28 | 0.25 | 0.32 |
| $sv$ | 0.90 | 0.78 | 1.02 | 0.72 | 0.60 | 0.84 |
| $sz_{Acc}$ | 0.17 | 0.02 | 0.29 | 0.25 | 0.15 | 0.33 |
| $sz_{Spd}$ | 0.65 | 0.60 | 0.68 | 0.67 | 0.64 | 0.71 |
| $st$ | 0.09 | 0.08 | 0.09 | 0.07 | 0.06 | 0.07 |
| $z$ | 0.54 | 0.52 | 0.56 | 0.50 | 0.48 | 0.52 |

Table 2.3.: Posterior distributions for the estimated parameters of the selected DDM. 2.5% and 97.5% represent boundaries of CrI of the mean. Note that $T_{er}$ is on the second, rather than millisecond, scale.

and 10 ms (2.5% = 5 ms, 97.5% = 14 ms), respectively, in Experiments 1 and 2. If we were to assume a direct link between $T_{er}$ and $MT$, we would need to account for the observation that the effect on $T_{er}$ was approximately twice the effect observed on $MT$. One *ad-hoc* assumption would be that encoding processes are also affected by the SAT instructions, to a comparable extent, however we cannot yet exclude the possibility that the SAT effect could also/partially be contained in motor processes that precede the EMG activity onset.
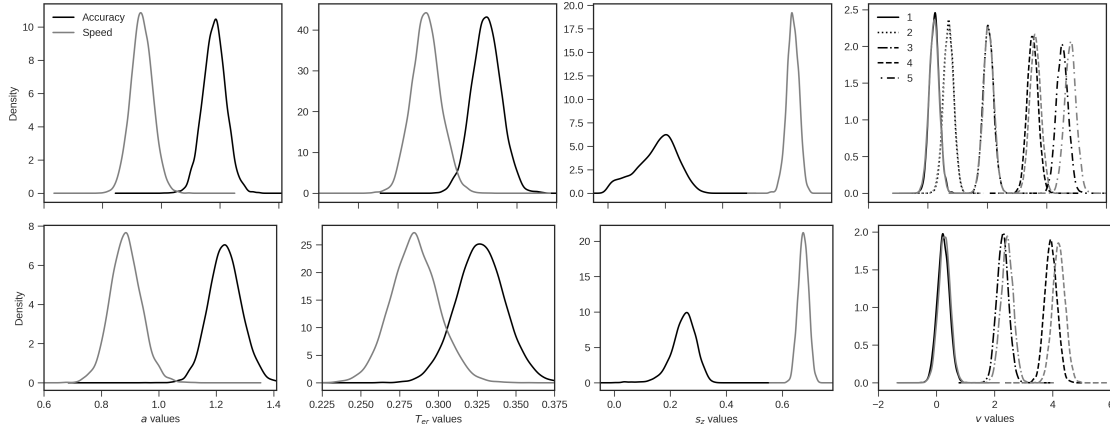
Figure 2.5.: Posterior distributions of the parameters left free to vary across conditions. Top row is for Experiment 1, bottom row for Experiment 2. $a$, $T_{er}$, $s_z$, $v$ stand for boundary separation, non-decision time, inter-trail variability of the starting point and drift rate. For the drift rate (last panel), numbers in the legend represent contrast levels.

### 2.6.2.2. Between-participant dependence of $MT$ and $T_{er}$

Having estimated a $T_{er}$ parameter for each SAT condition, we can perform the plausible population correlation separately in each condition and for each experiment. In Experiment 1 the plausible population correlation was positive and high, but with high uncertainty in both the accuracy (m = 0.60, 2.5% = 0.17, 97.5% = 0.85) and the speed conditions (m = 0.56, 2.5% = 0.11, 97.5% = 0.83). In Experiment 2, in contrast, the plausible population correlation was lower and its CrI contained 0 in both accuracy (m = 0.24, 2.5% =-0.24, 97.5% = 0.63) and speed conditions (m = 0.20, 2.5% =-0.28, 97.5% = 0.60). Although posterior distributions of plausible correlation values clearly overlap between Experiment 1 and 2 (Figure 2.6), hindering any strong conclusion, we speculate that the higher values in Experiment 1 could be linked to the higher force requirement therein. We also observed that the pattern is consistent within the experiments where values are always higher for accuracy (which yields longer $MT$) than speed conditions. Hence, when a modeller is searching to interpret the $T_{er}$, it is highly likely that the amount of information about the motor system one can get by comparing $T_{er}$ parameter between participants depends on the amount of time spent on executing the response.
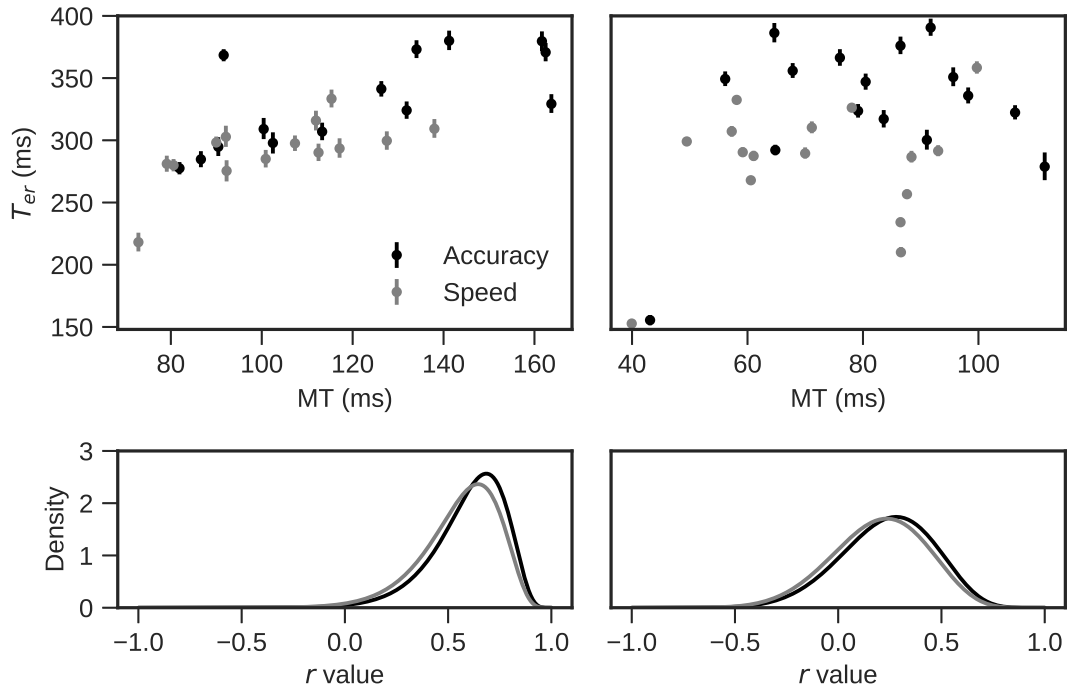
Figure 2.6.: Correlation between the estimated $T_{er}$ and the measured $MT$, computed
by participant.
Upper row: scatterplots of the $MT$ and $T_{er}$ pairs for each participant; lower
row: plausible population correlation density. Left column: Experiment 1;
right column: Experiment 2.

### 2.6.2.3. Within-participant link between $MT$ and $T_{er}$

We added the $MT$ as a covariate of the $T_{er}$ to the models selected for Experiments 1
and 2. The estimated slope linking $T_{er}$ and $MT$ indicates that $MT$ weakly but con-
sistently covaries with $T_{er}$ in Experiment 1 ($\beta = 0.22$, 2.5q = 0.10, 97.5q = 0.34) and in
Experiment 2 ($\beta = 0.27$, 2.5q = 0.12, 97.5q = 0.41). While the population estimate is
rather concentrated, at the individual level the mean of the posterior distribution on
the $\beta$ coefficient varies from -0.10 to .78 across both experiments. This would indicate
that, for some participants at least, measured $MT$ does not covary with estimated $T_{er}$,
although it should be noted that this could also be due to measurement error.

### 2.6.2.4. Source of between participant variation in the by-trial covariance

The previous section showed that there is a high variability, between participants, in
the degree to which $T_{er}$ and by-trial $MT$ covary. We thought to link this variability
with the previous finding from the EMG section which shows that, for fast decision,
$PMT$ and $MT$ are negatively correlated. Such correlation violates the independence

assumption of the model. One can hence assume that the larger this violation, the lower the capacity of the model to fit the data and to correctly recover the parameters. As a consequence, the relationship between $T_{er}$ and $MT$ should be all the weaker that the negative correlation is high. In order to test this hypothesis we took the average Spearman correlation value between $PMT$ and $MT$ across all trials in both SAT conditions for each participant. We then test the plausible population correlation between the estimated $\beta$ of the linear regression between $T_{er}$ and $MT$ and the correlation between $PMT$-$MT$.

From this analysis (see Figure 2.7) we see that there is a high association between the regression $\beta_{MT}$ and the value of the correlation between $PMT$ and $MT$, both in Experiment 1 (m = 0.50, 2.5% =0.03, 97.5% = 0.81) and even stronger in Experiment 2 (m = 0.70, 2.5% = 0.36, 97.5% = 0.89). Hence this analysis suggest that the less negative the correlation between $PMT$ and $MT$ is, the better $T_{er}$ recovers $MT$.
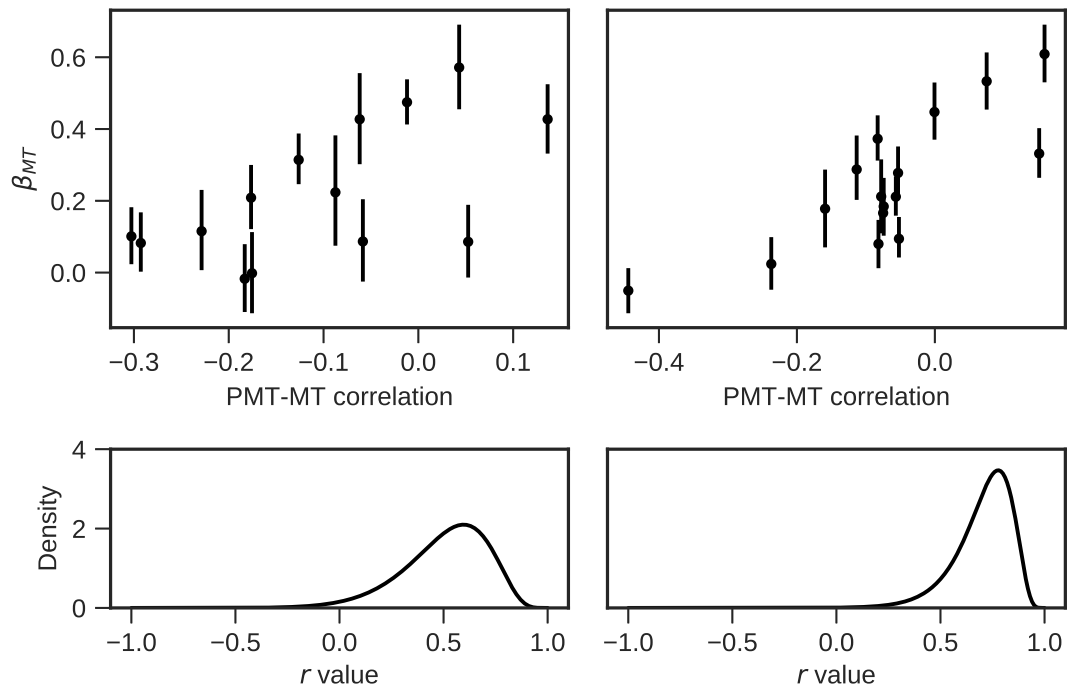


Figure 2.7.: Representation of the correlation between the regression coefficient of $MT$ over $T_{er}$ plotted against the correlation between $PMT$ and $MT$. Upper panels depicts scatterplots of both variables while lower panels represent plausible population correlation density for Experiment 1 (left panel) and Experiment 2 (right panel).

## 2.7. General Discussion

Our understanding of mental functions is largely owed to chronometric analyses of performance (Luce, 1986). Both experimental and mathematical psychologists assume that response times ($RT$s) result from different subprocesses, although the nature and the temporal relationship between these processes has long been, and still is, largely debated (e.g. Dubarry, Llorens, Trébuchon, et al., 2017; McClelland, 1979; Miller, 1988; Requin, Riehle, & Seal, 1988; Sternberg, 1969). A minimal, commonly accepted form of $RT$ decomposition assumes that response latencies result from (at least) two components: a decision time ($DT$) and a residual non-decision time ($T_{er}$), where the $T_{er}$ itself can be decomposed into an encoding time ($T_e$) and a motor execution time ($T_r$). In current and widely-used practice, these components and subcomponents are assumed to be additive and independent (see Eq. 1.4 and 1.8), two subsidiary assumptions embraced by many formal models of decision making since their original formulation (Stone, 1960).

While experimental/mathematical psychologists and neuroscientists aim to understand the processes leading to behavior, the main dependent variable, the $RT$, only gives access to the overall outcome of all those processes, and does not reveal, by itself, the individual underlying processes. While analyzing adequate experimental manipulations, and their interaction patterns, has proved very powerful in $RT$ research see, e.g. Sternberg, 1969, the validity of the inferences one can do on the individual processes is difficult to assess see, e.g. Pieters, 1983. To better study separately the underlying cognitive processes, two main approaches have been followed. The first consists in using physiological variables e.g. Coles, 1989; Requin, Riehle, and Seal, 1988 which are thought to better reflect the intermediate processes. Such approach, however, requires to explicitly define "linking functions" (Schall, 2004; Teller, 1984) mapping physiological measures into cognitive processes, which is not as easy as it may seem (Schall, 2019).

The second approach is to use process models that are fitted to empirical data ($RT$s and error rates) to derive inferences about latent cognitive processes, which may be reflected in the estimated parameters. The validity of the inferences made hinges on the validity of the process model assumptions themselves. Despite the importance of assessing such validity, however, few empirical studies have directly put these assumptions to the test.

In recent years, several studies have yoked behavioral and physiological markers to constrain, and, hopefully, improve the recovery of underlying cognitive processes (Turner, van Maanen, & Forstmann, 2015; van Ravenzwaaij, Provost, & Brown, 2017). The perspective followed in the present work is (partially) different and motivated by the vast majority of studies where formal decision-making models are fitted to data solely on the basis of behavioral measures. We aimed at estimating the information provided by the non-decision parameter of a decision making model (namely, the DDM fitted on response times) about the motor processes captured through EMG decomposition. In other words, we used physiological data as an external validator,

to assess whether the inferences drawn from model fits are compatible with the independent information provided by the EMG decomposition.

As argued in the Introduction, for such validation to be effective, there must be a strong link between the motor execution part of the $T_{er}$ and the $MT$. Any discrepancy will either be generated by an incorrect linking function between $MT$ and $T_{er}$ (which is unlikely) or by an inappropriately specified model. Importantly, identifying a mis-estimation of the $T_{er}$ signals the risk of a mis-specification of the other parameters, because the correct estimation of any parameter depends on the correct estimation of the others.

In the present study, the two $RT$ decompositions, one physiological and the other model-based, were first performed separately, and then combined. We will hence first discuss the conclusions one may reach based on each approach individually, and then evaluate the added value of combining them.

## 2.7.1. Relationship between decisional and motor processes based on physiological fractionation of $RT$

Recent studies analyzing movement trajectories in reaching tasks have revealed that motor execution can be affected by decision processes (Buc Calderon, Verguts, & Gevers, 2015; Dotan, Meyniel, & Dehaene, 2018; Resulaj, Kiani, Wolpert, & Shadlen, 2009), thus questioning the assumption of independence between these two levels. Despite their theoretical importance, these prior results were obtained in experimental settings (reaching behaviors) that depart from the vast majority of studies (simple key-pressing) in which formal decision making models are fitted, and on which many cognitive inferences have been made. Such experimental settings can afford strategic adjustments like optimizing the distance to the targets according to the on-line evolution of the decision parameter.

In the present study, we tested the relationship between decisional and motoric processes using conventional setups but added a neurophysiological measure on each trial.

In the two experiments, we manipulated perceptual difficulty and SAT conditions, and, between experiments, the required response force was adjusted to further verify our findings. All manipulations were successful and produced clear and expected effects on the $RT$s. With the noticeable exception of response side (affecting performance in Experiment 1 but not in Experiment 2), all factors affecting the $RT$s did so by affecting $PMT$s. This was not surprising. More notably, $MT$ was also affected by several experimental manipulations that we now detail.

Stimulus contrast and SAT impacted not only $PMT$, but also $MT$; the effects on the two variables were in the same direction (i.e., shorter $PMT$s go with shorter $MT$s). This replicates the observation that SAT has an impact (approximately 10% of the SAT effect on RT and 14% of the mean $MT$) on motor execution (Spieser, Servant, Hasbroucq, & Burle, 2017; Steinemann, O'Connell, & Kelly, 2018). While effects of SAT on $T_{er}$ estimates are regularly observed, to date the genuine nature of the SAT effect

on $T_{er}$ (specifically motor processes) is still debated and yet to be validated (Lerche & Voss, 2018; Smith & Lilburn, 2020b). Our work, which assessed this phenomenon in a canonical, perceptual decision-making task with an isometric responding format, has provided a proof, or validated evidence by measures of muscular activity, that the motor processes assumed to be contained in the $T_{er}$ are indeed sensitive to SAT.

As already discussed above, the effect of stimulus contrast on $MT$ (approximately 7% of the effect of contrast on RT and 8% of the mean of $MT$) is somehow more surprising, but appears very reliable despite a small effect size. While the origin of this effect remains to be deciphered (as a "cognitive" or "energetic" process, see Sanders, 1983), one will retain here that stimulus contrast also modulates the two variables in the same direction. In contrast, replicating previous observations made in the context of conflict tasks (Allain, Carbonnell, Burle, Hasbroucq, & Vidal, 2004; Rochet, Spieser, Casini, Hasbroucq, & Burle, 2014; Śmigasiewicz, Ambrosi, Blaye, & Burle, 2020), correctness affected the two intervals in opposite directions: on errors, $PMT$ was shorter while $MT$ was longer, an effect that was enhanced by stimulus contrast. Comparing the two experiments, revealed that changing response force had a large effect on $MT$, but not on $PMT$ (if any, there would be a trend in the opposite direction). Finally, in Experiment 1 only, response side selectively affected $MT$, but spared $PMT$.

Two main conclusions emerge from this brief summary. First, the recorded $MT$ is not merely a constant factor, as it can be modulated by several experimental factors, including cognitive processes. Second, some factors selectively influence one of the two intervals, some affect both in a similar way, and others do so in an opposite way. Altogether, these results indicate that $PMT$ and $MT$ can be manipulated independently. In the literature, it has been repeatedly shown that certain experimental manipulations, such as stimulus-response compatibility, selectively affect $PMT$ and leave $MT$ unchanged (Burle, Possamaï, Vidal, Bonnet, & Hasbroucq, 2002; Hasbroucq, Possamaı, Bonnet, & Vidal, 1999; Spieser, Servant, Hasbroucq, & Burle, 2017). A mirror selective influence was recently observed: Dopaminergic-agonist treatment on Parkinson disease patients shortens $RT$ by selectively affecting $MT$, while leaving $PMT$ unchanged (Fluchère, Burle, Vidal, et al., 2018). Overall, these findings reveal a form of independence between the two intervals, and hence of the underlying processes they index. Stating such independence does not imply that $PMT$ and $MT$ are solely affected by disjoint groups of factors. As a matter of fact, contrast (*i.e.* stimulus quality) and SAT affect both $PMT$ and $MT$.

The independence between $PMT$ and $MT$ is further strengthened by the absence of correlation between the two intervals for most of the trials, as observed consistently in the two experiments. Crucially, however, this does not hold for the trials with very short $PMT$s, in both SAT conditions, where a negative correlation between the two intervals is observed. From a functional point of view, this variation in stochastic dependency suggests that, even within a condition, the temporal relationship between the underlying processes is variable, and hence that fast trials might not arise from the same generative model as slower ones. We will come back to this issue later.

### 2.7.2. Model-based decomposition and comparison with physiological fractionation

We first discuss the results from the model fit by themselves, irrespective of the information gained from EMG. As expected, contrast had a clear effect on the drift rate, and SAT on response threshold. Besides those canonical effects, SAT also affected the drift rate and $T_{er}$, confirming that SAT is not purely a threshold effect. Interestingly, the force manipulation between the two experiments was not captured by any difference in $T_{er}$, which, absent any $RT$ fractionation, would have lead to the conclusion that response force does not affect response execution, but rather upstream processes [10].

The herein observed presence of systematic effects on $MT$ puts into question the widely-practiced assumption that the $T_{er}$ parameter is insensitive to experimental manipulations which affect decisional processes, such as stimulus difficulty. Many modelers, out of convenience, implement the DDM with the assumption that $T_{er}$ does not vary across experimental conditions (*i.e.* this reduces the number of free parameters to estimate). The present results challenge this strategy. It is highly likely that when true effects are ignored by fixing non-decision time (e.g. Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998), the estimation of the other model parameters will be biased. While this is probably true for factors with moderate effect sizes such as SAT or response side on $MT$ in our experiment, it should be of lesser importance for factors with smaller effect sizes, such as contrast and correctness (although the two interact, leading to larger effect of correctness for the highest contrasts).

In our model selection procedure, $T_{er}$ was left free to vary as a function of SAT and response side for some models, but was always fixed for contrast and correctness. We note that current implementations of the DDM do not allow any parameter to vary between correct and error responses, because correctness is the product of decision related parameters. As a consequence, the model in its current state is not equipped to capture correctness effects on $MT$. In a similar vein, despite an effect of stimulus contrast on $MT$, $T_{er}$ was not allowed to vary as a function of contrast because i) this effect on $MT$ is small and the cost of 5 additional parameters was judged too high, and ii) contrast interacts with correctness and this interaction cannot be recovered. Hence, due to the model fitting method, we were limited to models with the non-decision parameter free to vary with SAT, response side and their interaction. We also tested for an effect of SAT on drift rate as well as the presence of a bias in the starting point. Although we do believe that our modeling strategy was the best possible, the empirical data indicates that it is not perfect.

After the initial model selection procedure suggested that boundaries, drift rate and non-decision time vary with the speed emphasis in our experiment, we showed that, the by-participant mean $MT$ and the by-participant $T_{er}$ estimated on behavior alone share a close relationship (Figure 2.6) in Experiment 1 but not in Experiment 2. At first sight, this could provide an argument against a the idea that $MT$ is indeed contained

---

[10]Unless one assumes that the effect in $T_r$ is perfectly compensated by an opposite effect on $T_e$ which, as argued in the introduction, is very unlikely.

in the $T_{er}$, which would be problematic for the model assumptions. However, a closer look leads to a more nuanced interpretation. Indeed, their actual covariation depends on how much of $T_{er}$ represents the motor processes. If one reasonably assumes that $T_e$ (encoding time) is approximately the same for the two experiments because the same stimuli are used, then the proportion of $T_r$ (execution time) into $T_{er}$ is lower in Experiment 2 (shorter $MT$ due to lower force) than in Experiment 1 (longer $MT$). The contribution of $T_e$ (that should not correlate with $MT$) is hence larger for Experiment 2 than Experiment 1.

It follows from these considerations that the pattern of substantial correlation for high force and much reduced for lower force is actually predicted by, or consistent with, the model. A potentially more problematic result is the observation that, despite a substantial reduction in the force needed to execute the response, captured by $MT$, as shown by the lower intercept for the LMM between Table 2.2 and Table 2.1, the mean non-decision time remains almost exactly the same in both experiments (see Table 2.3). This results implies that participants in Experiment 2, *i.e.* when the force needed to execute the response is lower, have a process inside the $PMT$ and captured by the $T_{er}$ parameter that is slower than it was for participants in Experiment 1. However in this study we cannot adjudicate between a genuine cognitive effect, a participant sampling bias, a consequence of the design modification between both experiments, or a misattribution of a motoric effect by the DDM. Hence, another study containing a within-participant manipulation of the force is needed to test these competing hypothesis. It is also important to note that the estimated SAT effect on $T_{er}$ was twice as big as the effect observed on $MT$. There are two possible explanations of this result: either the SAT effect could be contained in the remaining motor processes preceding the onset of EMG, or SAT could also affect early non-decision processes such as the $T_e$.

We do not have direct (physiological) measure of $T_e$. However, if we look at the value of $T_{er}$ and the mean values of $MT$ we see that subtracting $MT$ from $T_{er}$ would result in a residual time of approximately 200 ms, although the estimated value varies across experiments. This is fairly close to previous estimations of visual encoding times ($\simeq$ 200 msec) made by Roitman and Shadlen, 2002 using intra-cerebral recordings in monkeys or Nunez, Gosai, Vandekerckhove, and Srinivasan, 2019 using EEG in humans. Although this result has to be taken with caution, its consistency with previous estimates is noticeable.

We finally tested for a linear dependency between the EMG measured $MT$ and the $T_{er}$ by allowing a by-trial dependency between both measure in the estimation procedure. The corresponding fits show that the by-trial recorded $MT$ predicts the $T_{er}$. We do however observe that this linear dependency is variable across participants, with some participants presenting a low-to-null dependency. We linked this inter-individual variability in the dependency between $MT$ and $T_{er}$ with the inter-individual difference in correlation between $PMT$ and $MT$, showing that the more negatively the two intervals were correlated, the less there was a link between $MT$ and $T_{er}$. This is important as it shows that when we depart from the independence assumption of

the DDM, the $MT$ no longer shows a relationship with the modeling estimation of non-decision processes.

Taken together, our findings indicate a fair concordance between decision-making models (*e.g.* the DDM and its parameters) and EMG-derived chronometry measures. There are cases, however, where the model fitting completely missed, or misattributed, some of the empirically-observed effects. Although some coherent patterns can already be suspected, it is certainly too early to try to establish which types of effects can be appropriately recovered and which cannot. The present methodology, used systematically to probe other experimental manipulations may help to better delineate when the inferences based on model fitting are likely to be valid, and when one can doubt about them. In turn, this will certainly help to asses the strengths and weaknesses of the current models, and allow to improve them.

### 2.7.3. A single model with parametric modulations?

A last comment is in order. The core architecture of the DDM is intended to be generic, geared to account for data obtained in large variety of situations. These situations will induce changes in the parameters that capture different processes, always within the very same core architecture. The correlations observed between $PMT$ and $MT$, however, challenge this view. Indeed, while no correlation was observed between these two components in most of the trials, in agreement with the additive assumption (see Equation 1.4), a negative correlation was observed for short $PMT$s [11]. This has several consequences. These negatively correlated trials cannot easily be mapped into any parameter adjustment. Accordingly, participants' speed-up does not only reflect the modulation of one or another sub-process (speed of accumulation, response caution, etc.), but also partly reflects a change in the generative model of the decisions (see Dutilh, Wagenmakers, Visser, & van der Maas, 2011; Heitz, 2014, for discussions). The hypothesis of a different generative model is strengthened by the fact that the more negative the correlation between the two EMG intervals is, the less correspondence we observe between estimated non-decision time and motor time (Figure 2.7).

One approach that could accommodate these considerations might be a model in which decisions performed under SAT are generated from a mixture between fast-guesses and slow controlled decisions (an hypothesis formulated by Ollman, 1966, for a review see Heitz, 2014). The fact that trials identified as fast-guesses by the EWMA method (Vandekerckhove & Tuerlinckx, 2007) present a high negative correlation between $PMT$ and $MT$, while the other trials do not, supports this hypothesis. However additional analysis are needed to explore the nature of the generative model on these trials.

Even more than the difficulties the DDM had in accounting for some effects obtained on $MT$, such as the difference between correct and incorrect trials, a change in the architecture would strongly challenge the idea that DDM is a generic model that

---

[11]Importantly this is true even when considering only correct responses or computing correlations within each stimulus contrast condition

can account for the dynamics of information processing across response speeds. It must be noted, however, that despite this conceptual mismatch between the model's single process architecture and the presumed mixture of strategies the data provides evidence for, the model nonetheless efficiently fitted the data in the speed condition (see lower panels of Figure 2.4). This confirms that a good fit, while necessary, is not a sufficient argument for an agreement between the empirical observations and interpretative cognitive validity.

## 2.8. Conclusion

Through two experiments, we reliably showed that response execution time measured with EMG is far from being a ballistic movement subsequent to the decision. This conclusion is in contrast to what is assumed in major decision-making models and their fitting applications. Rather, response execution time is sensitive to the Speed-Accuracy Trade-off, stimulus strength, and correctness of the response. Importantly, EMG recording allows for the measurement of effect sizes, in turn allowing researchers to assess the impact of ignoring such variations. Additionally, we showed that a subsidiary assumption of most models of $RT$, the independence between decision and non-decision time, while potentially valid in most of the trials, is violated on the early quantiles (*i.e.* the fastest decision) of the $PMT$ distribution.

Despite such discrepancies between our observations and model assumptions, we show that although relying on only behavioral measures to estimate the non-decisional latencies, the DDM non-decision time parameter provides, under some conditions, an accurate representation of the EMG-observed response execution time. This observation is true both between participants, when the amount of motor processes in the $RT$ is fairly high, and within participants as is shown with the linear dependency between the by-trial EMG $MT$ latency and the estimated $T_{er}$. However, the by-trial negative correlation between pre-motor and motor time on faster trials, points to possible trade-offs between latency components that are dependent on task requirements. This could imply that parameter interpretability (or attribution) can change according to the validity of the independence assumption.

# 3. Bias and Motor processes in decision making

## Sommaire

## 3.1. Note

The following chapter is a manuscript under preparation with Gabriel Weindel, F.-Xavier Alario and Boris Burle as authors.

## 3.2. A primer on response execution and decision bias

In Experiment 1 of the previous chapter, fitting a DDM suggested that participants were biased to respond with their right hand (see the estimation of the starting point in Table 2.3) while no such bias was estimated to be present in the second experiment. We also observed, specifically in Experiment 1, a left/right difference in the motor time ($MT$) derived by EMG while no such difference was observed on the pre-motor time ($PMT$). This is surprising because we would expect that if DDM and EMG did decompose $RT$s in the same way (i.e. the decision time is contained prior to response initiation) a bias in the decision process should be associated with a difference in the $PMT$ rather than the $MT$. This mismatch could show that a standard DDM confuses a response execution bias as a decision bias.

A. Voss, Voss, and Klauer, 2010 provide evidence in support for this hypothesis. Using RT data simulated with a DDM, they show that if a difference in response execution ($T_R$) is induced in the generated data, the fit of a regular DDM on the data wrongly suggest a bias in the starting point (see Figure 3a in A. Voss, Voss, & Klauer, 2010). To fix this mis-estimation, the authors suggest to add a parameter to the DDM that is intended to capture the difference in the response execution times ($T_{er}$) referred to as the $d$ parameter. In the settings of Chapter 2 we have an empirical situation where such confusion might take place. We therefore seek to confirm that no pre-motor bias is present by fitting a DDM on the $PMT$ and test whether DDM and EMG decompositions align when using the parametrization suggested by A. Voss, Voss, and Klauer, 2010.

## 3.2.1. Methods

### 3.2.1.1. Data

We used the data from Experiment 1 of Chapter 2.

### 3.2.1.2. DDM fit

A full DDM (Ratcliff & Tuerlinckx, 2002) was fitted to the *RT*s. Contrary to previous studies in the thesis we used a frequentist estimation of the DDM parameters using the *fast-DM* program (A. Voss & Voss, 2007). Bayesian methods, especially hierarchical Bayesian methods, are usually considered a better option, in particular when estimating the full DDM (see Boehm, Annis, Frank, et al., 2018). However, our goal to use the implementation of the $d$ parameter described above imposed the use of *fast-dm 30* (A. Voss, Voss, & Lerche, 2015). For the model fit, we chose to use the Kolmogorov-Smirnov (KS) optimization criterion which has been shown to provide reliable recovery even with fast response contaminants (Lerche, Voss, & Nagler, 2017).

Overall, we fitted the same model as the selected model in Experiment 1 of Chapter 2 (see Table A.6 in Appendix A) with 29 free parameters[1] independently estimated for each participant.

For practical numerical purposes the lower threshold is equal to 0 and associated with a left response, the upper threshold is equal to the value of the estimated boundary parameter and associated with a right response. The starting point was divided by the value of the boundary parameter to indicate the relative distance between boundaries. With these features the *relative* starting point is contained in a range between 0 and 1. A value of 0.5 indicates an unbiased starting point while a value > 0.5 indicates a bias to the right response and a value < 0.5 a bias to the left response.

### 3.2.1.3. Statistical test of the starting point parameter

After estimating the DDM starting point parameter, we used the two-sided one sample Bayesian $t$-test implemented in *JASP* (JASP Team, 2020). We chose as null hypothesis $H0$ that the mean of the distribution of the starting point parameter lies between the boundary parameters hence at 0.5. The alternative hypothesis $H1$ was, then, that the mean of the population distribution of the starting point is biased towards one of the boundaries. The Bayesian $t$-test estimates a standardized effect size $\delta$ (the ratio between mean and standard deviation) for which we used the default JASP priors, namely a Cauchy prior with a scale of 0.707. We used the Bayes factor (BF) to assess the evidence towards $H0 : \delta = 0$ ($BF_{H0}$) and towards the alternative hypothesis $H1 : \delta \neq 0$ ($BF_{H1}$). To ensure for prior sensitivity of the BFs we performed a prior sensitivity analysis for each $t$-test (as implemented in JASP) and we report any significant difference

---

[1]In Chapter 2 there were only 21 parameters because 1) we estimated population inter-trial variability parameters and 2) we used the same drift for left and right responses see Footnote 9 there. We could not use these fitting strategies with *fast-dm*

among the default priors used. We used the convention that a BF located between $[0,1[, [1,3[, [3,10[$, or above 10 represents respectively no evidence, anecdotal, moderate, and strong evidence in favor of the considered hypothesis (Lee & Wagenmakers, 2013). For each statistical test we report the median of $\delta$ ($\bar{\delta}$)and the associated 95% credible interval (CrI, see method section of Chapter 2).

## 3.2.2. Results

Table 3.1 provides the descriptive statistics of starting point estimates, along with the associated BFs. Figure 3.1a provides a graphical summary of these estimates and their distributions.
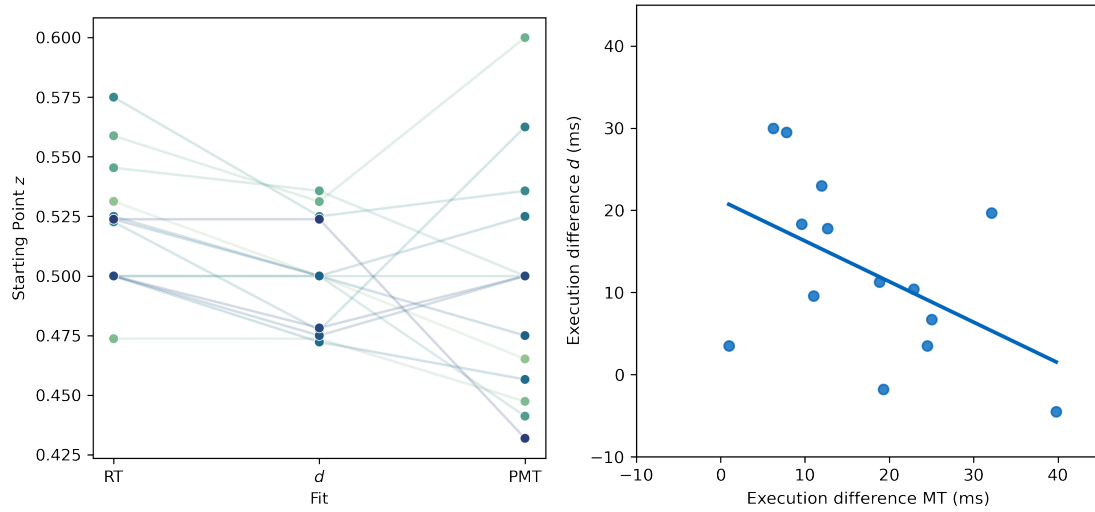


Figure 3.1.: a) Starting points estimated from $RT$, $PMT$ and $RT$ using the $d$ parametrization of A. Voss, Voss, and Klauer, 2010 for the 14 color-coded participants. b) Scatterplot and associated regression line of the estimation of response execution difference made with the model parameter $d$ and the measurement of the difference using $MT$.

From the estimates of the starting point we see that, as in Chapter 2, a fit on $RT$ suggests a bias towards the boundary associated with the right response ($\bar{\delta} = 0.63, CrI = [0.10, 1.23]$, $BF_{H1} = 3.84$). Fitting on $PMT$ removed the motor processes measured by EMG and suggested an unbiased starting point ($\bar{\delta} = -0.07, CrI = [-0.56, 0.39]$, $BF_{H0} = 3.53$). When we use the parametrization of the DDM including a response execution difference parameter $d$ along with a fit on $RT$ we also find an unbiased starting point ($\bar{\delta} = -0.02, CrI = [-0.49, 0.45]$, $BF_{H0} = 3.69$).

The $d$ parameter is thought to capture differences in response execution, therefore its value should be related to differences in response execution we measure with $MT$. At the group level, the mean difference in $MT$ between left and right hand is 17

| Model | Mean | SD | $BF_{H0}$ | $BF_{H1}$ |
|---|---|---|---|---|
| RT | 0.520 | 0.027 | 0.26 | 3.84 |
| PMT | 0.496 | 0.048 | 3.53 | 0.28 |
| $d$ | 0.499 | 0.022 | 3.69 | 0.27 |

Table 3.1.: Means and standard deviations for the starting point estimated from $RT$ and $PMT$ distributions using the standard full DDM, and from $RT$ using the $d$ parametrization of A. Voss, Voss, and Klauer, 2010. The two last columns provide the values of the BFs for the null (unbiased starting point) and the alternative hypothesis (biased starting point).

ms (SD = 10 ms) close to the estimation made by the models (mean = 13 md, SD = 10 ms). To test this hypothesis, we computed the $MT$ left-right difference for each participant yielding a "measured execution difference". Using a Bayesian correlation from *JASP* (JASP Team, 2020), we then computed a Pearson correlation between the estimated difference $d$ and this measured execution difference. This analysis shows a negative correlation, but with anecdotal evidence towards the alternative hypothesis ($r(14) = -0.49$, $BF_{H1} = 1.40$, see Figure 3.1b).

### 3.2.3. Discussion

Our re-assessment of Experiment 1 from Chapter 2 confirms that the decision bias inferred there from the starting point parameter of the DDM is linked to a difference in response execution. When the data are modelled either by removing the motor processes or by adding a parameter thought to account for response execution differences, no bias is found on the starting point. This is in line with the simulation results obtained by A. Voss, Voss, and Klauer, 2010, but now based on real data with the response execution measurements allowed by EMG. As argued by A. Voss, Voss, and Klauer, 2010, in applications where a response execution time difference is present, modellers might wrongly assume that participants have a decision bias. EMG of the responding muscles provides an efficient way to detect such execution differences.

It is noteworthy that the $d$ parameter and the measured response execution difference are not related in our data set; only an "anecdotal", negative if anything, correlation was observed. This might be because the amount of trials was not high enough to model the $d$ parameter accurately; against this view, the ~1000 trials available by response side and participant are close to the recommendations made in A. Voss, Voss, and Klauer, 2010. Alternatively, it is possible that DDM allows for appropriate inferences on the cognitive process when the correct parametrization is used (in this case, detecting a response execution bias),but that the individual estimates of cognitive process parameters are unreliable for assessing inter-individual differences. Lastly, the weak number of participants might also contribute to the lack of correlation

between DDM and EMG derived response execution difference.

## 3.3. Decision bias and response execution

The previous section shows that a response execution bias is wrongly interpreted as a decisional bias if a decision model that does not account for response execution differences is fitted. In contrast, not much is known about the effect of a genuine decision bias on response execution. If a decision bias induces a response execution bias, the DDM might not be able to separate these effects without the appropriate modeling strategy advocated by A. Voss, Voss, and Klauer, 2010 simulations. This is problematic as the interpretative validity of the starting point parameter of the DDM amongst other models has been established by selective influence tests where for example a manipulation of the response probabilities was assumed to translate selectively into a decision bias through a change in the starting point of the evidence accumulation (Leite & Ratcliff, 2011; Mulder, Wagenmakers, Ratcliff, Boekel, & Forstmann, 2012; Ratcliff & McKoon, 2008; A. Voss, Rothermund, & Voss, 2004). However if inducing a bias by *e.g.*, increasing the proportion of stimulus associated with one response, translates mainly in the motor processes, this would seriously question the interpretation of the starting point parameter of the DDM as reflecting bias in decisions.

Starns and Ma, 2018 have tested whether the non-decision components such as the motor times contributes to the effect of bias usually observed on RT. They first used a task designed to minimize decision processes time (responding to an arrow pointing left or right) and observed that high probability stimuli yielded faster RTs than low probability ones. Given the assumption that the task contained minimal decision related latencies, they attributed the observed effect mainly to motor non-decision processes. Note that this conclusion was reached without any model fitting. In another experiment, Starns and Ma, 2018 compared two canonical decision making tasks with a biased probability towards one alternative. In one of the tasks, the stimulus mapping was known in advance while in the other it was displayed 250 ms after stimulus onset. The bias effect on RT was found to be higher in the known mapping than the unknown mapping task. They attributed this increased effect on RT to the participation of motor execution bias when the response could be prepared in advance. The authors concluded that motor non-decisional processes significantly contribute to the bias effect observed on RT. They advised to take this effect into account in the modelling efforts of decision making. While the rationale of these experiments is very innovative, first, the interpretations rest on a non-trivial assumption, even in a simple arrow discrimination task, we do not know the relative variance explained by decision and non-decision processes, and the authors had no means to assess them. Second, some alternative interpretation can be made towards the higher bias in known *vs.*unknown mapping. For example one can easily imagine that fast guesses increase the effect of bias (more fast guesses when a cue on the probability of an upcoming stimlulus is given) but such fast guesses cannot occur when the stimulus-response mapping

only appears at 250 ms. Hence while Starns and Ma, 2018 shed light on a potential effect of bias on motor processes, a careful scrutiny of their study highlights how challenging it is to attribute an experimental effect to these latent motor processes based on behavior alone.

Physiology can provide a more direct access to the question as whether motor activity is impacted by a decisional bias. Calderon, Van Opstal, Peigneux, Verguts, and Gevers, 2018 tested the hypothesis that a prior cue on an upcoming motor plan would increase activity in the motor cortex M1 associated with the cued response side prior to the actual movement. To test this hypothesis the authors measured brain activity using fMRI in the context of a go/no-go task. On each trial the cue could indicate a trial with equal-probability of left or right response, the side of the response, or a no-go trial. The activity in the M1 cortex contralateral to the cued response in the go trials increased before the actual movement was required compared to the ipsilateral M1 cortex. In their study however the cue was either non-predictive (equal probability) or fully predictive of the response side. Hence, while the authors show that prior information about an upcoming motor response is integrated up to the motor cortex, we don't know whether this results generalizes to the case where the uncertainty in the upcoming response is merely biased towards one alternative. Additionally this study does not show wether this increased activity translates the actual response execution.

Weinberg, 2017 also suggested that prior information about an upcoming choice could bias motor processes. In a high powered series of three experiments, Weinberg used motor evoked potentials measured through transcranial magnetic stimulation as an index of cortico-spinal excitability. The choice in a 2AFC random dot kinematogram was biased using a cue with 5 levels ranging from 10% to 90% of right responses associated with the cue. The author hypothesized that cortico-spinal excitability of the cortex involved in the most likely response, as indicated by the cue, should increase at stimulus onset. However, the data did not provide evidence in favor of a bias in the motor processes as measured with muscular evoked potential (see also Hasbroucq, Osman, Possamaı, et al., 1999; Meckler, Allain, Carbonnell, et al., 2010). According to the author, either biasing a choice is not translated in motor areas at stimulus onset, or muscular evoked potentials are too variable to show the potentially small effect on motor processes.

### 3.3.1. The current study

Based on these conflicting results we do not know whether response execution in a 2AFC task is influenced by a manipulation of the decision bias. As shown in the previous section and by A. Voss, Voss, and Klauer, 2010, if a decision bias is associated with a response execution bias, the conclusions drawn from models such as the DDM (without a specific parametrization such as the one suggested by A. Voss, Voss, & Klauer, 2010) will be misleading on the way participants bias their decisions. The degree of mismatch between model and EMG decomposition will depend on the actual size of the effect of a decision bias on the response execution, suggested to be

high by Starns and Ma, 2018, compared to the size of the effect on decisional processes.

As in chapter 2, we can use the motor time ($MT$) derived from EMG to measure the actual effect on the durations of motor execution of an experimental factor such as a manipulation of response probabilities used by Starns and Ma, 2018. In addition to $MT$, we can also explore different properties of the EMG signal. For example, when assessing the effect of motor cueing, Possamaı, Burle, Osman, and Hasbroucq, 2002 showed that providing a fully predictive cue on the upcoming motor plan affect the shape of the response EMG burst. Such cues increase the steepness of the EMG burst leading to the response. They interpreted this effect as a more efficient motor command by a better synchronization of the motor units discharges. If an uncertain response probability manipulations have an effect on the motor processes, it could also be revealed by a more efficient motor command, hence a steeper EMG slope when the response was cued vs. not cued. Response probability could also affect other aspects of motor execution. For example, knowing that the left response is more likely than the right one could lead to a slight tension increase in the left muscle in order to speed-up the upcoming response. Such a subtle pre-activation might not be visible at the single trial level, but could be evidenced after averaging. To do so the EMG signal will be rectified (taking the absolute value of the signal) and averaged. If a small tonic activity is present, it will sum-up and differences in the averaged values between condition will inform about the amount of tonic activity.

We designed a task including a manipulation of the response probabilities with a cue as used by Starns and Ma, 2018. Along with this manipulation we also manipulated SAT as we hypothesized that different SAT instructions will change the reliance on the cue and hence the probable impact of the cue on the motor system.

## 3.3.2. Methods

### 3.3.2.1. Participants

Sixteen participants (4 men and 12 women, mean age = 23 years, 2 left-handed), all students at Aix-Marseille University, were recruited for this study. All participants reported having normal or corrected vision, and no neurological disorders. The experiment was approved by the ethical experimental committee of Aix-Marseille University, and by the "Comité de Protection des Personnes Sud Méditerrannée 1" (Approval n° 1041). Participants gave their informed written consent, according to the declaration of Helsinki. They were compensated at a rate of €15 per hour.

### 3.3.2.2. Apparatus

Participants performed the experiment in a dark and sound-shielded Faraday cage. They were seated in a comfortable chair about 100 cm away from a 15 inch CRT monitor that had a refresh rate of 75 Hz. The CRT monitor was gamma corrected by a psychophysical procedure provided by the software PsychoPy (Peirce, 2007). Responses were given by pressing either a left or a right button with the corresponding

thumb. The buttons were fixed on top of two cylinders (3 cm in diameter, 7.5 cm in height) that were fixed on a tablet and separated by a distance of 20 cm. The buttons were mounted on force sensors that recorded a continuous measure of the force produced at a sampling rate of 2048 Hz. The behavioral response was recorded when a force threshold was exceeded (1.8 N, 180 g, which is close to the force used in Experiment 2 of Chapter 2). Response triggers were transmitted to the parallel port of the recording computer. At button press, participants received a 3 ms sound feedback (1000 Hz pure tone).

The participants' forearms and hypothenar muscles rested comfortably on the table, to minimize muscle recruitment during response execution. We measured the EMG activity of the flexor pollicis brevis of both hands with two electrodes placed 2 cm apart on the thenar eminences. This activity was recorded using a BioSemi Active II system (BioSemi Instrumentation, Amsterdam, the Netherlands) at a sampling rate of 2048 Hz.

### 3.3.2.3. Stimuli

Stimulus presentation was controlled by the software PsychoPy (Peirce, 2007). The stimuli were composed of one central vertically oriented Gabor patch of 2.5 visual angle degrees with a spatial frequency of 1.2 cycles per degree. At each trial, the Gabor patch was randomly tilted to the right or to the left. The tilting angle was defined by the performance of the participant on a prior task described below. The participants' task was to indicate whether the Gabor patch was tilted to the right or the left with the corresponding button press.

### 3.3.2.4. Experimental manipulations

**Deviation angle** In order to adjust task difficulty to each participant, a calibration task was performed prior to the main task. The calibration was composed of 240 trials (approximately 6 minutes on task) with 6 tilting angles of the Gabor patch to the vertical line (0.5°, 0.75°, 1°, 1.25°, 1.5°, 1.75°) selected based on a pilot study. The degrees were randomly subtracted or added (calling left and right responses respectively) to the vertically oriented Gabor patch. The degree of deviation was randomly chosen at each trial.

Once the participant completed the 240 trials a low parameter version of the DDM, the PRDM proposed by Palmer, Huk, and Shadlen, 2005a, was fitted on the participants RT and choices.

We first excluded RTs slower than 2500 ms and faster than 100 ms. We then computed the mean RT and proportion of correct responses for the participant for each absolute value of deviation with left- and right-oriented stimuli combined.

PRDM assumes that the proportion of correct responses ($P_c$) for each absolute deviation $d$ is

$$P_c(d) = 1 - \frac{1}{1 + e^{-2apd}} \tag{3.1}$$

where $a$ represents the boundaries and $p$ the coefficient that multiplied by stimulus intensity $d$ provides the drift rate of the diffusion model. The mean RT is assumed to follow the function

$$\overline{RT}(d) = \frac{a}{pd} \tanh(apd) + T_{er} \tag{3.2}$$

Where $T_{er}$ is the non-decision time of the diffusion model. $a$, $p$ and $T_{er}$ are the three free parameters of the model fitted on the mean RT and proportion correct of the 6 levels of deviation.

We then used the "minimize" function of the *scipy* python module (Oliphant, 2007) to run a Nelder-Mead optimization. The maximized function is the combined log-likelihoods of the mean RT and correct response proportion ($P_c$) given by Palmer, Huk, and Shadlen, 2005b summed over the 6 deviation degrees :

$$\log_n(L) = \sum_d \log_n[L_{RT}(d)] + \log_n[L_{Pc}(d)], \tag{3.3}$$

where $L_{RT}$ and $L_{Pc}$ are respectively the likelihoods of the mean RT and the proportion correct given by the Gaussian and binomial probability density functions

$$L_{RT}(d) = \frac{1}{\sigma_{\overline{RT}}\sqrt{2\pi}} e^{-[\widehat{RT}(d)-\overline{RT}(d)]^2/(2\sigma_{\overline{RT}}^2)} \tag{3.4}$$

$$L_{Pc}(d) = \frac{n!}{r!(n-r)!} \widehat{P_c}(d)^r (1-\widehat{P_c}(d))^{n-r} \tag{3.5}$$

where $\widehat{RT}(d)$ and $\widehat{P_c}(d)$ are respectively the predicted mean response time and proportion of correct responses for deviation $d$ given by Equation 3.2 and Equation 3.1. The term $\sigma_{\overline{RT}}$ is the predicted standard error of the mean (see Appendix of Palmer, Huk, & Shadlen, 2005a). The participant was presented the main task after selection of the level of deviation that corresponded to a predicted accuracy of 90% by the PRDM (See Appendix B.1).

### 3.3.2.5. Procedure

All participants performed a single experimental session with 16 blocks of 120 trials each. Session duration was approximately 1h30 including a training session of 15 minutes and self-paced breaks between each block. Participants were asked to keep their gaze on the central fixation cross throughout each block, and to respond to the visual stimulus according to the corresponding SAT instruction (SAT instructions where implemented as in Chapter 2, see Section 2.3.1.4). The probability cues displayed before the stimulus indicated the probability of the right *vs.*left orientation. Three levels where chosen, right more likely, even and right less likely associated respectively with a probability of 0.70, 0.50 and 0.30 of an upcoming right-oriented stimulus. These cues were displayed as two histograms side-by-side whose height represented the probability of the ipsilateral orientation (see bottom screen in Figure 3.2).

The training session started with 3 blocks of 10 trials without specific SAT instructions but with the presentation of the probability cue before each trial. At the end of each block, the participant was given feedback about their accuracy and speed. These 3 blocks were followed by 2 blocks of 10 trials in the Speed condition, followed by 2 blocks in the Accuracy condition, and finally 2 blocks of 10 trials with alternating instructions. During the experimental session, SAT instructions alternated every three consecutive blocks, with the order of the SAT instructions counterbalanced across participants. Within each block, the cue and the orientation of the stimulus were pseudo-randomly presented. All cues were presented with equal frequency and the upcoming stimulus was determined by the probability associated with the cue.

Each trial started with the appearance of the cue displayed during 250ms, replaced by a fixation cross during 500 ms and itself replaced by the stimulus until a response was recorded. Each trial ended with a Gabor patch with a spatial frequency of 0 cycles per degree (white) during 100 ms after the response to avoid sequential effects followed by a blank screen displayed during 400 ms (see Figure 3.2).
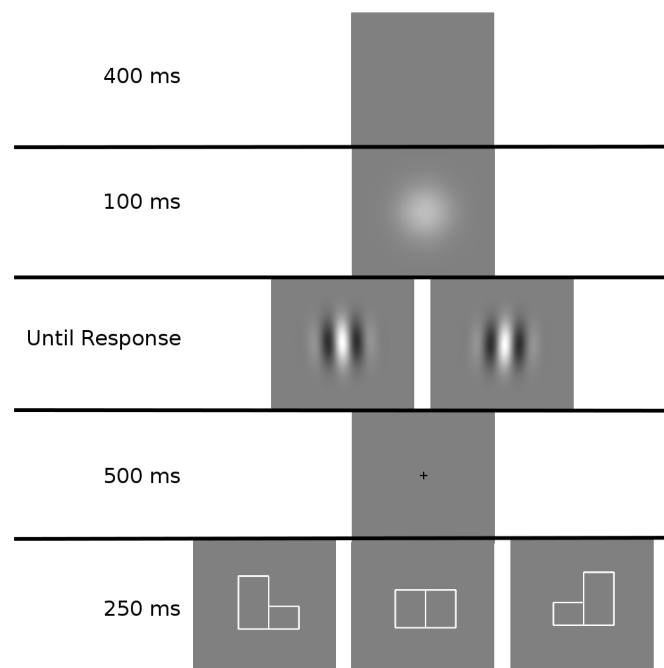


Figure 3.2.: Representation of the 5 events defining the time course of a trial (start of the trial bottom, end top) and their durations. The three screens on the bottom line represent the cue seen by the participant at trial start, each of these screen were equally likely to appear at each trial. These cues announced an upcoming stimulus (Gabor patch tilted to the left or to the right as presented on the middle line) with a probability of being tilted to the right of, respectively from the left to the right, 0.30, 0.50 and 0.70.

### 3.3.2.6. Fitting method

A full DDM (Ratcliff & McKoon, 2008) was fitted to the $RT$ distributions using the *fast-DM* program along with the the Kolmogorov-Smirnov (KS) optimization criterion (A. Voss & Voss, 2007).

We designed a model with 16 free parameters, independently estimated for each participant. The starting point parameter was free to vary with the response cue and with SAT, as we expect that reliance on the cue will depend on the speed stress[2]. Boundary, non-decision time $T_{er}$ and inter-trial variability of the starting point parameters were free to vary with SAT instructions as the previous chapter suggests that these parameters are highly influenced by this factor. Drift rate was free to vary with stimulus orientation only. Additionally, we estimated inter-trial variability parameters of the drift rate and the non-decision time.

As in the previous section (Section 3.2.1.2) the starting point ranges from 0 to 1 and indicates the relative distance between boundaries (with lower and upper threshold associated with left and right responses respectively) with a value of 0.5 representing an unbiased decision.

### 3.3.2.7. EMG Analysis

The methods used to record the EMG, to detect the onset of the response EMG, and to decompose each $RT$ into $PMT$ and $MT$ at each trial is the same that was used in Chapter 2 (Section 2.3.1.5). For this experiment, we removed trials that presented multiple EMG activities because, as pointed out in the previous chapter, the presence of multiple activities might change the activation dynamics of the cortical and spinal neurons.

In order to extract the baseline of the EMG prior to the processing of the stimulus by the participant, we defined a window between 850 ms before and 50 ms after the stimulus onset, hence including the EMG activity at the presentation of the cue occurring 750ms before the stimulus onset. As EMG bursts are extremely high compared to the baseline (see examples in Figure A.1), we excluded all trials that presented an EMG burst during the defined window as those trials would strongly bias the estimation of baseline activity. For that purpose, we first applied the same detection algorithm based on the integrated profile of the EMG signal as described in section 2.3.1.5. An EMG burst was detected if the EMG signal was above 3 SD of the baseline, computed on the first 100ms of the window, during 25 recorded consecutive signal samples. The resulting detection was then inspected and corrected in case EMG burst remained present after the variance based automatic detection (*e.g.* this detection method will usually fail to detect a subsequent burst if one is present during the computation of the baseline). The proportion of trials presenting an EMG burst was 0.24 [3]. After this

---

[2]Note that we estimate the relative starting point, namely that we estimate the distance of the starting point for the normalized boundaries. A starting point of 0.5 means that the decision is unbiased

[3]This is probably an overestimation of the number of trials with an EMG burst before stimulus onset but given their impact on the estimation of the baseline we preferred to use a low detection threshold

trial rejection procedure, we performed a running mean computation on the absolute value of the 10Hz high pass filtered EMG signal separately for both hands starting 850 ms before stimulus onset and ending 50ms after it. We chose to compute the running mean on a running window of 102 samples (approximately 50 ms with a sampling rate of 2048Hz). These running means were then used to plot the average evolution of EMG baseline with the time course of the trial for each condition and the last 102 points of the running means before stimulus onset for each trial was used with a LMM to search for differences between conditions (see below for a description of the LMM).

To compute the slope of the EMG burst leading to the response, as for the computation of the baseline EMG, we high-pass filtered the data at 10Hz then computed the rectified EMG signal by taking the absolute value of the signal. To study the effect of validity and SAT on the slope of the EMG while avoiding a potential confounding factor of error responses we only computed the slope of the EMG bursts that lead to a correct response (see Rochet, Spieser, Casini, Hasbroucq, & Burle, 2014, for an analysis of the EMG slopes difference between correct and error responses). To compute the slope of the rectified EMG signal of the responding hand, we computed a cumulative sum of the first 61 samples following the EMG onset (approximately 30 ms with a sampling rate of 2048Hz as done by Rochet, Spieser, Casini, Hasbroucq, & Burle, 2014). We then computed the slope of the EMG between the cumulative sum of the EMG signal in $mV$ and the time in milliseconds (from 0 to approximately 30ms after stimulus onset) by computing an ordinary least square regression using the *scipy* python package (Oliphant, 2007). The by-trial slope of the EMG was then analyzed with a LMM described below.

### 3.3.2.8. Statistical analysis

All linear mixed models (LMMs) applied to the data are based on the method that was described in Chapter 2 (see Section 2.3.1.6). For each LMM we report the mean of the posterior distribution of the population parameters (noted $\overline{\alpha}$ and $\overline{\beta}$ for intercepts and slopes, respectively) along with the 95 % Credible Interval (CrI). As in Chapter 2, we consider an effect to be significant when the CrI do not include 0. While this strategy is not necessarily optimal for a binary decision regarding the absence or presence of an effect, it does put the focus on the estimation of the effects sizes which is of particular importance in the present application.

For all analysis (except the starting point and baseline analysis, see below) the factor of the right-oriented probability cue was re-coded as valid (orientation of the stimulus of the trial corresponds to the cued side), even and invalid (orientation contralateral to the cued side) and used as a predictor along with the SAT and the interaction between both factors.

The proportion of correct responses were analyzed with generalized linear mixed models (GLMM) using binomial link functions. The LMMs on RT, PMT and MT we first log-transformed the data as in Chapter 2 and included the correctness of the

___

to favor sensitivity over specificity.

response (correct *vs.*errors) and its interactions with the other factors as predictors. The estimates of these GLMM and LMMs were back-transformed to the original scale (proportion correct and ms, respectively) as in Chapter 2 (Section 2.3.1.6).

The distribution of EMG slopes presented an important right skew and were therefore first log transformed. The EMG baseline was modelled as a Gaussian distributed variable and we used the right-oriented probability cue, the SAT instructions, the EMG channel (left or right) and their interactions as predictors. Using the right-oriented probability cue along with the EMG channel side we expect that these factor should interact as, e.g. an increase in the probability of a right-oriented stimulus should increase the activity in the right channel only and conversely an increase in the probability of the left-oriented response increase baseline in the left channel.

For the starting point parameter, out of simplicity and for comparison with the previous section of the chapter, we used the point estimate given by the *fast-DM* program and applied a LMM, assuming a Gaussian distribution. SAT, right-oriented probability cue and their interaction were included as predictors[4]. Although the number of observations is low compared to the single trial analysis (one starting point parameter for each of the 16 participants in the 6 experimental cells), we still estimate the full random effect structure as advocated by Barr, 2013 because we rely on Bayesian estimation of the parameters of the LMM.

### 3.3.2.9. Factor Coding

For all regression analysis the SAT factor was coded as a treatment factor (0 for speed and 1 for accuracy) because we expected that participants will tend to use the cue especially when speed stress is high. The validity factor was coded as a sum contrast with invalid, even and valid located at -0.5, 0, 0.5, respectively[5]. The correctness factor was coded as 1 for errors and 0 for correct responses. The right-oriented probability cue factor with the three levels, 0.3, 0.5 and 0.7 were coded as respectively -0.5, 0 and 0.5. The EMG channel for the EMG baseline LMM was coded as 0 for left and 1 for right channels. See 3.6 for a summary of the contrasts used in the study and Section 2.3.1.6 for a guide on how to interpret linear model coefficients with different contrasts.

$$\text{SAT} = \begin{pmatrix} \text{Speed} & \text{Accuracy} \\ 0 & 1 \end{pmatrix}$$

$$\text{Validity} = \begin{pmatrix} \text{Invalid} & \text{even} & \text{valid} \\ -0.5 & 0 & 0.5 \end{pmatrix}$$

$$\text{Correctness} = \begin{pmatrix} \text{Correct} & \text{Incorrect} \\ 0 & 1 \end{pmatrix} \tag{3.6}$$

---

[4]While the starting point is bounded between 0 and 1, its estimated value usually lies far from these boundaries. We therefore treated the starting point as an unbounded continuous variable

[5]Using this coding scheme imposes a constrain of linearity between the three levels, a later inspection of the fits compared to raw data suggests that this assumption is reasonable except for the result on proportion correct, see Figure 3.3

$$p_{Right} \text{ Cue} = \begin{pmatrix} 0.3 & 0.5 & 0.7 \\ -0.5 & 0 & 0.5 \end{pmatrix}$$

$$\text{EMG channel} = \begin{pmatrix} \text{Left} & \text{Right} \\ 0 & 1 \end{pmatrix}$$

### 3.3.2.10. Priors for the regression analysis

The priors used for the GLMMs on proportion correct, the RT, PMT and MT are described in Appendix A.2. The priors of the LMMs on the EMG parameters and starting point of the DDM are given in Appendix B.2; they were determined based on the visual inspection of distributions of the data, their adequacy was evaluated by performing a prior predictive check (see Schad, Betancourt, & Vasishth, in press, for an example of prior predictive checks applied to reading times and Appendix B.3 for an example applied on the baseline EMG).

## 3.3.3. Modeling results on the behavioral and EMG chronometric variables

In addition to the trials with multiple EMG activities (28.6 % of the data) and trials where EMG onset detection was compromised by tonic EMG artifacts (1.5 % of data), we removed trials with RTs higher than 2000 ms for the behavioral analysis (2% of the data). During the experimental sessions, three participants were interrupted because of a recording failure but completed more than half of the trials of the task. We chose to keep these participants. After all exclusions there were on average 210 trials (min = 104, SD = 50) for each experimental cell : participant × SAT × Cue.

As in the previous chapter, the results of the G/LMMs are presented congruently to the factor coding design chosen (see Section 3.3.2.9). Therefore, e.g. the effect of validity on *RT* is the difference between invalid and valid cues when SAT instructions are to emphasize speed and the response was correct. We report the interaction terms to show how the other levels of the SAT and correctness factors change the pattern of effects.

### 3.3.3.1. GLMMs on proportion correct and RT

Figure 3.3 represents the effect of SAT and cue validity on the proportion correct. Participants made more correct responses when asked to favor Accuracy ($\overline{\beta}_{SAT} = 0.13$, $CrI = [0.08, 0.17]$). When speed was emphasized, a valid cue tended to be associated with a higher number of correct responses although the CrI loosely included 0 ($\overline{\beta}_{Validity} = 0.08$, $CrI = [-0.02, 0.15]$)[6]. The interaction with SAT shows that the effect of validity is centered on 0 when participants are asked to favor accuracy ($\overline{\beta}_{SAT \times Validity} = -0.08$, $CrI = [-0.21, 0.04]$).

---

[6]Note that the effect of validity is probably under-estimated, see legend of Figure 3.3
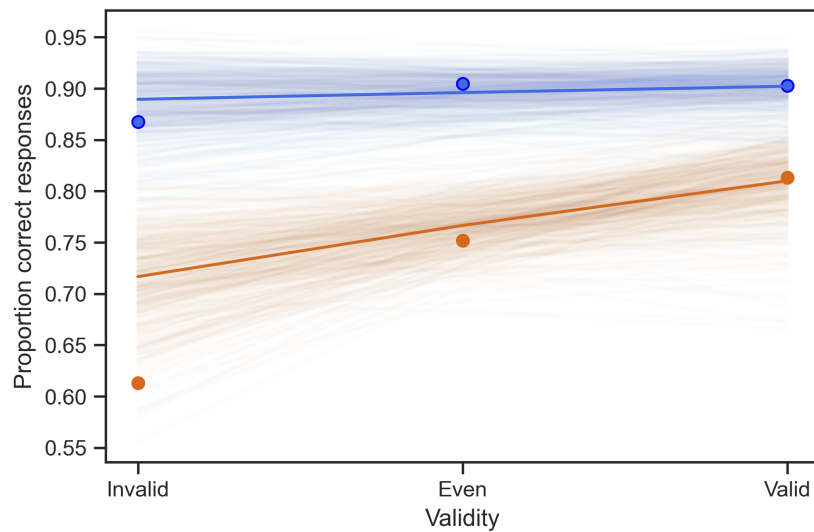
Figure 3.3.: Group averaged mean values of proportion of correct responses (dots) plotted for each SAT condition and against validity levels. The lines represent 1000 random draws from the joint posteriors of the combined MCMC chains of the GLMM fit. The thick line represent the predicted regression line with all parameters set at the mean of their respective marginal posterior distribution. Note that we observe a clear non-linear pattern across validity levels in both SAT conditions resulting in a misfit of the GLMM (e.g. difference between the data and the lines of the GLMM in the invalid condition especially). This misfit likely translates in a underestimation of the difference between invalid and even cue but given that the results on proportion correct are not the focus of the present study, we maintain the same modelling strategy as for the rest of the variables.

For the RT, coherent with the effect on proportion of correct responses, asking participants to favor accuracy did slow the RTs ($\overline{\beta}_{SAT} = 195$, $CrI = [137, 255]$). SAT and validity interacted: a valid cue shortened RTs in the "speed" instructions condition ($\overline{\beta}_{Validity} = -57$, $CrI = [-82, -31]$) and to a much lesser extent in the "accuracy" condition, as indicated by the interaction term ($\overline{\beta}_{SAT \times Validity} = 52$, $CrI = [18, 89]$). SAT and error responses also interacted: errors were associated with faster RTs ($\overline{\beta}_{Correctness} = -58$, $CrI = [-70, -45]$) in speed but not in accuracy ($\overline{\beta}_{SAT \times Correctness} = 85$, $CrI = [57, 114]$, see Figure 3.5 left column). Validity interacts with trials type (correct vs. error) showing that the relative speed of an error compared to a correct response changed according to the validity of the cue, faster when invalid and slower when valid ($\overline{\beta}_{Correctness \times Validity} = 196$, $CrI = [129, 265]$, see Figure 3.5 top left panel). Finally the three-way interaction term proved to be significant and of opposite sign to the other interaction terms. This shows that the interaction between correctness and validity was reduced when accuracy was emphasized ($\overline{\beta}_{SAT \times Correctness \times Validity} = -111$, $CrI = [-147, -76]$) (see Figure 3.5 left column).

### 3.3.3.2. Drift Diffusion Model

Table 3.2 provides a summary of the estimated parameters (see also Appendix .5 for an evaluation of the goodness of fit) and Figure 3.4 a representation of the effect of the right-oriented probability cue on the starting point.

Interestingly the estimation of the intercept of the LMM (i.e. predicted value of the starting point when speed is emphasized and $p_{Right}$ Cue is equal to 0.5) on starting point suggests that participants were slightly biased towards right responses even though the cue indicated an equal right/left probability ($\overline{\alpha} = 0.530$, $CrI = [0.501, 0.558]$). The SAT instruction did not affect the starting point ($\overline{\beta}_{SAT} = -0.026$, $CrI = [-0.068, 0.012]$) although most of the probability mass of the coefficient suggest that the bias to the right pointed by the intercept is reduced in the accuracy condition. The SAT instruction interacted with the right probability cue : when speed was emphasized participants increased their bias toward the right response threshold according to the increase in right probability cue ($\overline{\beta}_{pRight} = 0.218$, $CrI = [0.113, 0.321]$) but less so in the accuracy condition ($\overline{\beta}_{SAT \times pRight} = -0.159$, $CrI = [-0.256, -0.056]$, see Figure 3.4).

### 3.3.3.3. EMG derived chronometric variables

The PMT results (see Table 3.3) mirrored the one found on RT (see Figure 3.5 middle vs left column).

As in Chapter 2 the MT proved to be slower when accuracy was emphasized ($\overline{\beta}_{SAT} = 16.66$, $CrI = [11.22, 23.11]$) and, in a lesser extent, when an error was made ($\overline{\beta}_{Correctness} = 3.25$, $CrI = [1.93, 4.64]$). Validity did have a really small but consistent effect on MT ($\overline{\beta}_{Validity} = 1.44$, $CrI = [0.18, 2.80]$) and opposite to the one found on RT and PMT. None of the factors seemed to interact (see Table 3.3 and Figure 3.5 right column).

| | $a_{Acc}$ | $a_{Spd}$ | $v_{right}$ | $v_{left}$ | $T_{erAcc}$ | $T_{erSpd}$ | $sv$ | $sz_{Acc}$ | $sz_{Spd}$ | $st$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Mean | 1.25 | 0.79 | 2.31 | -2.17 | 0.36 | 0.28 | 0.69 | 0.24 | 0.45 | 0.18 |
| SD | 0.26 | 0.15 | 0.61 | 0.82 | 0.07 | 0.05 | 0.34 | 0.19 | 0.25 | 0.06 |

| | | $z30_{Acc}$ | $z50_{Acc}$ | $z70_{Acc}$ | $z30_{Spd}$ | $z50_{Spd}$ | $z70_{Spd}$ |
|---|---|---|---|---|---|---|---|
| | Mean | 0.47 | 0.51 | 0.53 | 0.40 | 0.56 | 0.63 |
| | SD | 0.10 | 0.06 | 0.10 | 0.19 | 0.11 | 0.18 |

Table 3.2.: Means and standard deviations for the 16 parameters estimated with the DDM.
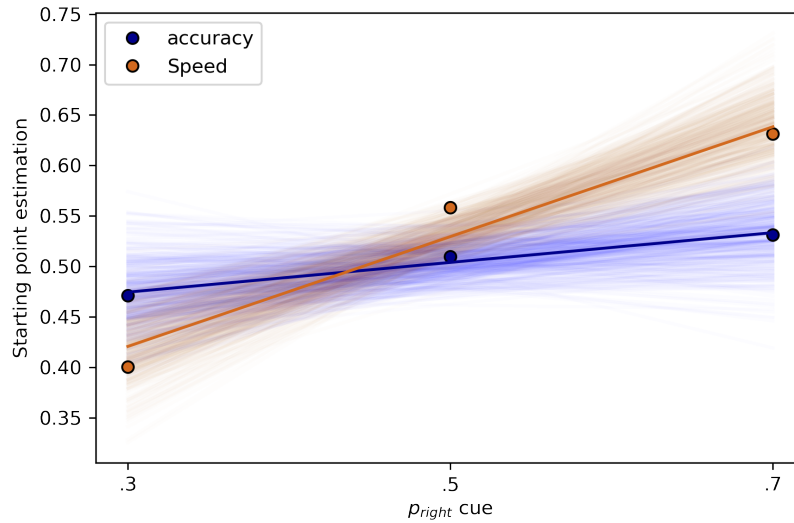


Figure 3.4.: Average estimation of the starting point of the DDM for each SAT and right probability cue levels. The lines represent 1000 random draws from the joint posteriors of the combined MCMC chains of the corresponding LMM fits. The thick line represent the predicted regression line with all parameters set at the mean of their respective marginal posterior distribution.
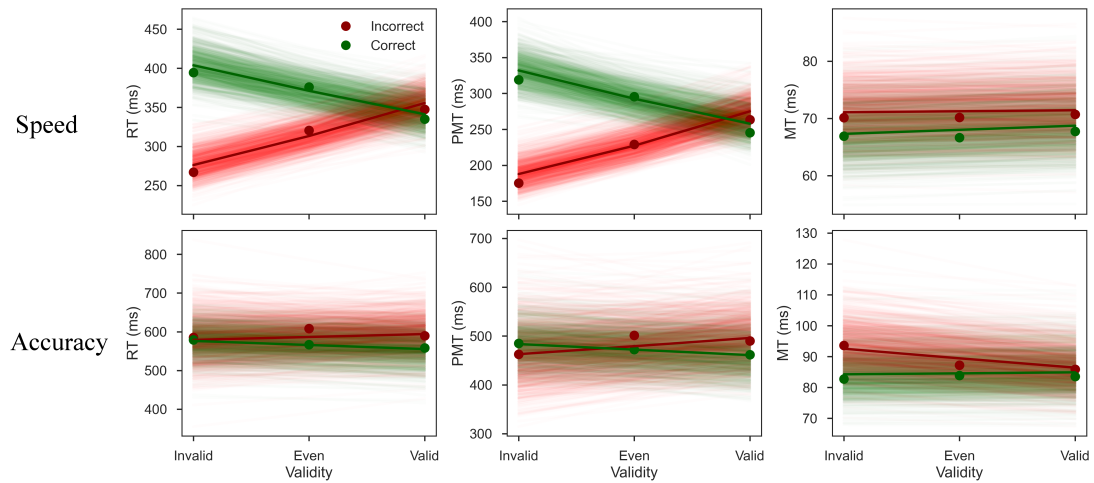
Figure 3.5.: Group average mean values for $RT$, $PMT$, and $MT$ (columns from left to right) plotted for each SAT condition (speed on the top row and accuracy on the bottom row), plotted against validity levels and separated for correct and error responses. The lines represent 1000 random draws from the joint posteriors of the combined MCMC chains of the corresponding LMM fits. The thick line represent the predicted regression line with all parameters set at the mean of their respective marginal posterior distribution.

| | PMT | | | | MT | | | |
|---|---|---|---|---|---|---|---|---|
| Predictor | Coeff. | SE | 2.5 | 97.5% | Coeff. | SE | 2.5% | 97.5% |
| Intercept | 293.14 | 17.59 | 261.83 | 330.74 | 68.12 | 4.10 | 60.28 | 76.26 |
| SAT | 180.20 | 27.37 | 127.85 | 236.05 | 16.66 | 3.01 | 11.22 | 23.11 |
| Validity | -64.75 | 14.36 | -93.64 | -37.29 | 1.44 | 0.66 | 0.18 | 2.80 |
| Correctness | -65.28 | 7.25 | -79.74 | -51 | 3.25 | 0.68 | 1.93 | 4.64 |
| SAT × Validity | 66.92 | 21.77 | 24.27 | 111.03 | -0.90 | 0.87 | -2.58 | 0.79 |
| SAT × Correctness | 90.28 | 15.60 | 62.94 | 124.24 | 0.64 | 1.09 | -1.41 | 2.86 |
| Validity × Correctness | 261.00 | 50.55 | 160.55 | 360.80 | -1.06 | 1.34 | -3.78 | 1.50 |
| SAT × Contr. × Corr. | -117.57 | 16.40 | -150.15 | -85.50 | -3.93 | 2.19 | -8.40 | 0.16 |

Table 3.3.: Results of the LMM models performed on PMT and MT .
Coeff. represents estimated coefficient of the LMM fitted on the log scale and back-transformed to the millisecond scale. SE represent standard error for the estimate; 2.5% and 97.5% represent, respectively, the lower and upper CrI around the estimate.

## 3.3.4. EMG signal analysis

Data rejection for the EMG baseline was restricted to trials presenting an EMG burst before the stimulus (see Section 3.3.2.7). The data rejection criterion's for the computation of the EMG slope was the same as for the behavioral analysis except that we also excluded error trials (18% of the data).

### 3.3.4.1. EMG baseline activity

The average time course of the running mean performed on the rectified EMG signal for both hands is displayed in Figure 3.6 along with the regression estimates from the LMM on the single trial baseline activity at stimulus onset.

Compared to the intercept of 2.502 mV ($CrI = [2.229, 2.800]$), the SAT had a small effect, showing that speed instructions induces a small tonic activity although the CrI barely excluded 0 ($\overline{\beta}_{SAT} = -0.105$, $CrI = [-0.215, -0.001]$). There was no effect of the right-oriented probability factor ($\overline{\beta}_{p_{Right}} = -0.002$, $CrI = [-0.025, 0.022]$) nor did it interact with the EMG side ($\overline{\beta}_{EMG \times p_{Right}} = 0.011$, $CrI = [-0.022, 0.043]$). The EMG side factor did have a numerical strong effect compared to the other effects, suggesting a lower baseline for the right EMG channel, but CrI were quite broad and included 0 ($\overline{\beta}_{EMG} = -0.152$, $CrI = [-0.360, 0.052]$). All interactions including SAT were centered on 0.
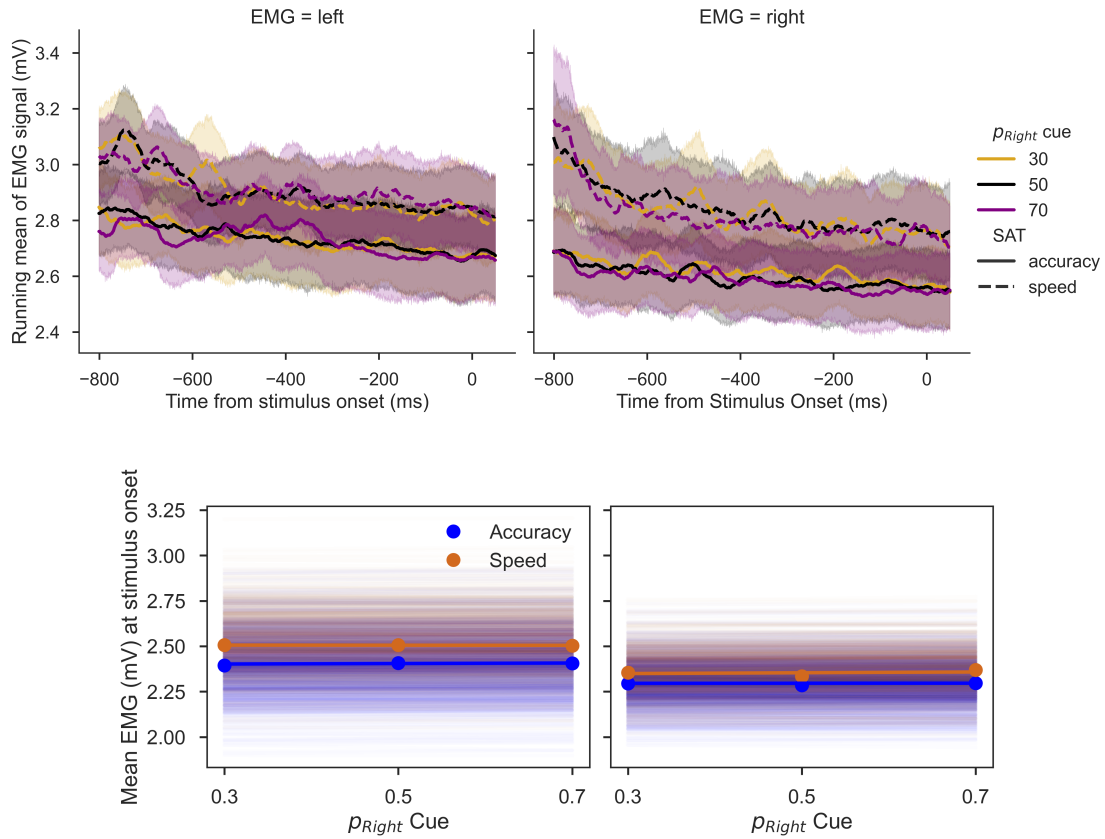
Figure 3.6.: Top : Average time course of the running mean of the rectified EMG signal across SAT and right probability cue conditions for left and right EMG channels. The shaded areas represent one SD of the running mean. The progressive reduction across time in the running mean is probably linked to the proximity of the previous EMG burst leading to the response (preceding stimulus and cue by respectively 1250 ms and 500 ms)

Bottom : LMM estimated on the mean baseline at stimulus onset across SAT conditions, right probability cue and EMG channel (left : left EMG, right : right EMG). The lines represent 1000 random draws from the joint posteriors of the combined MCMC chains. The thick line represent the predicted regression line with all parameters set at the mean of their respective marginal posterior distribution.

### 3.3.4.2. Response EMG slope

The average time course of the cumulative sum of the rectified EMG signal following the onset of the response EMG is displayed in Figure 3.6 along with the regression estimates from the LMM on the single trial estimated slope of the cumulative sum.
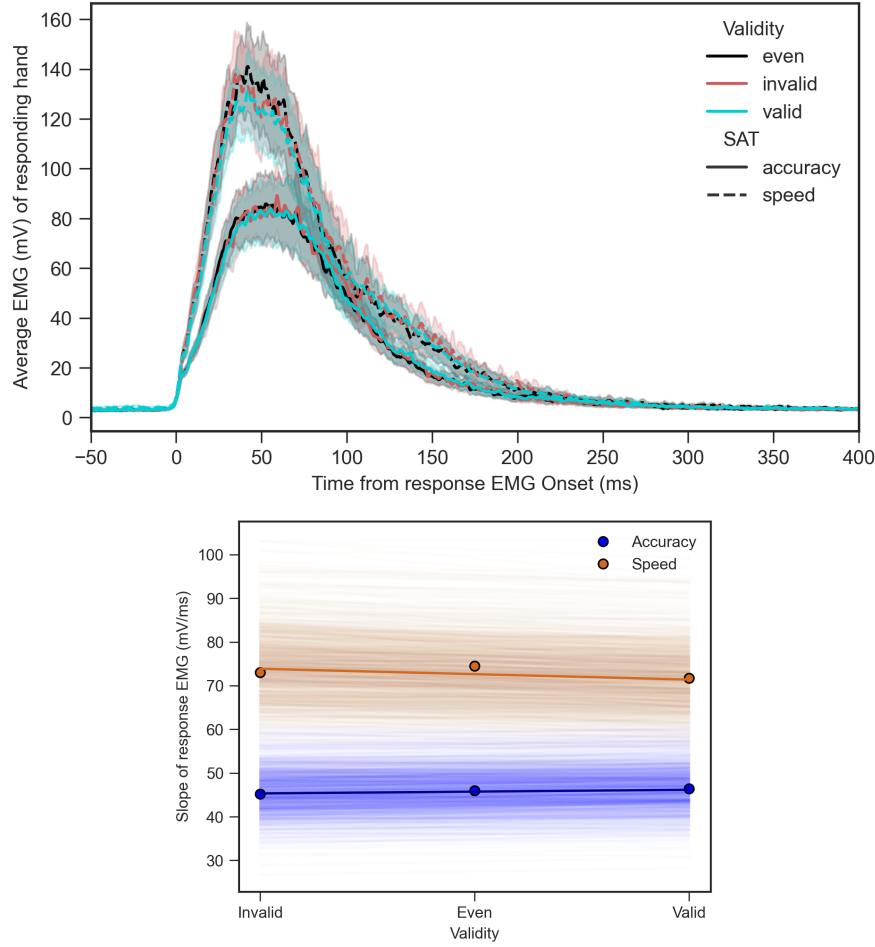
Figure 3.7.: Top: the average time course of the EMG signal of the responding hand after response EMG onset.
Bottom : fit from the LMM on the by-trial regression slopes performed on the cumulative sum of the 62 samples redressed EMG signal following EMG onset.

Compared to the intercept ($\overline{\alpha} = 73.06$, $CrI = [58.79, 90.33]$) the SAT had a high effect on the slope ($\overline{\beta}_{SAT} = -27.01$, $CrI = [-36.98, -16.64]$). However we observe no effect of the validity factor ($\overline{\beta}_{Validity} = -2.43$, $CrI = [-6.64, 1.59]$) nor its interaction with SAT ($\overline{\beta}_{SAT \times Validity} = 3.97$, $CrI = [-1.18, 9.01]$).

## 3.3.5. Discussion

LMM on RT and proportion correct showed that participants did use the cue about the upcoming stimulus by adjusting their starting point. The DDM indicated that they did so by varying the starting point of the accumulation. However, both of these results seems true only in the speed condition, given that the effects of cuing are

mostly cancelled out (or at least greatly reduced) when participants are instructed to favor accuracy over speed. This interaction show that participants did weakly, if at-all, rely on the cue to achieve the required accuracy level in the accuracy condition however they used the cue to achieve the required speed in the speed condition. The weak reliance on the cue in the accuracy condition could be because the task was not designed to be hard enough to encourage participants to use the cue in the accuracy condition [7].

### 3.3.5.1. Effect of validity on motor processes

Concerning motor processes, we find that the validity of the cue has an effect on the motor processes, as suggested by the behavioral experiment by Starns and Ma, 2018. However, contrary to what was suggested by the authors, and as assessed using the MT, the effect is small and of opposite sign compared to the effect of validity on PMT and RT. The nature of this unexpected effect remains to be found and we suggest an interpretation in the general discussion of the thesis. Although not conclusive because the uncertainty of the effect includes 0, the effect on the slope of the EMG burst leading to the response is coherent with the sign of the effect on MT. A steeper slope when the cue was invalid would produce a faster execution. We excepted that the cue could have an effect on preparatory motor processes such as the increase in tonic activity in the cued hand. However, the baseline of EMG at stimulus presentation did not seem to vary with the probability of the response suggesting. As the result of Weinberg, 2017, response probability does not affect response readiness of the effector muscle measured with transcranial magnetic stimulation or EMG baseline.

Concerning the sign of the effect of validity while RT and PMT proved to be shorter when the participant was given a valid cue (and under speed stress), MT proved to be longer. Another factor had an opposite effect on PMT and MT, the correctness of the response, first observed in Chapter 2 (and already observed by Rochet, Spieser, Casini, Hasbroucq, and Burle, 2014 in the context of a conflict task) was replicated in the present study. Allain, Carbonnell, Burle, Hasbroucq, and Vidal, 2004; Śmigasiewicz, Ambrosi, Blaye, and Burle, 2020 interpreted this result as reflecting evidence for an online executive control where participants try to stop the ongoing response. This interpretation does however not seem to explain the slow-down found in MT when the cue was valid. This opposition between PMT and MT, both on the correctness and the cue validity effects, is of primary importance on the role of motor processes in decision making and we come back to it in the General Discussion of the thesis (Section 6.1).

One additional surprising property of the results of validity on MT is the stability of the effect across participants. As allowed by hierarchical models such as LMMs, we have access to the estimation of individual parameters along the population parame-

---

[7]The average proportion of correct response in the accuracy condition was close to the level aimed with the prior task using the PRDM (0.88 *vs.* 0.90); we speculate that choosing a lower criterion would probably have increased the reliance on the cue in the accuracy condition.

ters (although these individual parameters are not independent due to the hierarchical nature of these models). Looking at the individual slopes of the effect of validity when instructions where to emphasize speed, we see that among the variables of Proportion correct, RT, PMT and MT only the MT shows an estimated effect of the same sign across all participants. This implies that, assuming that the individual parameters reflect the "real" effect of the participant (*i.e.* no measurement error), some individuals do not use the cue to make a decision but still display an effect on MT (see bottom right panel of Figure 3.8). Additionally, if we rank these individual estimates according to the effect size on the proportion correct we see that the rank of the participant is more or less preserved on RT and PMT but not on MT (see Figure 3.8).
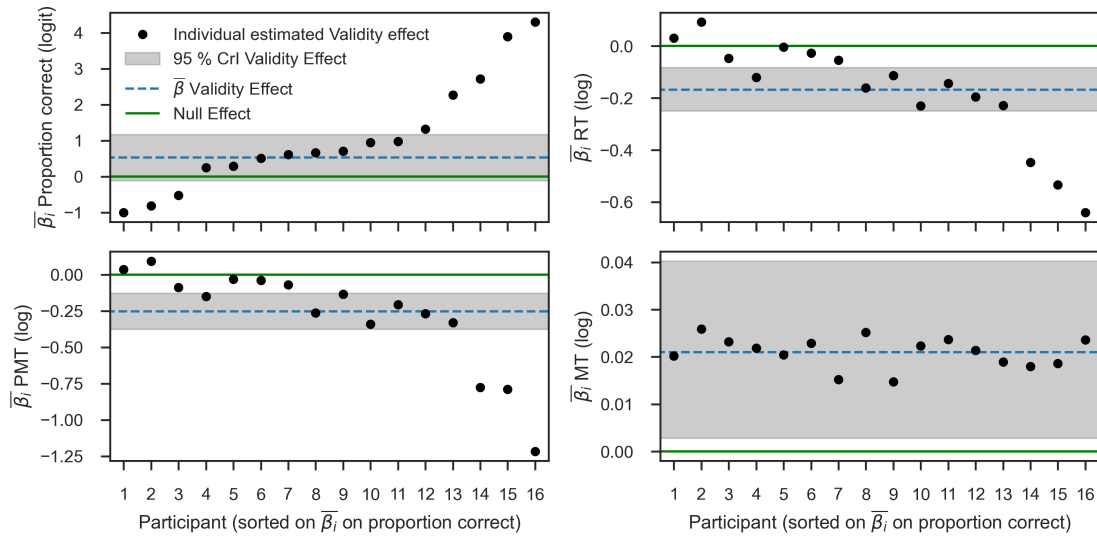


Figure 3.8.: Means of the marginal posterior distribution of the by-participant validity effect ($\overline{\beta_i}$) from the LMMs on proportion correct, RT, PMT and MT. The participants are sorted on the effect on proportion correct across all subplots. The green line shows the null effect while the blue dashed line shows the mean from the posterior distribution of the population parameter validity effect ($\overline{\beta}_{Validity}$ in the result section). The shaded area shows the CrI associated with the population parameter of validity effect. Note that contrary to all other reports of the results, all effects are displayed in the transformed scale of the variable (logit for proportion correct and log for RT, PMT and MT.

While drawing conclusions on variability of the effects would require more participants and a more adapted modelling effort (for a recently proposed method see Haaf & Rouder, 2017) we can still suggest an interpretation of the stability of the validity effect on MT across participants. Assuming this stability generalizes to larger samples and that the effect is genuinely linked to the presence of the cue. We propose that priming a motor response always (weakly) affects the motor component irrespective

of whether the participant actually uses the cue to make a decision (in line with the observation that SAT does not seem to modulate the effect of validity on MT).

### 3.3.5.2. Additional evidence on the effect of SAT on motor processes

One final observation should be highlighted. We observe for the third time in this thesis the effect of SAT on motor processes, first shown using EMG by Spieser, Servant, Hasbroucq, and Burle, 2017. In the present study we also replicate the finding of Spieser, Servant, Hasbroucq, and Burle, 2017 that the SAT does affect the slope of the EMG burst leading to the response. We do also show that the SAT affects the baseline EMG with a higher baseline activity when speed is stressed over accuracy. Our result show that SAT adjustments are present before the presentation of the stimulus as participants slightly pre-activate their muscles. This can, functionally be linked to a lower decision threshold (smaller distance between starting point and response threshold) and therefore that the motor system directly participates in the SAT adjustments. We come back to this issue in the general discussion of the thesis.

## 3.4. Conclusion

This chapter first demonstrated that when using the wrong parametrization of a decision making model such as the DDM one can draw false interpretations on the cognitive processes affected by a manipulation. While this was for example shown by A. Voss, Voss, and Klauer, 2010 using simulations, the present chapter shows the same conclusion but with empirical data and an actual measure of such cognitive processes. This result calls for the second part of the chapter. The standard DDM being incapable to separate a cognitive bias from a response execution difference, if a manipulation of the former generates the latter this would seriously question the validity of the starting point established though such manipulations (Leite & Ratcliff, 2011; Mulder, Wagenmakers, Ratcliff, Boekel, & Forstmann, 2012; Ratcliff & McKoon, 2008; A. Voss, Rothermund, & Voss, 2004). However, while the effect on motor processes as measured by the MT is present, it is of low effect size compared to the effect size on RT (approximately 1.8% of the absolute effect on RT). It is hence unlikely that motor processes have an important impact on bias manipulations. The properties of this effect, along with the one on error responses and the possible contribution of motor processes to the decision threshold reduction in SAT, really question the role of motor processes in decision.

# 4. The Decisive Role of Non-Decision Time to Understand Decision Making Model Parameters

## Sommaire

# 4.1. Note

# 4.2. Decomposing RTs in decision and non-decision times

The DDM used in the previous chapters, as well as all of the decision making models presented in the thesis (S. Brown & Heathcote, 2005; Heathcote & Love, 2012; Ratcliff, 1978; Ratcliff & Rouder, 1998; Stone, 1960; Usher & McClelland, 2001), decompose $RT$s into "decision time" ($T_D$) and a time outside the decision, or "non-decision time" ($T_{er}$).

One way to provide evidence for this decomposition is to use selective influence tests on the $T_D$ and $T_{er}$. For instance, one expects that manipulating the difficulty of the perceptual decision will specifically increase $T_D$ through the drift rate. This result is indeed observed in several studies (Gomez, Ratcliff, & Childers, 2015; Palmer,

Huk, & Shadlen, 2005b; A. Voss, Rothermund, & Voss, 2004). Another of such selective influence test is manipulating the SAT, thus presumably selectively lengthen the $T_D$ through an increase in the boundaries. It has been shown that while $T_D$ is indeed lengthened, SAT also affects the estimation of $T_{er}$ (Palmer, Huk, & Shadlen, 2005b; Ratcliff, 2006; A. Voss, Rothermund, & Voss, 2004). Similarly, decreasing the brightness of a visual stimulus (Servant, White, Montagnini, & Burle, 2016) has been shown to lengthen $T_{er}$ as expected, but also to impact the parameters defining $T_D$ such as the *drift*. Same for increasing the force needed to produce a response (Gomez, Ratcliff, & Childers, 2015; Ho, Brown, & Serences, 2009; A. Voss, Rothermund, & Voss, 2004) shown to lengthen $T_{er}$ as expected, but also to impact the parameters defining $T_D$ such as the *bias* (A. Voss, Rothermund, & Voss, 2004) and *boundaries* (Gomez, Ratcliff, & Childers, 2015; Ho, Brown, & Serences, 2009).. These results may receive two alternative explanations. On one hand, they might result from the fact that the experimental manipulations under scrutiny do not have the expected selective influence on cognitive processes. On the other hand, they may reflect a failure of the model and/or of the fitting process to separate decision from non-decision processes. Hence, to date, the use of experimental manipulations with behavioral data alone does not provide evidence for interpreting $T_D$ and $T_{er}$ as decision and non-decision times.

## 4.3. Present study

We tested how faithfully the non-decision time $T_{er}$ estimated in the Drift Diffusion Model (DDM, Ratcliff & Tuerlinckx, 2002) reflects the cognitive delays it is intended to capture: $T_{encoding}$ plus $T_{response}$. To that end, we designed a sensori-motor decision task with two specific features: response time decomposition and targeted experimental manipulations. The $RT$s of every trial were decomposed into pre-motor and motor times ($PMT$ and $MT$) as in previous chapter, based on muscle activity onsets revealed by the EMG of the responding hand. Following Chapter 2, we assumed a tight link between MT and $T_{response}$, which could thus be measured rather than estimated. As a result, the remainder $PMT$ was thought to correspond to the sum of $T_{encoding}$ and $T_D$. The DDM could then be fitted either to $RT$s, to estimate $T_{er}$ overall, or be fitted to $PMT$s, to provide a reasonable estimate of $T_{encoding}$. Assuming that the decomposition of RT using EMG corresponds to the decomposition estimated by the DDM we expect that the parameters governing $T_D$ (drift rate, boundaries and bias) should be the same whether estimated on $RT$ or $PMT$.

There were three experimental manipulations. The two first ones were predicted to primarily affect $T_{er}$: a manipulation of stimulus contrast was predicted to affect the encoding process ($T_e$), and a manipulation of the response force required was predicted to affect the response execution process ($T_r$). Finally, the participants' strategy was manipulated with a classic Speed-Accuracy trade-off. The details for these three manipulations are as follows.

### 4.3.1. Affecting encoding processes

Participants had to choose which of two sinusoidal gratings (Gabor patches) had the highest contrast. We manipulated the mean contrast of the gratings, while keeping their absolute difference constant. It has been shown that stimuli become more accessible, and are processed faster, when their contrast increases e.g. Harwerth and Levi, 1978, in a simple reaction time task. We thus expected shorter $T_{encoding}$ for higher contrasts. Conversely, according to Fechner's law on sensation, the relative (perceived) contrast difference between the gratings should decrease as their mean contrasts increase, because the absolute difference was kept constant. We thus expected that discriminating between the two Gabor patches will be more difficult for higher contrasts (e.g. see Ratcliff, Voskuilen, & Teodorescu, 2018), thereby inducing longer $T_D$ through a decrease in the *drift* parameter as in Chapter 2. In sum, the contrast manipulation was expected to have an opposite effects on $T_{encoding}$ and on $T_D$.

### 4.3.2. Affecting motor processes

We manipulated the force required to produce the responses, a manipulation known to affect processes related to motor execution (Burle, Possamaï, Vidal, Bonnet, & Hasbroucq, 2002). We expected that an increase in the required response force will result in an increase of $T_{response}$. Consistently, we expected that the $T_{er}$ fitted on PMT (i.e. putatively $T_{encoding}$), will not be affected by the Force manipulation. As mentioned earlier, it has been reported that force requirements can also influence decision related parameters (Bias and threshold, see Gomez, Ratcliff, & Childers, 2015; A. Voss, Rothermund, & Voss, 2004) as estimated by a DDM but no interpretation has been offered. Our EMG decomposition and the hypothesized absence of decision latency in $MT$ allows the following subsidiary expectations. If these effects are truly linked to decision processes happening before response muscle triggering, they should be invariant whether the fit is performed on RT or on PMT. Conversely, observing that the effects vary between both fits would suggest an incompatible separation of $T_D$ and $T_{er}$ with the EMG decomposition.

### 4.3.3. Speed accuracy trade-off (SAT)

Finally, we manipulated the SAT level required from participants through verbal instructions. This manipulation is classically/mainly described as an adjustment of the level of evidence needed before a decision is taken, hence most often linked to a change in the *boundary* parameter of the DDM. However, the manipulation of SAT has also been suggested to modulate the speed of encoding processes (Steinemann, O'Connell, & Kelly, 2018) and of motor execution (Spieser, Servant, Hasbroucq, & Burle, 2017; Steinemann, O'Connell, & Kelly, 2018, and the two previous chapters of the present thesis).

We therefore seek to verify that the predictions from both contrast and force manipulations are true independent of the SAT level as, a violation of these predictions in any SAT level would question the interpretation of the $T_{er}$ parameter and the way human participants achieve SAT.

## 4.4. Methods

### 4.4.0.1. Participants

Sixteen participants (6 men and 10 women, mean age = 24.5 years, 2 left-handed) that were students in Aix-Marseille University, were recruited for this study. All participants reported having normal or corrected vision, and no neurological disorders. The experiment was approved by the ethical experimental committee of Aix-Marseille University, and by the "Comité de Protection des Personnes Sud Méditerrannée 1" (Approval n° 1041). Participants gave their informed written consent, according to the declaration of Helsinki. They received a compensation at a rate of 15€per hour.

### 4.4.0.2. Apparatus

Participants performed the experiment in a dark and sound-shielded Faraday cage. They were seated in a comfortable chair about 100 cm away from a 15 inch CRT monitor that had a refresh rate of 75 Hz. The CRT monitor was gamma corrected by a psychophysical procedure provided by the software PsychoPy (Peirce, 2007). Responses were given by pressing either a left or a right button with the corresponding thumb. The buttons were fixed on top of two cylinders (3 cm in diameter, 7.5 cm in height). The cylinders were fixed on a tablet and separated by a distance of 20 cm. The buttons were mounted on force sensors that recorded a continuous measure of the force produced at a sampling rate of 2048 Hz. The behavioral response was recorded when a force threshold was exceeded. The device allowed adjusting the force threshold needed for a response to be received. The threshold was manipulated across conditions, as described below. Response signals were transmitted to the parallel port of the recording computer. At button press, participants received a 3ms sound feedback (1000 Hz pure tone).

The participants' forearms and hypothenar muscles rested comfortably on the table, to minimize muscle recruitment during response execution. We measured the EMG activity of the flexor pollicis brevis of both hands with two electrodes placed 2 cm apart on the thenar eminences. This activity was recorded using a BioSemi Active II system (BioSemi Instrumentation, Amsterdam, the Netherlands). The sampling rate was 2048 Hz.

### 4.4.0.3. Stimuli

Stimulus presentation was controlled by the software PsychoPy (Peirce, 2007). Each stimulus was composed of two vertical oriented Gabor patches, on the left and right of a fixation cross separated by 1.4 visual angle degrees. The Gabor patches had a spatial frequency of 1.2 cycles per visual angle degree and a size of 2.5 visual angle degrees each. At each trial, the same amount of contrast (7%) was subtracted to the randomly assigned incorrect Gabor patch and added to the correct one, resulting in a 14% contrast difference. The task of the participant was to select the patch with the highest contrast.

### 4.4.0.4. Experimental manipulations

**Contrast**   To manipulate contrast, we chose to manipulate the mean contrast of both Gabor patches while keeping the difference of 14% constant. Six levels of mean contrast (23%, 37%, 51%, 65%, 79%, 83%) where selected based on a pilot study, targeting a performance that would span from near-perfect accuracy to near chance level. The mean contrast across both patches was randomly chosen at each trial with a fixed rate of occurrence (1/6) within every block.

**Force level settings**   The Force level factor had two levels: strong and weak. These levels were tailored to each participant before the experiment started. Participants were asked to press twice the right and then left button, with the maximum force they could apply. The maximum voluntary force was defined as the maximum between the two trials from the weakest of the two arms. Defining maximum voluntary force based on the weakest arm was chosen to avoid muscular fatigue from the non-dominant hand. The actual force levels for the strong and weak conditions were then defined as, respectively, 2% and 20% of this maximum voluntary force level (generating force levels respectively around 120 and 1200N with some participants up to 230 and 2300N).

**SAT**   The speed-accuracy trade-off (SAT) instruction was manipulated between blocks. Participants were instructed that "Speed" instructions required a mean reaction time near 400 ms and that "Accuracy" instructions required a percentage of correct responses near 90% while maintaining RTs below 800 ms. Each block started with the presentation on the center of the screen of its corresponding instruction: the French word for Speed ("Vitesse") or Accuracy ("Précision"). The end of each block was followed by feedback about mean reaction time and mean accuracy, along with oral feedback from the experimenter, if the participant had not satisfied the condition goals of the block.

### 4.4.0.5. Procedure

All participants performed a single experimental session with 24 blocks of 100 trials each. Session duration was approximately 1h30 including a training session of 15

minutes and self-paced breaks between each block. Participants were asked to keep their gaze on the central fixation cross throughout each block, and to respond to the visual stimuli according to the corresponding SAT instruction.

The training session started with 40 trials without specific SAT instructions, followed by 2 blocks of 10 trials in the Speed condition, followed by 2 blocks in Accuracy condition, and ended with 2 blocks of 10 trials with alternating instructions. During the experimental session, SAT instructions alternated every three consecutive blocks. The force settings varied every six blocks, with an on-screen message to inform the participant beforehand. The order of the SAT instructions and the force requirement was counterbalanced across participants so that every possible order combination was presented to 4 participants. Within each block, the 6 levels of mean contrast value were fully randomized across trials. No response deadline was applied, and the inter-trial interval was fixed to 1000 ms from button press to next stimulus onset.

### 4.4.0.6. EMG processing

The EMG recordings were read in Python using the MNE module (Gramfort, Luessi, Larson, et al., 2013). The signal was filtered using a Butterworth 3rd order high pass filter at 10Hz from the scipy Python module (Oliphant, 2007), then segmented by-trial in windows between 150 ms before and 1500 ms after stimulus onset. We used a variance-based method to detect whether EMG activity was significantly above threshold in either hands' channels. The precise burst onset was then identified with an algorithm based on the "Integrated Profile" of the EMG burst (see Liu & Liu, 2016; Santello & Mcdonagh, 1998, for details). If the algorithm failed to locate or detect the EMG burst onset, the experimenter corrected or added them manually. At this stage of signal processing, the experimenter was unaware of the trial type he was correcting for, to avoid any bias. Every muscular event (above-threshold change in the signal followed by a return to the baseline) in the trial was marked, even when the activation did not lead to an overt response. This provided information about how many times the muscle was triggered in the course of every trial.

Motor time ($MT$) was defined as the time between the onset of EMG burst of the responding hand's recorded muscles and the force threshold crossing recorded at the button press. Pre-motor time ($PMT$) was defined as the time between stimulus onset and the EMG burst onset. We observed multiple EMGs in 21% of trials. Such observations are not new (e.g. Chapters 2 and 3), but a precise account of these multiple activities is still lacking. Minimally, they show that participants were not always engaged in a pure sequential encoding-decision-execution process. Therefore, we removed these trials from all the analysis in the study. Studies developing a generative model of these trials and how they relate to sequential sampling models are needed.

## 4.4.1. Statistical procedure

### 4.4.1.1. Bayesian Statistics

All analysis were performed in a Bayesian framework. Bayesian methods try to estimate an unknown parameter (or set of parameters) and the uncertainty around it. More explicitly, Bayesian methods combine prior information and Bayes' rule to quantify the likelihood of the parameters by generating a posterior distribution for each of them. This posterior distribution can be naturally interpreted as the probability of any given parameter value given the priors, the data, and the tested model. In our study we summarize the posterior distribution using the mean and the Credible Interval (CrI), the 95% Highest Probability Density interval (HPD; Kruschke, 2010).

### 4.4.1.2. Hierarchical Modelling

All models, including linear models, were constrained to follow a hierarchical structure with parameters from each participants as units assumed to be drawn from a population distribution. This parametrization allows to estimate population parameters (*e.g.*the slope of the effect of stimulus contrast on *RT*s) along with individual parameters (*e.g.*the inter-individual differences in the slope of contrast with RT), often referred to respectively as fixed and random effects in the case of linear models. Hierarchical modelling remains Bayesian. Thus, it preserves the uncertainties associated with parameter values while sharing information between participants to estimate individual parameters. This parametrization allowed directly testing our hypotheses, by comparing the posterior distributions for the population effects across conditions.

### 4.4.1.3. Estimated difference between condition levels $\hat{d}$

In order to estimate the magnitude of the difference ($\hat{d}$) between the levels of the experimental factors SAT, Force and contrast we chose to use the predictions of the fitted linear models (described below). For each dependent variable with these three factors as predictors we first computed the predicted difference between both SAT level within other predictors kept at 0 (see Appendix C.1). We then computed the predicted differences between the lowest and highest contrast level as well as the weak and high force condition for each SAT instruction separately. The report of the result is then composed of the effect of SAT, the effects of force and contrast in each SAT condition and the difference of these effects between SAT conditions. Due to the Bayesian nature of the analysis, the uncertainties associated with the regression parameters are preserved in the predictions. Thanks to the hierarchical nature of the regression models, we directly estimate a population difference. Both methods therefore allow to directly infer the population level effect size with its uncertainty using the posterior distribution of the predicted difference. The strength of evidence in favor of the presence or absence of an effect was determined based on the likeliest values of the differences as provided by the mean and the 95% CrI of the posterior

distribution.

### 4.4.1.4. Linear Mixed Models

We used linear mixed models (LMM) on the log transformed $RT$, $PMT$, and $MT$, and generalized linear mixed models (gLMM) for the proportion of correct responses. Formally, LMM model the dependent variable as drawn from a normal distribution who's parameters are constrained by the experimental design (e.g. the mean of the normal distribution changes with SAT instructions).

Given our analysis plan, we derived generic LMMs for $RT$, $PMT$ and $MT$ where all fixed effects and all random effects were estimated. The effects of the experimental factors were modeled on the mean parameter for normally distributed dependent variables (DV), assuming equal variance across conditions.

$$y_{ji} \sim \mathcal{N}(\mu_j, \sigma^2) \tag{4.1}$$

$$\begin{aligned} \mu_j = \alpha_j + \ & \beta_{1j}\ SAT + \ \beta_{2j}\ FC + \ \beta_{3j}\ Con. \\ & + \beta_{12j} SAT \times FC + \ \beta_{23j}\ SAT \times Con. \\ + \beta_{13j}\ & FC \times Con. + \ \beta_{123j}\ FC \times SAT \times Con. \end{aligned} \tag{4.2}$$

Where $y_{ji}$ represents the modeled DV ($RT$, $PMT$ or $MT$) on the $i$th trial for the $j$th participant and is assumed to be normally distributed with mean $\mu_j$ and standard deviation $\sigma^2$. As seen in Equation 4.2, $\mu$ is dependent of the experimental factors (Con. for mean contrast, FC for Force Condition and SAT for speed accuracy trade-off instructions). The LMM where only fitted on correct responses as we already reported the effect on these variables on errors in Chapter 2 and, for the sake of simplicity, to limit the analysis to three factors.

Accuracy was modelled with a gLMM, whereby proportion of correct responses is hypothesized to follow a Bernoulli distribution, modulated by the same factors than in Equation 4.2:

$$p(response_{ij} = 1) \sim Bernoulli(logit(\mu_j)) \tag{4.3}$$

To reiterate, in the LMMs and the gLMM, the intercept and all factors and interactions are modeled as random effects :

$$\alpha_j \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2) \tag{4.4}$$

$$\beta_{xj} \sim \mathcal{N}(\mu_{\beta_x}, \sigma_{\beta_x}^2) \tag{4.5}$$

where $\mu_\alpha$ and $\mu_{\beta_x}$ are the population estimated intercept and regression coefficient and $\sigma_\alpha^2$ and $\sigma_{\beta_x}^2$ the estimated random effect of the population sampling.

**Fitting procedure for the g/LMMs**   For each LMM and gLMM, six MCMC sampling processes were run in parallel, each composed of 2000 iterations among which the first 1000 samples were discarded as warm-up samples. We assessed convergence

of the MCMC chains both by computing the potential scale reduction factor ($\hat{R}$, see Gelman & Rubin, 1992) and by means of visual inspection of the MCMC chains. We visually checked the assumptions of the linear regression by inspecting the normality of the residuals through QQ-plots and assessment of homeoscedasticity. The LMM and generalized LMM were fitted with a custom Stan code, available in the online repository, inspired from the code provided by Nicenboim, Vasishth, Engelmann, and Suckow, 2018 and using the `pystan` package (Stan Development Team, 2018). The summary statistics and plots of the parameters were created using `arviz` python package (version 0.4.1, Kumar, Carroll, Hartikainen, & Martin, 2019).

**Priors for the g/LMMs**   The priors for the LMM and gLMM are intended to ease the fitting procedure, we chose to use the relatively broad informative priors used in Chapter 2 and described in Appendix A.2.

### 4.4.1.5. Drift Diffusion Modelling

**Model fitting procedure**   We used the implementation of a hierarchical Bayesian DDM provided in the *HDDM* python package (Wiecki, Sofer, & Frank, 2013). Note that *HDDM* uses the diffusion coefficient (See Ratcliff & McKoon, 2008, for a review of the DDM parameters) as a scaling parameter by fixing it to a value of 1 (contrary to a value of 0.1 in some applications of the DDM). For each model on $RT$ and on $PMT$, both in the "Model selection" section below and for the model including co-variates, we ran 32500 burn-in samples and 2500 actual recorded samples across four Markov chains Monte-Carlo (MCMC). We inspected each parameter of each chain visually to assess whether they reached their stationary distribution, and whether the $\hat{R}$ (Gelman & Rubin, 1992) was under the conventional threshold of 1.01. Additionally, we examined the auto-correlation of each chain to ensure that samples were drawn independently. For the priors, because our design is canonical and in order to ease convergence, we used the default informative priors used in HDDM based on the work of Matzke and Wagenmakers, 2009. Almost all parameters were estimated individually with the constrain of being drawn from a common normal distribution (or half-normal depending on the boundaries, *e.g.* variability parameters cannot have a negative value). Only the inter-trial variability parameters of the drift rate, of the bias, and of the non-decision time were estimated at the group-level because they are notoriously difficult to estimate (Boehm, Annis, Frank, et al., 2018; Wiecki, Sofer, & Frank, 2016).

**Model selection**   We designed a base model and added parameters according to our hypotheses. The base model was chosen based on previous studies. For this base model, the boundary parameter was free to vary with SAT instructions. The drift rate was free to vary with the contrast[1], as this parameter has been shown to be associated

---

[1]As boundaries were coded as right and left responses (respectively upper and lower threshold), and in order to avoid estimating one drift for each combination of stimulus side and contrast, the model was coded to take negative drift value when the correct stimulus was on the left

with stimulus strength. The $T_{er}$ was free to vary with SAT, as it has been observed that this parameter also varies with SAT conditions (Palmer, Huk, & Shadlen, 2005b; Ratcliff, 2006; A. Voss, Rothermund, & Voss, 2004), with the Force Condition and with the contrast factor, as all three are the factors of interest in the study of $T_{er}$ [2]. The accumulation bias was free to vary for each participant. We also added inter-trial variability of the drift rate and the non-decision time, because of their ability to reduce the influence of contaminant fast-trials (Lerche, Voss, & Nagler, 2017). Finally, we added the inter-trial variability of the starting point parameter which was free to vary with SAT instructions, because it is often reported that the latency contrast between errors and correct response does change according to the SAT condition and that this pattern is captured by a different bias variability.

In addition to the base model, we tested the following hypothesis, and combinations thereof: whether the drift rate also varies with SAT (Rae, Heathcote, Donkin, Averell, & Brown, 2014), or with the Force Condition (A. Voss, Rothermund, & Voss, 2004); and whether the bias and the boundaries are variable between Force conditions (Gomez, Ratcliff, & Childers, 2015; Ho, Brown, & Serences, 2009; A. Voss, Rothermund, & Voss, 2004). The 16 possible combinations of hypotheses are summarized in Table .4.

We used the deviance information criterion (DIC) to select among competing models. The DIC is an analog to the Akaiake information criterion (AIC) generalized to the hierarchical Bayesian estimation method, in which the improvement of the log-likelihood is weighted against the cost of additional parameters. Because, it has repeatedly been shown that DIC tends to select over-fitted models, we also report for each model the Bayesian predictive information criterion (Ando, 2007, BPIC). BPIC is intended to correct DIC's bias in favor of over-fitted models by increasing the penalty term for the number of parameters. For all these measures, a lower value of DIC or BPIC indicates a preferred model.

**DDM regression analysis**  Once the best-fitting model was identified and selected, the effects of the experimental factors on the parameters were assessed by further embedding a hierarchical regression in the model fitting procedure (Boehm, Marsman, Matzke, & Wagenmakers, 2018). The three experimental factors and their interactions were included as predictors in the regression, on the condition that had been left free to vary across conditions in the model selection procedure. Each parameter that is free to vary with one or more factors was estimated with one intercept and one slope for each factor and interactions. This allowed to use the posterior distribution of intercept and slopes to test directly for the presence and the direction of an effect by inspecting whether 0 is included in the posterior distribution. We compared the results of joint DDM-regression fits on the $PMT$ and $RT$.

The hierarchical nature of the data is preserved in these models because each intercept and slope parameters is estimated as being drawn from a population distribution.

---

[2]We started by including models that do not allow the parameter $T_{er}$ to take different values for the different contrast levels but these models failed to converge probably because of the effects reported in the behavioral results section

The parameters that do not vary with experimental factors (*i.e.*inter-trial variability of the drift rate and the non-decision time) are estimated as described in the model selection section. The inter-trial variability of the bias was free to vary between SAT instructions but the corresponding effect size was not estimated with a regression. This is because, first, we only have one estimate for the population due to the difficulty to estimate it and, second, we do not have specific hypothesis about this parameter. As in the model selection procedure, the models were fitted using the $HDDM$ python package (Wiecki, Sofer, & Frank, 2013).

### 4.4.1.6. Fast-guess detection and removal

Fast guess trials can be problematic when studying decision making in the context of evidence accumulation models. Before performing any statistical analysis, we applied an exponentially weighted moving average filter (EWMA; Vandekerckhove & Tuerlinckx, 2007) as in the previous chapters (see Section 2.3.1.6). Contrary to previous chapters, the EWMA filter was applied for each participant's $PMT$ distribution, separately in the speed and accuracy conditions. $PMT$s rather that $RT$s were used for the EWMA, first because Chapter 2 showed a high reliabilitiy when the method was applied separately to $PMT$ and $RT$, and second because trials that do not appear very fast on $RT$ can sometimes be fast on the $PMT$ and therefore be problematic trials when fitting a DDM on the $PMT$ as done here. The figures illustrating these rejection procedures can be found in the online repository.

## 4.5. Results

The method implemented for detecting EMG onsets imposed an $RT$ upper limit of 1500 ms, whereby 1% of the trials were excluded. Trials with low signal-to-noise ratio or with high spontaneous tonic activity that resulted in uncertain EMG onset detection were excluded (7%). Trials that presented more than one EMG activity (see Method section on EMG analysis) were also excluded (21%). Finally, the trim criterion derived from the fast-guess detection method lead to the exclusion of 8% of the data. Thus, the combined EMG and statistical criteria resulted in the exclusion of 37% of the trials. Censoring errors, for the LMM analysis of RT, PMT, and MT, removed 13% of the remaining data. On average, 1513 trials (SD = 310) were available per participant. All estimated differences ($\hat{d}$) are presented on the data scale, milliseconds for chronometric variables and proportion correct for accuracy.

### 4.5.1. Behavioral results (response times and error rates)

#### 4.5.1.1. Linear mixed models

The following observations are illustrated in Figure 4.1 and in Figure 4.2 (two left columns). When speed is emphasized, $RT$ decreases, $\hat{d}_{SAT} = -165, CrI = [-197, -134]$,
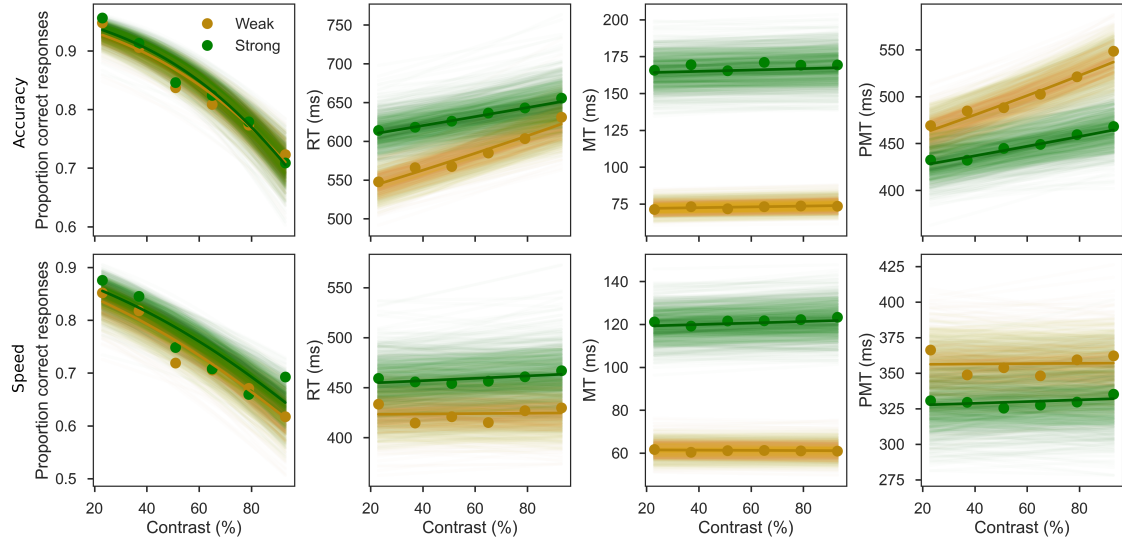
Figure 4.1.: Average values for proportion correct, $RT$, $MT$, and $PMT$ (columns from left to right) plotted for each SAT condition (accuracy on the top row and speed on the bottom row), broken down by contrast levels and by force condition. The lines represent 1000 random draws from the joint posteriors of the combined MCMC chains of the corresponding G/LMM fits. The thick line represent the predicted regression line with all parameters set at their maximum a posteriori value.
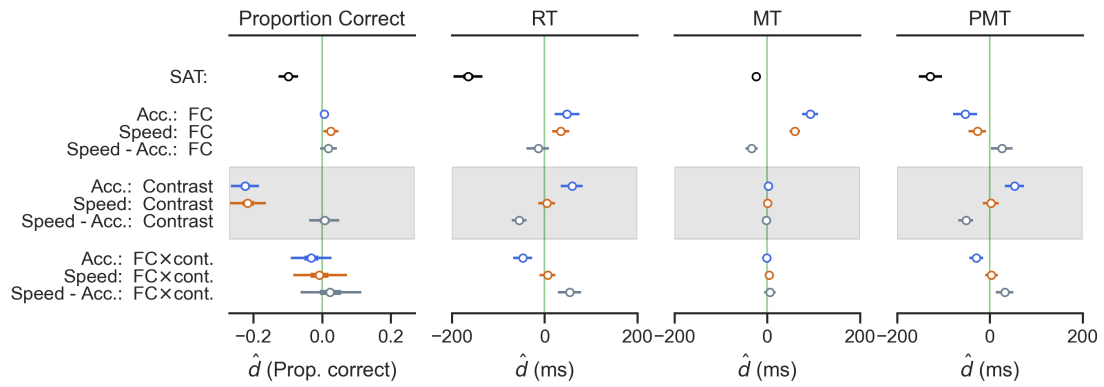


Figure 4.2.: Estimated differences between condition levels ($\hat{d}$) for SAT, Force (FC), Contrast and their interactions on the raw scale of the data in the Accuracy condition (blue), the Speed condition (chocolate), and the difference between both conditions (i.e. interaction; grey). Dots represent the maximum a posteriori, and bars the 2.5% and 97.5 % HPD of the corresponding marginal posterior distributions.

and so does the accuracy rate, $\hat{d}_{SAT} = -0.10$, $CrI = [-0.13, -0.07]$. When participants are asked the emphasize their accuracy, an increase in contrast lengthens $RT$s, $\hat{d}_{Contrast:Acc.} = 60$, $CrI = [35, 83]$, and reduces response accuracy, $\hat{d}_{Contrast:Acc.} = -0.22$, $CrI = [-0.27, -0.18]$. This effect of contrast on $RT$ proved to be different between SAT instructions $\hat{d}_{Contrast:Speed-Acc.} = -55$, $CrI = [-70, -39]$ , indicating that the contrast effect is essentially canceled when participants are asked to speed their responses, $\hat{d}_{Contrast:Speed} = 5$, $CrI = [-13, 22]$. The proportion of correct responses however displayed similar effects of Contrast in both the Speed and Accuracy conditions, $\hat{d}_{Contrast:Speed} = -0.22$, $CrI = [-0.27, -0.16]$, $\hat{d}_{Contrast:Speed-Acc.} = 0.02$, $CrI = [-0.01, -0.04]$.

When force requirements are higher, $RT$ increases both in the Accuracy, $\hat{d}_{Force:Acc.} = 48$, $CrI = [22, 75]$ and the Speed conditions, $\hat{d}_{Force:Speed} = 35$, $CrI = [16, 53]$. The proportion of correct response is not affected by the Force factor neither in accuracy $\hat{d}_{Force:Acc} = 0.01$, $CrI = [-0.01, 0.02]$ nor in the speed condition $\hat{d}_{Force:Speed} = 0.03$, $CrI = [0.00, 0.05]$ although the CrI barely included 0.

Unexpectedly, the interaction between Force and Contrast had an effect on $RT$ selectively in the Accuracy condition, $\hat{d}_{Force \times Contrast:Acc.} = -47$, $CrI = [-68, -27]$ but not in the Speed condition $\hat{d}_{Force \times Contrast:Speed} = 7$, $CrI = [-11, 23]$. The proportion of correct responses was not sensitive to the interaction between force and contrast neither in Accuracy $\hat{d}_{Force \times Contrast:Acc} = -0.03$, $CrI = [-0.09, 0.03]$) nor in Speed, $\hat{d}_{Force \times Contrast:Speed} = -0.01$, $CrI = [-0.08, 0.07]$.

### 4.5.1.2. Drift Diffusion Model selection

The model selection procedure is fully described in Appendix C.2, with DIC and BPIC estimates summarized in Table .4. In the model that was ultimately selected, one boundary parameter was estimated for each combination of SAT and force condition levels, one drift for each level of contrast, one starting point for each level of force level, and one non-decision time for each experimental cell of the three factors SAT × Force × Contrast (see Table .4). The effects of the experimental factors on these model parameters are summarized in Table 4.1 and spelled out below.

### 4.5.1.3. Effects on $T_{er}$

Emphasizing speed over accuracy revealed a strong effect of SAT on $T_{er}$. SAT and force interacted: increasing force had a strong effect on $T_{er}$ in the accuracy condition, and the interaction term indicated a smaller yet reliable effect in the speed condition (see left column of Figure 4.3). SAT and Contrast also interacted: there was no evidence for an effect of contrast on $T_{er}$ in the accuracy condition, but the interaction term indicated a negative effect when speed was emphasized (see left column, middle and bottom panels, of Figure 4.4). Finally, there was no evidence for an interaction between Force and Contrast, nor for the three way interaction (although in this latter case the CrI barely included 0). These estimates are summarized in Table 4.1.
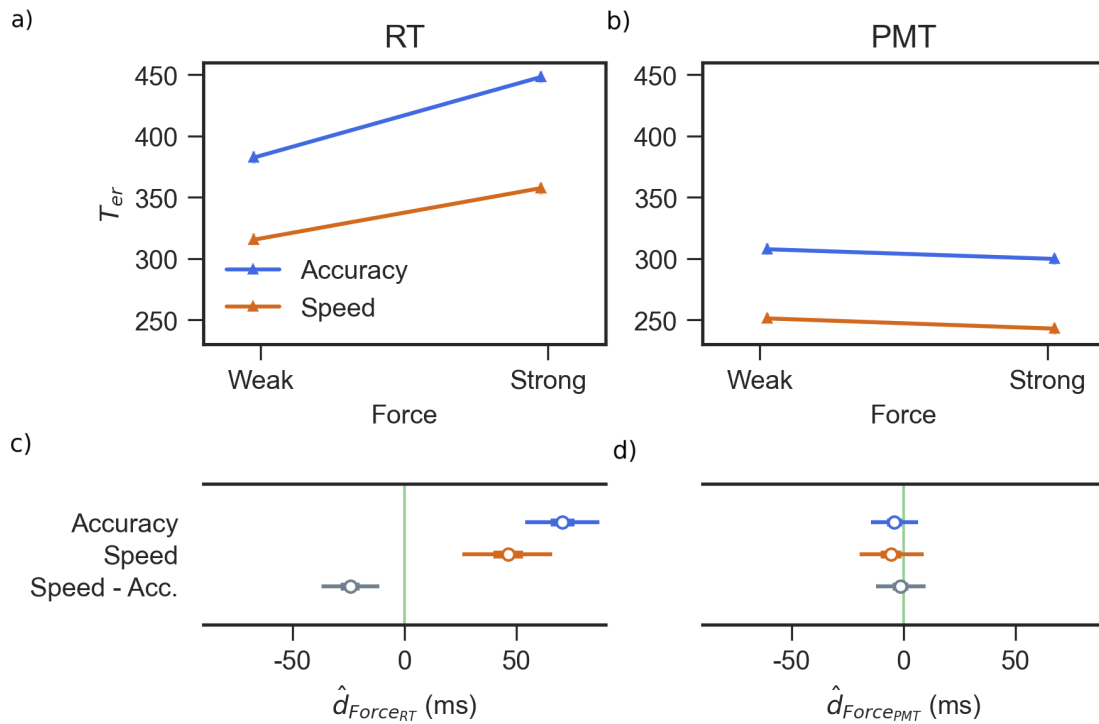
Figure 4.3.: *Effect of Force on the estimated parameter $T_{er}$ estimated on $RTs$ (left panel, a and c) vs. on $PMTs$ (right panel, b and d).*
Upper row: estimated mean values and one SD as shaded lines for a fit on $RT$ (a) and $PMT$ (b).
Bottom row: posterior distribution of the DDM parameter for the effect of Force on $T_{er}$ at each level of SAT, and of its interaction with SAT (see $T_{er}$ SAT × FC in Table 4.1) for a fit on $RT$ (c) and $PMT$ (d).
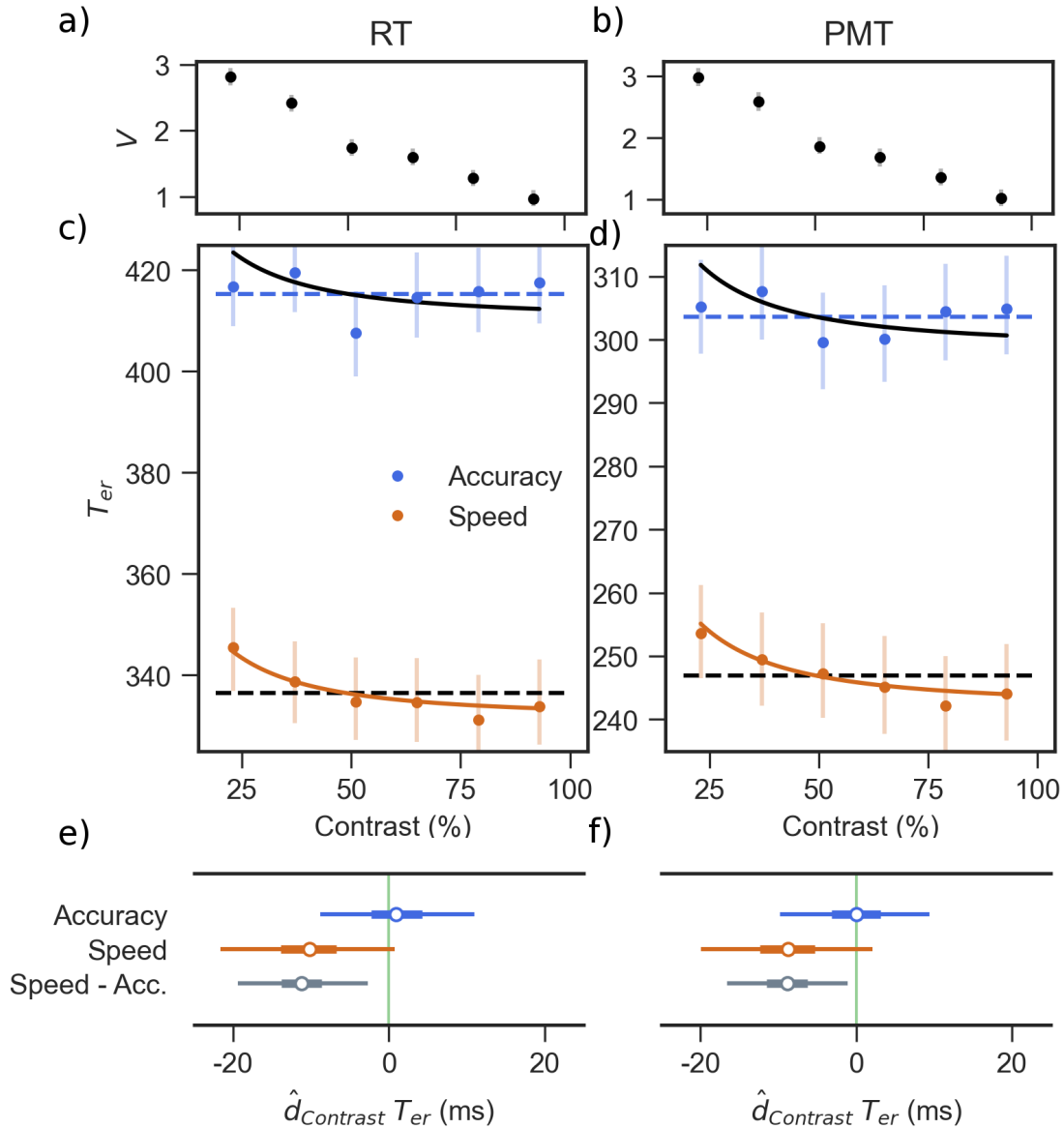
Figure 4.4.: *Effect of Contrast on the parameters of the drift rate (V) and $T_{er}$ estimated with DDM on RT (left column: a, c and e) and PMT (right column: b, d and f) in the speed and accuracy conditions.*

Panels in the top row show the estimated drift ($V$) mean values and one SD as shaded lines along with the contrast levels for a fit on $RT$ (a) and $PMT$ (b).

Panels in the middle show the estimated $T_{er}$ mean values and one SD as shaded lines for a fit on $RT$ (c) and $PMT$ (d). In the General Discussion section we describe external data whose fit is represented by the plain line and compared to the mean $T_{er}$ of each condition represented by the dashed lines. The best fit, as assessed using an $R^2$, between both lines is colored according to the color code of the condition.

Panels in the bottom row show the posterior distribution of the effect of contrast in accuracy and speed along with the difference between both conditions ($T_{er}$ SAT × FC in Table 4.1) for a fit on $RT$ (e) and $PMT$ (f).

### 4.5.1.4. Effects on decision related parameters

The model selection procedure revealed that the drift rate was affected by Contrast (see left column top panel of Figure 4.4 and Table 4.1), but not by SAT nor Force (for details, see model selection results in Table .4). The boundary parameter was affected by SAT, being smaller for the speed condition, and it was not reliably affected by Force in either SAT condition (see left panel in Figure 4.5). The starting point parameter revealed no evidence for an effect of Force (with the provision that the corresponding CrI barely included 0; Table 4.1).
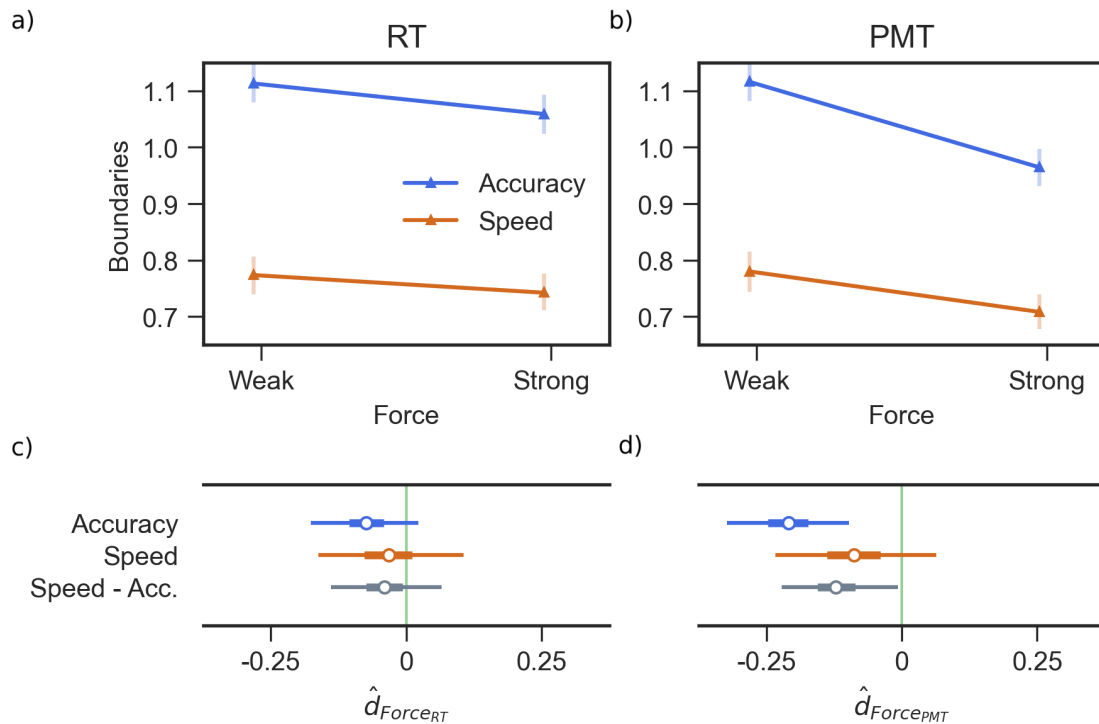


Figure 4.5.: *Effect of Force on the boundary parameter* a *when estimated on RTs (left panel : a and c) vs. on PMTs (right panel: b and d).*
Upper row: mean values and one SD as shaded lines for a fit on *RT* (a) and *PMT* (b).
Bottom row: posterior distribution of DDM regression coefficients for the effect on *a* of Force in accuracy, speed and their difference on either chronometric variable (see Boundaries in Table 4.1) for a fit on *RT* (c) and *PMT* (d).

### 4.5.1.5. Summary and discussion of behavioral observations and their DDM fit

Regarding the targeted encoding processes, performance decreased with increasing contrast. This variation was captured by the model as the predicted negative relation between contrast and $T_{er}$, as well as the predicted negative relation between contrast and drift – the latter resulting in a positive relation between contrast and $T_D$[3]. However, the effect of Contrast on $T_{er}$ was only present when participants were asked to emphasize speed over accuracy. Because $T_{er}$ aggregates $T_{encoding}$ and $T_{response}$, which cannot be estimated separately, a *post-hoc* account of the unexpected absence of Contrast effect in the Accuracy condition would be to hypothesize opposite effects on the two components of $T_{er}$. We show in the next sections on $MT$ and $PMT$ measures that this account is very unlikely.

Regarding the targeted motor processes, $RT$ increased with higher Force demands, while the rate of correct responses was not affected by Force. The variation in $RT$s was captured by the DDM as a rather selective increase in $T_{er}$ that left all other parameters unaffected (with the possible exception of bias). However, there was a discrepancy between the effects of Force estimated on $RT$ and on the $T_{er}$ fitted values. In the $RT$ analysis, the effect size of Force was 48 ms in the accuracy condition and 35 ms in the speed condition (Force did not interact with SAT); in the $T_{er}$ fitted values, Force interacted with SAT and its values for the Accuracy and Speed conditions were 71 ms and 46 ms, respectively. In the next section, we quantify factor effects on $MT$ and we build on our assumption linking $MT$ to $T_{response}$ to assess whether the more faithful capture of motor processes comes from linear regression or DDM fitting.

Regarding the parameters affecting the decision time $T_D$, the observed effects were exactly as predicted: the drift was selectively affected by the Contrast manipulation and the boundary was selectively affected by the SAT manipulation.

## 4.5.2. Motor times ($MT$)

### 4.5.2.1. Linear mixed model

The following observations are illustrated in Figure 4.1 and in Figure 4.2 (third columns). The SAT condition "speed" reduced $MT$, $\hat{d}_{SAT} = -24$, $CrI = [-30, -17]$. A higher force requirement increased $MT$, both in the accuracy, $\hat{d}_{FC:Acc.} = 93$, $CrI = [75, 110]$, and the speed conditions, $\hat{d}_{FC:Speed} = 59$, $CrI = [48, 71]$. The interaction term confirmed that the effect of Force was indeed smaller when speed was emphasized $\hat{d}_{FC:Speed-Acc.} = -34$, $CrI = [-47, -20]$.

Contrast and its interaction with Force had no reliable effect on MT across any of the SAT conditions, as indicated by coefficients restricted to low effect sizes and $CrI$s

---

[3]Note that when speed is emphasized the effect of contrast is reduced on $RT$. This reduction is not surprising as a DDM predicts that a factor influencing the drift rate will have its effect on mean RT scaled according to the height of the boundaries (e.g. see Figure 2 of E. J. Wagenmakers, Grasman, & Molenaar, 2005)

| DDM Par. | Factor | RT | | | PMT | | |
|---|---|---|---|---|---|---|---|
| | | $\hat{d}_{RT}$ | $2.5\%_{RT}$ | $97.5\%_{RT}$ | $\hat{d}_{PMT}$ | $2.5\%_{PMT}$ | $97.5\%_{PMT}$ |
| Drift | (Intercept) | 1.79 | 2.03 | 1.51 | 1.91 | 2.19 | 1.62 |
| | Contrast | -1.77 | -1.46 | -2.08 | -1.90 | -1.55 | -2.23 |
| Boundaries | (Intercept) | 1.08 | 1.02 | 1.15 | 1.05 | 0.99 | 1.10 |
| | SAT | -0.33 | -0.40 | -0.25 | -0.30 | -0.38 | -0.22 |
| | FC:Acc. | -0.06 | -0.14 | 0.02 | -0.17 | -0.26 | -0.08 |
| | FC:Speed | -0.03 | -0.13 | 0.08 | -0.07 | -0.19 | 0.05 |
| | FC:Speed-Acc. | 0.03 | -0.05 | 0.11 | 0.10 | 0.01 | 0.18 |
| Bias | (Intercept) | 0.50 | 0.48 | 0.53 | 0.49 | 0.47 | 0.51 |
| | FC | 0.004 | -0.001 | 0.01 | 0.002 | -0.002 | 0.006 |
| $T_{er}$ (ms) | (Intercept) | 414.6 | 391.0 | 439.1 | 301.2 | 280.0 | 322.8 |
| | SAT | -77.7 | -106.1 | -49.0 | -51.5 | -77.7 | -22.7 |
| | FC:Acc. | 70.6 | 54.1 | 87.2 | -4.2 | -14.6 | 6.6 |
| | FC:Speed | 46.3 | 26.0 | 66.2 | -5.6 | -19.7 | 9.1 |
| | FC:Speed-Acc. | -24.3 | -37.2 | -11.3 | -1.4 | -12.4 | 9.8 |
| | Contrast:Acc. | 1.0 | -8.8 | 11.0 | 0.0 | -9.7 | 9.4 |
| | Contrast:Speed | -10.2 | -21.6 | 0.7 | -8.9 | -20.0 | 2.0 |
| | Contrast:Speed-Acc. | -11.2 | -19.3 | -2.7 | -8.8 | -16.6 | -1.1 |
| | FC × Contrast:Acc. | -6.0 | -16.0 | 4.1 | -2.5 | -11.9 | 6.8 |
| | FC × Contrast:Speed | 5.5 | 3.6 | 14.7 | 6.9 | -1.6 | 15.5 |
| | FC × Contrast:Speed-Acc. | 11.5 | -1.3 | 24.7 | 9.3 | -4.1 | 21.5 |

Table 4.1.: Comparison of Estimated differences between conditions levels ($\hat{d}$) across fits on *RT* and on *PMT*. Columns labelled $\hat{d}$ refer to the maxima a posteriori from the corresponding marginal posterior distribution. Columns labelled 2.5% and 97.5% refer to the CrI intervals.

containing 0.

### 4.5.2.2. Summary and discussion of $MT$ results

We found no evidence for an effect of contrast on $MT$ nor of its interaction with SAT. In the section on behavioral results ($RT$ and error rates), we reported an interaction between Contrast and SAT, and speculated that Contrast may have opposite effects on $T_{encoding}$ and $T_{response}$, the two components of $T_{er}$, in the Accuracy condition. The current analysis of $MT$ shows that this hypothesis is implausible. One alternative possibility is that response mechanisms in the Speed and Accuracy conditions are different enough that they are differently sensitive to the encoding of contrast. We come back to this issue in the General Discussion.

The expected effect of Force on $MT$ was clear, but its size (93 ms and 60 ms in the Accuracy and Speed conditions, respectively) was much larger than that observed on $RT$ (48 ms and 35 ms). Because each $RT$ is the sum of its corresponding $MT$ and $PMT$, this discrepancy can only be explained by an opposite effect of Force on $PMT$. We pursue this issue in the next section on $PMT$.

Another remarkable observation was that the effect size of Force on $MT$ (93 ms and 60 ms in the Accuracy and Speed conditions, respectively) was higher than the effect size estimated in the previous section on $T_{er}$ (71 and 47 ms). Our assumption linking $MT$ to the $T_{response}$ component of $T_{er}$ invites, again, a tentative compensation account in which the other component of $T_{er}$, namely $T_{encoding}$, would be sensitive to Force in the opposite direction. While this hypothesis may seem counter-intuitive, it can be directly tested in our framework by fitting $DDM$ to $PMT$ distributions, as we do in the next section.

## 4.5.3. Pre-Motor times ($PMT$)

### 4.5.3.1. Linear Mixed Model

The results for $PMT$ were very similar to those reported above for $RT$ with the important exception of the effects of Force. This can be appreciated by comparing the second and fourth columns in Figure 4.1 and Figure 4.2. The effect of Force on $PMT$ was opposite to that observed on $RT$ both in accuracy, $\hat{d}_{FC:Acc} = -53$, $CrI = [-79, -27]$ and speed conditions, $\hat{d}_{FC:Speed} = -27$, $CrI = [-46, -8]$. Here the effects of Force were reliably different across SAT conditions, $\hat{d}_{FC:Speed-Acc} = 26$, $CrI = [2, 50]$.

### 4.5.3.2. Drift Diffusion Model selection

The model selection procedure applied to $PMT$ selected the same model structure that was selected when the procedure was applied to $RT$, namely M13 in Table .4.

### 4.5.3.3. Effects on $T_{er}$

The patterns of effects were similar for both $PMT$ and $RT$ fits, with the following important exceptions (compare the two panels on Table 4.1). The main difference is that neither Force nor its interaction with SAT appeared to affect the $T_{er}$ estimated on $PMT$ (Figure 4.3 right panel). In addition, we highlight in Figure 4.4 that the effect of contrast on $T_{er}$ interacted with SAT. The effect was in the expected direction in the Speed condition (although the CrI included 0) but centered on 0 in the Accuracy condition.

### 4.5.3.4. Effects on decision related parameters

The boundary parameter was affected by Force and this effect interacted with SAT. Increasing Force resulted in a lower boundary parameter (therefore an opposite effect than the one observed on $RT$), an effect that was much reduced (if not absent) in the Speed condition (Figure 4.5 right panel). The other decision parameters, drift and bias, were roughly similar when derived from fits on $PMT$ and $RT$ (compare the two panels on Table 4.1).

### 4.5.3.5. Summary and discussion of $PMT$ observations and their DDM fit.

$PMT$ does not include $MT$, which we hypothesized to be strongly linked to $T_{response}$. Therefore, in this analysis, $T_{er}$ provides a reasonable estimate of $T_{encoding}$. The effects of Contrast were consistent with those observed on $RT$, revealing "opposite" effects on encoding processes and drift, including the fact that Contrast interacts with SAT instructions showing that the expected negative effect is present in the Speed condition only (Figure 4.4).

Conversely, the effect of Force was remarkably different for $RT$ and for $PMT$. The linear models revealed Force effects of opposite signs on $RT$ and $PMT$, whereby $PMT$s were shorter with stronger force (Figure 4.1). The DDM fit of $PMT$ did not attribute this effect on $T_{er}$ (Figure 4.3). This absence of effect is consistent with the assumption that, here, $T_{er}$ indexes force-independent encoding processes (i.e. $T_{encoding}$). It undermines the tentative hypothesis (from the section on $MT$) that opposite effects of Force compensate one another on the two components of $T_{er}$, $T_{encoding}$ and $T_{response}$. Instead, the DDM fit attributed the effect of Force to the boundary parameter. Boundary decreased with increasing Force requirements, in the Accuracy condition only (Figure 4.5). This means that the DDM fits on $RT$ and on $PMT$ lead to different attributions of the Force effect, particularly in the Accuracy condition suggesting that EMG and DDMs decomposition differ on the locus of the force effect.

### 4.5.4. Summary of summaries

First, SAT modulations of RT and proportion correct were associated with a change in both the boundaries and the $T_{er}$ of the DDM. Second, the contrast effect was captured simultaneously as a decrease in the drift rate and the encoding processes in the speed condition. However the effect of contrast on these encoding processes was absent in the accuracy condition.

Thirdly, the Force factor resulted in opposite effects on $RT$ and $PMT$. Fitting a DDM on both variables yielded inconsistent parameter attributions of the Force effect. A DDM fit on $RT$ shows a selective effect on $T_{er}$ in both SAT conditions, while a fit on $PMT$, hence without motor processes captured by $MT$, shows a selective effect on the boundaries mainly in the accuracy condition. The latter observation predicts more errors from the participants when a high force is required. This is clearly not found when analyzing the number of correct response. Lastly the Force effect proved to be larger on $MT$ than on $T_{er}$.

## 4.6. General Discussion

### 4.6.1. Decisional and motor processes

We found that $T_{er}$ depends on the force required to respond. Since motor processes are clearly impacted by this manipulation, this result might be due either to the fact that $T_{er}$ embeds motor processes, or to the fact that the response force influence other cognitive processes embedded in $T_{er}$. However, this last explanation can be rejected, as we found no effect of force on $T_{encoding}$ ($T_{er}$ from the DDM fitted on $PMT$). Moreover, we found a similar interaction effect between force and SAT factors on $T_{er}$ and $MT$, while no interaction effect was found on $T_{encoding}$. All together these results support the idea that $T_{er}$ actually captures motor processes.

The force effect estimated on the motor time extracted through EMG was higher than the one estimated on $T_{er}$. Additionally the time preceding motor onset, $PMT$, proved to be shorter when a high force was required. Burle, Possamaï, Vidal, Bonnet, and Hasbroucq, 2002 also observed this result on $PMT$ and interpreted it as an increased caution when response is easily executed. At first sight, fitting a DDM on $RT$ seems to contradict this hypothesis (although a tendency for a boundary decrease with an increase in the force is observed). However, fitting a DDM on the $PMT$ does support this view, and clearly shows that the boundaries decreases as force increases.

The conundrum is then as follows, EMG and the DDM when fitted on $PMT$ suggest an increased response caution when the force to produce is low, but the analysis of the performance of the participants show that they do not make less errors in that case (if anything they make more errors, see Figure 4.1 and the discrepancy between $PMT$ model fit and data in Figure .6 of Appendix C). The only explanation therefore seems to be that participants continue to sample evidence while the motor plan is ongoing[4],

---

[4]Also congruent with the presence of covert EMG activities

allowing them to be equally correct between force conditions and explaining why boundary values are inconsistent between a fit on $RT$ and a fit on $PMT$. Note that all these results are mainly found in the accuracy condition. This interaction with SAT levels could point to a strategical adjustments of the participant or to a failure to reliably detect these effects when $PMT$ and the boundaries are at extremely low values (e.g. floor effect or low power due to low effect size).

Interestingly if we compare values of the boundaries estimated by the DDM between a fit on $RT$ and on $PMT$ we observe that boundary values are mainly changed in the high force condition (see figure 4.5) this suggests that not much, if any, decision is contained in the weak force condition and that participants are mainly continuing to sample when a high force is required.

To summarize, our results show that $T_{er}$ reflects changes in motor processes, as postulated by DDM. However they also suggest that some decisional processes occurs during the motor execution as measured with EMG. Hence, while the motor part of $T_{er}$ is definitely contained in $MT$, $MT$ is not fully contained in $T_{er}$. This poses a problem as if motor processes included in $MT$ (e.g. bringing the applied pressure or force close to response triggering) happen simultaneously to the decision, a model such as the DDM assuming serial execution will necessarily underestimate either the decision time or the motor part of the non-decision time. Note that this does not imply that $MT$ contains discussion, see Section 6.2.2.1.

## 4.6.2. Encoding processes

We found, as expected, a negative relation between contrast and $T_{encoding}$. However, this is no longer true when accuracy is emphasized. In other words, $T_{encoding}$ seems to fail to reflect encoding processes when participants favor accuracy over speed. An alternative interpretation could be that contrast has genuinely no effect on $T_{encoding}$ in Accuracy. This last interpretation would be surprising and clearly invalidate the selective influence of SAT on decision thresholds but the next Chapter tests this hypothesis.

We undertook additional analysis to better understand this unexpected result. To this end we used the data set collected by Reynaud, Masson, and Chavane, 2012 on the latency of visual neurons in awake monkeys responding to stimuli similar to those used in the present study (see Appendix C.3). These data give precise predictions on the onset of activity of neurons in cortex area V1 (the minimal definition we could make of encoding time) as a function of the contrast of the presented Gabor patch. We extracted these predictions for the contrast levels used in our study. As the predicted latencies are faster than the one we observed, we centered them by subtracting their mean and adding the mean of the $T_{er}$ or $T_{encoding}$ computed over all contrast levels and participants. We then calculated a $R^2$ between the predictions (non-linear lines in Figure 4.4 or see Figure .8 in Appendix C.3 for a close-up) and the actual maximum a posteriori estimates of the population $T_{encoding}$ for each SAT and force conditions.

When computing the $R^2$ for the condition where speed is emphasized and force

required is weak we observe an almost perfect adjustment ($R^2 = 0.99$) between the predictions from V1 and the $T_{encoding}$. When speed is emphasized but force is high we observe a lower but still satisfactory agreement ($R^2 = 0.63$). However this is no longer true when considering conditions were the accuracy was emphasized, both in the low force condition ($R^2 = -0.71$) and in the high force condition ($R^2 = -1.14$)[5].

As the adequacy between monkey V1 neurons and DDMs $T_{encoding}$ was not predicted *a priori*, we sought to replicate this finding with the data from the next chapter. We therefore performed the same calculations but on an independent data set collected for a follow-up study using electro-encephalography (EEG) and the same design with 20 new participants. The only change between the original and replication study was that contrast levels were sub-sampled to three of the levels used in the study for practical purpose on the EEG processing. We observed the same pattern of results with a high agreement between V1 neurons predictions and the DDM estimates in the speed condition (low force:$R^2 = 0.93$; high force : $R^2 = 0.85$), but not in the accuracy condition (low force : $R^2 = 0.52$; high force : $R^2 = -5.04$)[6].

To summarize, we observed the predicted relation between $T_{encoding}$ and stimuli contrast in the speed condition. Moreover, secondary analysis show that the pattern of this relation is very close to that observed between stimuli contrast and the activity of neurons in V1 in monkeys. At first sight, these results suggest that $T_{encoding}$ actually reflects encoding processes. However, none of these results hold when participants are asked to favor accuracy over speed.

Two interpretations are at hand. The DDM could simply fail to achieve a good decomposition between decision and non-decision times when participants favor accuracy, which would question the validity of the inference made from the DDM parameters in a canonical SAT instruction. Alternatively, it might be the case that the effect of contrast on encoding processes is changed by SAT instructions. Put differently, human participants achieve SAT by other means than a parametric modulation within the DDM as previously suggested by some neuro-physiological (see Heitz, 2014, for a review on the neurophysiological evidence) and computational (Rafiei & Rahnev, 2020; Verdonck, Loossens, & Philiastides, 2020) studies.

### 4.6.3. Effect of SAT on non-decision processes

The $T_{er}$ proved to be sensitive to SAT, once we account for the motor processes by removing the $MT$s we observe that the putative $T_{encoding}$ estimate of the DDM is still sensitive to SAT instruction with a difference between speed minus accuracy level of -50ms. This effect represents approximately 1/3 of the effect found on $RT$. Therefore according to the DDMs decomposition 1/3 of the effect of SAT on $RT$ is to

---

[5]Note that when computing the $R^2$ on $T_{er}$ estimated from the RT we find the same result : $R^2 = 0.91$ and $R^2 = 0.88$ in speed, low and high force respectively *vs.* $R^2 = -1.11$ and $R^2 = -0.02$ in accuracy, low and high force.

[6]And same on $T_{er}$ from a DDM fitted on RT : $R^2 = 0.92$ and $R^2 = 0.53$ in speed, low and high force respectively *vs.* $R^2 = 0.44$ and $R^2 = 0.21$ in accuracy, low and high force.

be attributed to non-decision components happening before muscles initiation as recorded with EMG. Following the logic of the chapter this would say that the latency to encode the stimuli is reduced of 50ms when participants are asked to speed their response. This either sheds a novel light on the locus of the SAT effect or once again shows that the DDM mis-attributes what is decisional *vs.* non-decisional.

## 4.7. Conclusion

In conclusion, our prediction, on the negative relationship between contrast and encoding processes, and external physiological data suggest that the non-decision time as estimated by a decision making model such as the DDM contains stimulus encoding processes when the participants are asked to favor speed over accuracy. However these observations do not hold when participants are asked to favor accuracy. The nature of this discrepancy remains to be clarified in the next chapter where we use a single trial EEG estimate of encoding time to observe 1) whether the contrast has the expected effect in both SAT conditions 2) whether these estimates account for the 50 ms difference in SAT levels observed on $T_{encoding}$ and 3) whether this single trial estimate covaries with the estimation made by the DDM.

About motor processes while we do observe that the DDM interprets the Force manipulation to be specifically of motor origin, the EMG data shows that pre-motor processes are also impacted by the Force manipulation but in an opposite direction. Fitting a DDM on these pre-motor processes suggests that participants modulate their caution of response according to the force required to trigger a response but therefore wrongly predicts the amount of correct responses in each force level. This leads to the conclusion that, at least given a low speed pressure and a long force execution time, some portion of decision is happening during the initiation and the execution of the response. This leads necessarily to an underestimation either of the decision time or the non-decision time by serial models such as the DDM.

# 5. Testing the recovery of processing stages using electrophysiology

## Sommaire

# 5.1. Note

# 5.2. Introduction

The previous chapter suggests that participants are deciding while they are executing the response instead of deciding prior to the engagement of the muscles as often assumed using serial models of decision making. It is important to note that such results has already been observed by previous studies (Buc Calderon, Verguts, & Gevers, 2015; Dotan, Meyniel, & Dehaene, 2018; Resulaj, Kiani, Wolpert, & Shadlen, 2009; Selen, Shadlen, & Wolpert, 2012) but these studies all used reaching movements and the same effector muscles. As already argued in the general discussion of Chapter 2 these settings do not represent the vast majority of decision making tasks on which most inferences from EAMs are made. Additionally and as fueled by the observation of the previous chapter, these settings are prone to strategic adjustments (e.g. using intermediate movement could be optimal for the decision maker, see Haith, Huberdeau, & Krakauer, 2015) making unclear how these results generalize to standard settings.

Therefore, testing whether deviations from a serial decision making model in a standard setting has some actual consequences on the estimation of cognitive processes remains to be shown.

The previous chapter also casts doubt on the nature of the encoding processes estimated by a decision making model such as the DDM (Ratcliff, 1978). Either the genuine encoding processes are highly changed between SAT conditions, both in terms of magnitude and of sensitivity to the contrast experimental factor, or the DDM mis-attributes decisional and non-decisional processes. In any case either the parametric modulation of decision thresholds as an explanation of SAT is insufficient or the DDM does not achieve a satisfactory decomposition between decision and non-decision times, questioning the interpretations made from its parameters. In order to test the extent of the discrepancy between predictions and model, we seek to replicate the results of the previous chapter while adding a single trial electrophysiologically derived measure of the encoding time.

## 5.2.1. Definition of the encoding time

At first, for the purpose of the present chapter we have to define what is assumed as being the encoding time. We start as before by stating that $T_{er}$ is, as in Equation 1.8, the sum of pre and post evidence accumulation non-decision times (which we assume to be encoding and motor processes). However the definition can be further refined.

For example, the non-motor related part of the non decision time has been defined by Luce, 1986 as :

> 1) The time required for the physical signal to be transduced from physical energy into the neural spike trains, which appears to be the information code of the nervous system; 2) the transit time for such pulses to pass from the output of the transducer to that part of the brain where the information arriving on the entire fiber bundle (the optic and auditory nerves being the most thoroughly studied) is processed and interpreted[...]

This definition owe to be the strict minimal definition we can make of the pre-evidence accumulation non-decision times, a translation into a neural code and a conduction to the "decision center"(Luce, 1986) in the brain. But other definitions can also be made. In his seminal paper written in the context of a model for memory retrieval, Ratcliff, 1978 defines the non-motor part of the non-decision time as "*probe encoding, preparation for comparison and decision processes*". Finally another quote from Ratcliff, Smith, Brown, and Mckoon, 2016 define this time as "*extracting the dimension(s) of the stimulus that form the basis of the decision from the stimulus or memory*".

Based on these definitions it appears clearly that the non-motor part of the non-decision time is the duration from stimulus onset to the time at which the evidence starts entering the decision process. However this time will obviously be dependent both on the stimuli and the task, E.-J. Wagenmakers, 2009 raises the following example :

> For instance, suppose people have to judge whether a word represents an object that is bigger or smaller than a television. Also, suppose that each decision requires that people first construct a mental image of a television, then construct the mental image associated with the presented word, and finally engage in some sort of comparison process. The time that is taken up by the process of constructing the mental image of a television does not depend on the nature of the imperative stimulus, and is therefore clearly part of Ter [non-decision time].

Therefore encoding time can be defined as solely the time of the physical transduction of the presented stimuli (in this case the physical stimulus of the presented word) or the whole time preceding the evidence accumulation. Based on this overview, the definition one can make about encoding time seems rather imprecise and tightly linked with the task at hand and how the decision process is initiated.

### 5.2.2. Interdependence between encoding and decision times

Having stated encoding time as pre-evidence accumulation time, one already has implied a serial dependency between encoding and decision times, i.e the decision stage is assumed to start after completion of the encoding stage. This assumption is also made by the diffusion models used in the present thesis (Ratcliff & Smith, 2010). But alternative mechanism could take place to encode the stimuli and accumulate the evidence from it. For example in the cascade model suggested by McClelland, 1979 it is suggested that processes can feed-forward information to the following process at any stage of completion. In such case the encoding and decision processes do not operate serially but in parallel, the decision process will evolve in time with the variation in the encoding processes amongst other sources of noise. If the aforementioned mechanism (or any other non-serial mechanism) is the genuine link between decision and encoding times in a considered task, a serial model such as the DDM will fail to capture the respective encoding and decision stages. In their collaborative study, Dutilh, Annis, Brown, et al., 2019 have suggested that non-decision time of EAMs such as the DDM pervasively lead to failure of selective influence. Smith and Lilburn, 2020b interpreted this failure as resulting from the unusual perceptual characteristics of the task used for the selective influence test, a random dot kinetogram. They provide support for their interpretation by fitting a model with a dynamic evolving accumulation rate rather than a constant one as assumed by most EAMs. The authors end up suggesting that studies probing the conformity of their predictions with model parameters should account for the visual properties of the evidence entering the decision process.

Throughout this thesis, except for Chapter 3, we used sinusoidal gratings that the participants need to compare in order to decide. Hence the encoding time in our experiments has to be the time at which the relevant dimension of the contrast of both Gabor patches is entirely perceived for the difference to be accumulated. We therefore think that our design is appropriate for a serial model such as the DDM as we assume

the encoding time is a process happening serially (as any premature accumulation of the difference before completion should be driven by noise) and linked to the visual perception of the gratings.

### 5.2.3. Expected latency of the encoding time

In order to extract a physiological marker of the above defined encoding time we have to provide some estimate about the duration of that putative latency. Recall from Chapter 2 that removing the EMG derived motor execution time ($MT$) from the non decision time estimated by a Drift Diffusion Model (DDM) suggested a residual time of approximately 200 ms. In Chapter 4 the intercept of the non-decision time fitted on the pre-motor time ($PMT$) was of 301ms when participants emphasized accuracy and 250ms when they emphasized speed. Hence if we once again assume that the estimation of non-decision processes, once we account for the motor execution, is the time needed for encoding we expect it to be around 200 and 300ms.

Previous studies also suggest a latency with a duration close to the one we found. Vogel, Woodman, and Luck, 2006 for example provide some evidence about the time needed to manipulate visual information. Using a memory array task in which participants were asked to compare a prior screen, replaced by a mask, to a following test array they showed that visual information storage reached asymptotic capacities when the mask was displayed later than 200 to 300ms. In a different task, Strijkers, Costa, and Thierry, 2009 have shown that lexical access takes place around 180ms, suggesting that all visual information processing needed to access the meaning of a word is achieved at that time. In another different task and even a different species, Roitman and Shadlen, 2002 have found using single cell recordings in awake monkeys that accumulation like activity in the lateral intra-parietal area, started 200 ms after onset of a random dot kinetogram.

All these evidences suggest that some operations on a visual input is necessary prior to the decision processes on a visual stimulus and that these operations take a time around 200ms. We therefore assume that the encoding time we are looking for should be close to 200ms.

### 5.2.4. Linking propositions : $T_e$, $T_r$ and $T_D$

To summarize, when looking for a physiological candidate for the encoding time in our experiment we are looking for a physiological marker happening around 200 ms and linked to the visual perception of the sinusoidal gratings. In addition, recall from Chapter 4 that we observed a surprisingly high agreement between the estimates of encoding time from the DDM and the predictions from the cortical dynamics in V1 in response to sinusoidal gratingsThis does also suggest that we should look for a physiological marker that is linked to visual processing as found in visual cortical areas.

One classical studied EEG component does correspond to the given criteria, a negative ongoing component happening at posterior electrodes and at a time of approximately 200ms, the N200. This component has already been linked to visual information arriving in extrastriate areas (Zhang, Walsh, & Anderson, 2017). Loughnane, Newman, Bellgrove, et al., 2016 showed that an EEG component spatially and temporally consistent with a N200 is consistently associated with the onset of relevant evidence for forthcoming decision even when spatial and temporal certainty was high. Martin, Huxlin, and Kavcic, 2010 compared the properties of this event related potential (ERP) and linked it to non-decision time estimate from the EZ-diffusion (E.-J. Wagenmakers, van der Maas, & Grasman, 2007) and found a positive relationship between the onset of the N200 and the non-decision time.

These studies point to the link between the N200 ERP and the encoding time as defined earlier. Problem is that studying ERPs often comes with the limitations raised in the introduction of the thesis, namely noisy signal and averaging that distort the underlying single trial processes. But Nunez, Gosai, Vandekerckhove, and Srinivasan, 2019 have suggested a way to measure single trial N200 derived from the averaged N200 ERP. Using a singular value decomposition method they extracted the electrodes that most contribute to the trial averaged N200. Afterwards they reconstructed the EEG signal by using the weights of each electrode to the N200 therefore de-noising the single trial EEG signal and allowing to measure the onset and peak of the N200 at the single trial level. Moreover the authors report that the single trial estimate of the peak N200 latency relates both to the visual manipulations they use and to the DDM estimate of non-decision time. Based on these evidences we suggest the following linking proposition : single trial encoding time is indexed by the single trial peak of the N200.

The linking proposition between the response execution time and the motor time measured with EMG has already been spelled out in chapter 2 and is kept in the present chapter. As a logical consequence of the two previous linking propositions we can make a final one : For the present study the time between the N200 single trial peak ($N200$) and the single trial motor time ($MT$) is the single trial decision time ($DT$).

### 5.2.5. Present study

We build on the experimental design of the previous chapter because it has proven to be efficient in breaking up decision and non-decision related effects. Therefore we manipulate force requirements and average contrast of both sinusoidal gratings as in Chapter 4 to the exception that contrast manipulation was restricted to three levels. In addition like before we manipulate SAT instructions because we have shown that different SAT emphasis can change decision and non-decision related processes. In the next part of the section we derive how these factors should affect the measure made through electrophysiology *vs.*the estimation made using a DDM of these processes.

### 5.2.5.1. Predictions on the effect of contrast

As before we expect that increasing contrast will simultaneously increase the decision time (and increase the amount of errors) and decrease the encoding time. Therefore, on the measurements of these times we predict that the contrast factor will have opposite effects on $N200$[1] and $DT$ while sparing $MT$. About the estimation of processes, we expect that the decision time will increase, through a decrease in the drift rate parameter, and that the non-decision time parameter, $T_{er}$ will decrease. The previous chapter showed that the contrast factor interacted with SAT, suggesting no contrast effect in the condition emphasizing accuracy on estimated encoding processes. We expect that this interaction on the estimation made by the DDM will replicate and if estimation is congruent with measurement, that the $N200$ will also present no effect of contrast in the accuracy condition.

Additionally we test our measurements of encoding and decision latencies using the Wagenmakers-Brown law. As described in Section 1.1.2.1, the Wagenmakers-Brown law is the prediction of a linear relationship between mean and standard deviation of the $RT$ whenever task difficulty is varied. As the effect of contrast is supposed to affect encoding and decision times, and given the assumption that the former only changes mean RT, the Wagenmakers-Brown law should hold only on the measured $DT$ and not on the pre-motor time containing both intervals. Therefore we expect that Wagenmakers-Brown law should hold for the effect of contrast on $DT$ but not for the time containing $N200$ and $DT$ (i.e. $PMT$ in the previous chapter).

### 5.2.5.2. Predictions on the effect of force

Based on the results from the previous chapter we expect manipulating force will have opposite effect on $DT$ and $MT$ while sparing $N200$ and the proportion of correct responses. We expect that when force requirements are high $DT$ will be shorted and $MT$ longer than when force requirements are low. About the estimation we predict that only the $T_{er}$ will be impacted by the force manipulation.

### 5.2.5.3. Predictions on the effect of SAT

First as suggested by Steinemann, O'Connell, and Kelly, 2018 using another EEG component and the DDM fit on $PMT$ in the previous chapter, we expect that $N200$ will be impacted by SAT instructions as $MT$ and $DT$.

### 5.2.5.4. Expected co-variance between estimation and measurement

We will end by directly comparing the inter-individual variations for the measured and estimated $DT$ (see method section for how we derive $DT$ from the DDM estimates) and non-decision times $NDT$ ($T_{er}$ for the DDM and the sum of $N200$ and $MT$ for

---

[1]Note that although we have seen in the previous chapter that the effect of contrast might not be linear on the encoding processes we still rely on a linear link as the approximation seems reasonable

the measurement) both for the absolute values and the experimental factor differences. We expect this analysis to further refine the comparison of the locus of the experimental effects.

## 5.3. Methods

### 5.3.1. Participants

Twenty participants (8 men and 12 women, mean age = 23 years, 3 left-handed) that were students in Aix-Marseille University, were recruited for this study. All participants reported having normal or corrected vision, and no neurological disorders. The experiment was approved by the ethical experimental committee of Aix-Marseille University, and by the "Comité de Protection des Personnes Sud Méditerrannée 1"(Approval n°1041). Participants gave their informed written consent, according to the declaration of Helsinki. They were compensated at a rate of 15 € per hour.

### 5.3.2. Experimental procedure

The apparatus, experimental factors, settings, randomization and design were the same as described in Chapter 4. The only differences was 1) that the contrast levels were reduced from six to three (23, 51 and 93% see method of Chapter 4) and 2) the force settings varied every three block rather than every six block. The first modification was intended to increase the number of trials per contrast levels to compensate for the high rejection rate expected with the estimation of single trial N200. The increase in force alternation frequency was chosen to better control for possible confounding effects of fatigue and learning. It was not possible for the previous experiment for technical reasons, that were solved in the current one[2].

### 5.3.3. Electrophysiological recordings

EEG activities were continuously sampled from 64 scalp electrodes with a location conforming to the 10–20 positioning system (Chatrian, Lettich, & Nelson, 1988). The EMG was recorded from two electrodes pasted 2 cm apart on the thenar eminence, above the flexor pollicis brevis. EEG and EMG were recorded with a sampling rate of 1024 Hz using a BioSemi Active II system (BioSemi Instrumentation, Amsterdam, the Netherlands). Electrodes for vertical and horizontal electrooculogram were respectively located at the Fp2 and below the right eye, and near each outer canthus. Electrophysiological data were filtered offline (high-pass = 0.01 Hz). EEG data were first re-referenced to the left mastoid, and EMG electrodes were bipolarly referenced.

---

[2]We thank Didier Louber for allowing to change the force settings by the computer rather than requiring an intervention of the experimenter inside the Faraday cage

## 5.3.4. EEG preprocessing

All preprocessing applied to each participant can be found at https://github.com/ GWeindel/EEG_thesis_chapter/tree/main/preprocessing.

The EEG data was preprocessed using the *mne* python module (Gramfort, Luessi, Larson, et al., 2013). For artifact detection EEG data was first re-referenced to the average of the electrodes. Only one electrode in one participant was detected as faulty because of a flat signal (FT7) and excluded from the data. To remove ocular artifacts as well as contaminant EMG noise we relied on an independent component analysis (ICA) with the *fastica* algorithm implemented in *mne*. Before applying the ICA we first excluded trials where the participant blinked at stimulus onset (rare) or trials with artifacts spreading across all electrodes (e.g. saturation of the amplifier or body and head movements). Additionally we removed breaks between blocks and high-pass filtered the data at 1Hz to avoid edges of the portions removed from the data. The ICA was fitted with 63 components on the 500Hz re-sampled filtered data. The resulting ICA unmixing matrix was then used on the unfiltered epoched data to study the components. Each component was inspected based on the time-course of the reconstructed signal during the epochs (-500 ms to 1000 ms after stimulus onset), the contribution of the electrodes to the component, the power spectrum and variance of the signal and the average ERP of the component. Rejection of the component was decided by the experimenter whenever contributing electrodes were far from the electrodes of interest (mainly the posterior electrodes for the measurement of the *N*200) and the spectrum displayed artifact like activities (e.g. high peak on alpha frequencies, peak at 50 and 100 Hz) or the time-course indicated blink like activity with contributing electrodes next to the eyes. Once component identified as artifacts were isolated, the EEG signal was reconstructed using the un-mixing matrix of the remaining components.

All the preprocessing of the EMG signal was similar to the previous chapters.

## 5.3.5. Single trial N200 estimation

To extract the single trial estimate of the N200 we relied on the method applied by Nunez, Gosai, Vandekerckhove, and Srinivasan, 2019 (see Figure 5.1 for a graphical summary of the method). All estimation for each participant can be found at https://github.com/GWeindel/EEG_thesis_chapter/tree/main/N200_single_trial_ estimate. For each participant the method was as follows : the data was first filtered between 1 and 10Hz to simultaneously attenuate the influence of later slower potentials such as the P300 and high frequency components or artifacts. The data was epoched in a window of 1100ms from -100 to 1000 ms after stimulus onset. The signal of each electrode was then centered by subtracting the mean of the epoch to all the data points present in the window and followed by a resampling of the data from 1024 to 1000 Hz to get a direct correspondence between EEG samples and millisecond units. We computed the ERP by computing the average of the signal across all trials for each

sample using all electrodes. We applied a principal component analysis (PCA) on the centered data using the *sklearn* python module with the singular value decomposition method. We then selected the first component of the PCA (the component that explained most of the variance in the ERP) by previously estimating whether that first component had the expected topography (maximal at posterior electrodes) and a clear negative peak around 200ms [3]. The data was then reconstructed based on the weights of each electrode to the first component therefore creating a matrix with one single channel for each trials. An empirically defined window of 175ms, from 125 to 250ms after stimulus onset, was then used to search for the minimum value of the PCA reconstructed EEG channel inside the defined window for each epoch[4].

## 5.3.6. Statistical analysis of the behavioral and EEG/EMG derived variables

### 5.3.6.1. Estimated difference between condition levels $\hat{d}$

As in the previous chapter we directly estimate the magnitude of the difference ($\hat{d}$, see Section C.1) from the predictions of the fitted linear models (described below).

**N200 analysis**   The effects of experimental factors on the N200 was analyzed at the single trial level only. In order to give a sense of the effects on traditional event related potentials, we display the group averaged mean signal between electrodes PO7 and PO8 across all trials. The LMM applied modeled the single trial N200 as a Gaussian distributed variable. The priors given were chosen based on the N200 search window defined in the previous section. The intercept was given a normal prior with mean 200 and SD 25. The priors for the effects of experimental factors, the random effects and the residual SD were given a normal distribution of mean 0 and SD 25. We chose an SD of 25 for intercept, fixed and random effects because it allows to capture all values given the search window of the by-trial $N200$.

**Decision Time analysis**   The LMM applied on the log-transformed $DT$ (the time left after subtraction of the by-trial $MT$ and $N200$ to the $RT$) modeled the variable as a normal distribution. The priors given were equal to those used on $RT$, $PMT$ and $MT$ in the previous chapter (see Appendix A.2) except the mean of the normally distributed prior for the intercept which was equal to 4.6 (approx. 100 ms) in order to lie closer to the observed distribution.

All other variables, proportion correct, $RT$ and $MT$ were analyzed as in Chapter 4 (including the transformation from coefficients to predicted differences as described

---

[3]For 2 participants ("S1" and "S17") the second principal component presented these features, we therefore chose this component for these two participants to compute the single trials N200

[4]The size of the tested windows was chosen in the ranges suggested by Nunez, Gosai, Vandekerckhove, et al. 2019 and we choose the window for which we had the lowest amount of single trial estimated N200 at the boundaries of the window
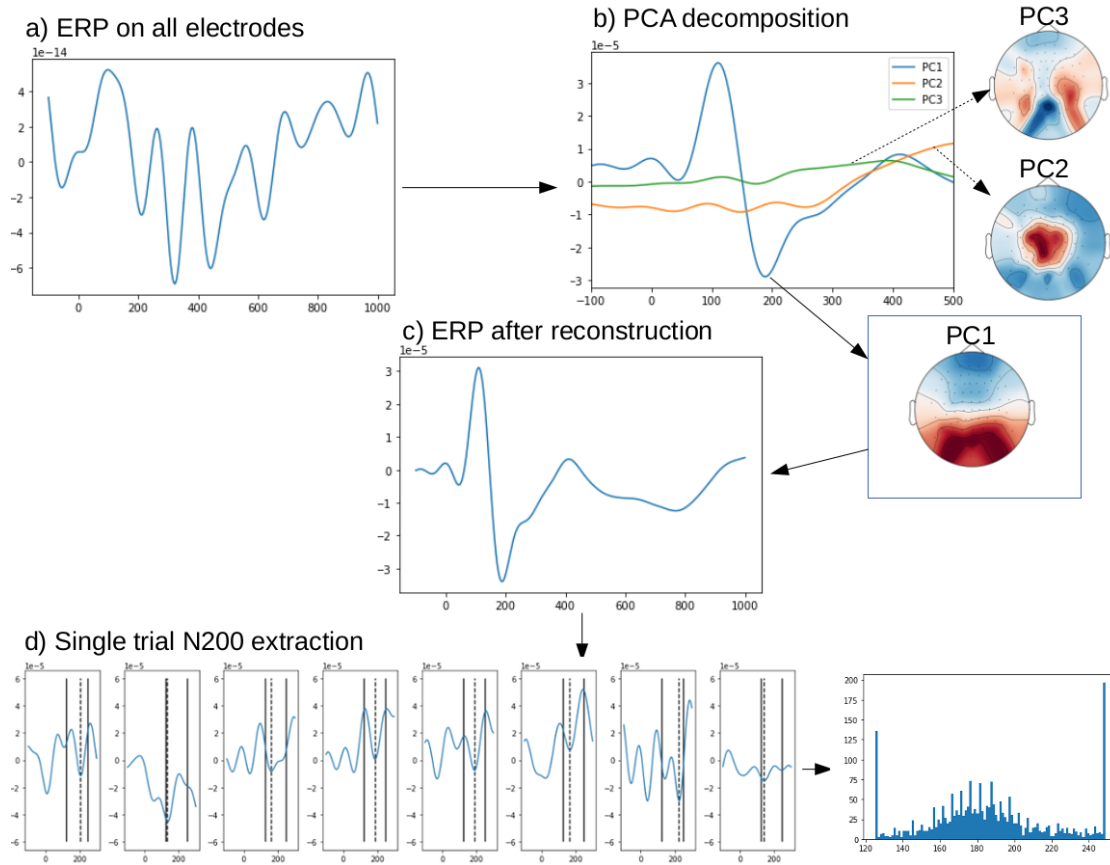
Figure 5.1.: Representation of the method used to extract the single trial *N*200s adapted from Nunez, Gosai, Vandekerckhove, et al. 2019. On all figures the x-axis shows the time from stimulus onset and the y-axis the value of the signal (except for the histogram which shows the frequencies of the corresponding time bin).
a) Single participant ERP computed on all trials across all electrodes for a window between 100 ms before and 1000 ms after stimulus onset.
b) ERP from the three principal components (PC) extracted from the PCA method and the topography of the electrode contributions (or weight matrix) to each PC on the side (blue negative contribution, red positive contribution). PC1 only corresponds to the criterion defined in the method and is therefore selected.
c) Same ERP as in a) expect computed on the reconstructed signal from PC1
d) Example of 8 randomly selected single trial N200 detection, solid line corresponds to the defined search window while the dashed line represents the detected local minimum (N200) of the reconstructed signal. The histogram for all single trial N200 is presented on the right side of the subplot.

in Appendix C.1) with the priors described in Appendix A.2. All estimations of the LMMs were performed with the same code as in the previous chapters, therefore the MCMC estimation was performed using the *pystan* module and based on 6 MCMC chains with 2000 iterations including 1000 warm-up samples.

## 5.3.7. Testing the Wagenmakers-Brown law in the DT

To test whether the Wagenmakers-Brown law holds for $DT$ only, we first computed the ratio between the mean and standard deviation (SD)[5] for each experimental cell and each dependent variable (DV) $DT$ and $PMT$ (*i.e.* $N200 + DT$) :

$$y = \frac{\text{mean}(DV)}{\text{SD}(DV)} \tag{5.1}$$

Using a LMM we then modeled the estimated ratio $y$ as a normally distributed variable with SAT, DV and contrast as predictors. We expect a linear relationship only between DT and contrast". We coded the DV variable as 0 when the ratio is computed on $DT$ and 1 when computed on $PMT$. We therefore expect that the difference ($\hat{d}$) in ratios for the lowest and highest contrast ($\hat{d}_{Contrast}$) should be equal to 0 when computed on $DT$. For $PMT$ we expect that the effect of contrast ($\hat{d}_{Contrast} + \hat{d}_{Contrast \times DV}$) on the ratio should be negative (as the mean should grow less than the SD because of the opposite effect of contrast on encoding times). It should however be noted that the use of only three difficulty levels might be too weak to properly test the Wagenmakers-Brown law (E. J. Wagenmakers, Grasman, & Molenaar, 2005).

## 5.3.8. Drift Diffusion modelling

We estimated the same model as selected in the previous chapter (see Table .4) using the *HDDM* python package with 4 MCMC chains and 15000 iterations including 12500 warm-up samples [6].

To estimate the effects of factors on parameters we again relied on the *HDDM* package and directly estimated a regression DDM with the same characteristic as described in Chapter 4. Note however that, because of several power failures and a difficult maintenance of the computer due to the COVID-19 pandemic, we reduced, compared to Chapter 4, the number of iteration to 4000 including 3000 samples as warm-up samples. This amount of samples does not guarantee well converged MCMC chains for all parameters. We therefore estimated the parameters using 10 MCMC chains, inspected the traces of the MCMC chains for each population parameter

---

[5]Note that usually the ratio is computed as $SD/mean$, however applying a LMM on this ratio proved to break the assumptions of normality and homoscedasticity from linear models. These deviations were not observed when we computed the ratio $mean/SD$, we therefore choose to retain this last computation.

[6]We reduced the number of iterations compared to Chapter 4 to reduce computation time and because 15000 iterations were enough for well converged MCMC chains

and dropped the chains that presented evidence for severe non-convergence (N = 5). Despite such efforts, the traces for the inter-trial variability parameters appeared non-converged but lied at reasonable values. Out of conscientiousness, the present analysis should be reproduced with a higher number of iterations. The results are however not expected to change based on an a posteriori comparison of the results with the previous chapter and on a frequentist estimation of the parameters (using a dual-stage approach as in Chapter 3, estimation using *fast-dm* followed by a LMM on the point estimates).

## 5.3.9. Estimation of the Decision Time of the DDM

To estimate the mean decision time for each condition we relied on the *rtdists* R package (Singmann, Brown, Gretton, et al., 2016). For each recorded MCMC iteration of the DDM model, we computed the mean predicted reaction time for the set of parameters except the non decision time and its inter-trial variability parameter therefore only estimating the decision time. We then obtain predictions for the mean decision time while keeping the uncertainty associated with the estimation of the parameter.

## 5.3.10. Correlation between RT stages estimated from the DDM and electrophysiological recordings

To estimate the inter-individual correlation between the mean decision and non-decision stages measured (respectively the $DT$ and the addition of $N200$ and $MT$) and estimated by the DDM (respectively the decision time described above and the $T_{er}$), we relied on the plausible values method developed by Ly, Boehm, Heathcote, et al., 2017 and described in Section 2.6.1.3. As we are also interested in the correlations of the experimental factor effects we computed for each experimental factor the differences between the levels. The SAT difference was computed as Accuracy minus speed, the force difference as high force minus low force and the contrast difference as highest contrast minus lowest contrast.

About these correlations it should be noted that the sample size of participants is rather low to compute robust inter-individual correlations, however this should be reflected in the CrI of the plausible value distributions. In addition, the summation of $N200$ and $MT$ is problematic as we are also summing measurement errors and therefore likely increasing the variance of the measured NDT. Despite these limitations, this method appeared to be the best strategy given the time left to finish the manuscript and the computing problems raised in Section 5.3.8.

## 5.4. Results

One participant was discarded prior to the analysis because the analysis of the force signal clearly showed that the change in force settings failed for this participant.

As in the previous chapters we excluded $RT$s higher than 1500 ms (1% of the data). The trials with multiple activities as well as trials with EMG artifacts were removed (respectively 16% and 3.8% of the data). In addition to these criterion we removed trials where N200 was estimated at the boundaries of the search windows (e.g. see first and last bar in histogram of Figure 5.1) and trials with EEG artifacts (respectively 22.6% and 6% of the data). Finally we removed all trials for which the physiological $DT$ was estimated to be negative or null (4.1% of the data) as these trials were clearly indicative of either a misestimation of the DT or a guessing strategy by the participant [7]. All combined rejection criteria lead to the rejection of 46.7% of the data, leaving on average 1330 trials per participant (SD = 182) for the analysis. As in Chapter 4 the LMM on $RT$, $N200$, $DT$ and $MT$ were only performed on correct responses (80% of the remaining trials). Given the unusual high rejection rate we replicated the LMM on $RT$ and proportion correct to ensure that the rejection criteria did not change the effects of the experimental factors on these measures (see Appendix D.1).

As in Chapter 4 for all gLMMs applied we provided the average values for each condition along with the fits (Figure 5.2) as well as the values of the predicted differences between each conditions (Figure 5.3). All estimated differences are reported in the natural scale (proportion correct and milliseconds).

Finally note that precisely describing the decomposed latency is not the scope of the study as we are interested in describing the locus of experimental effects. However we do provide a quick overview on the relationship of these latencies to the RT in Appendix D.4.

---

[7] observed accuracy for these trials was equal to 55%
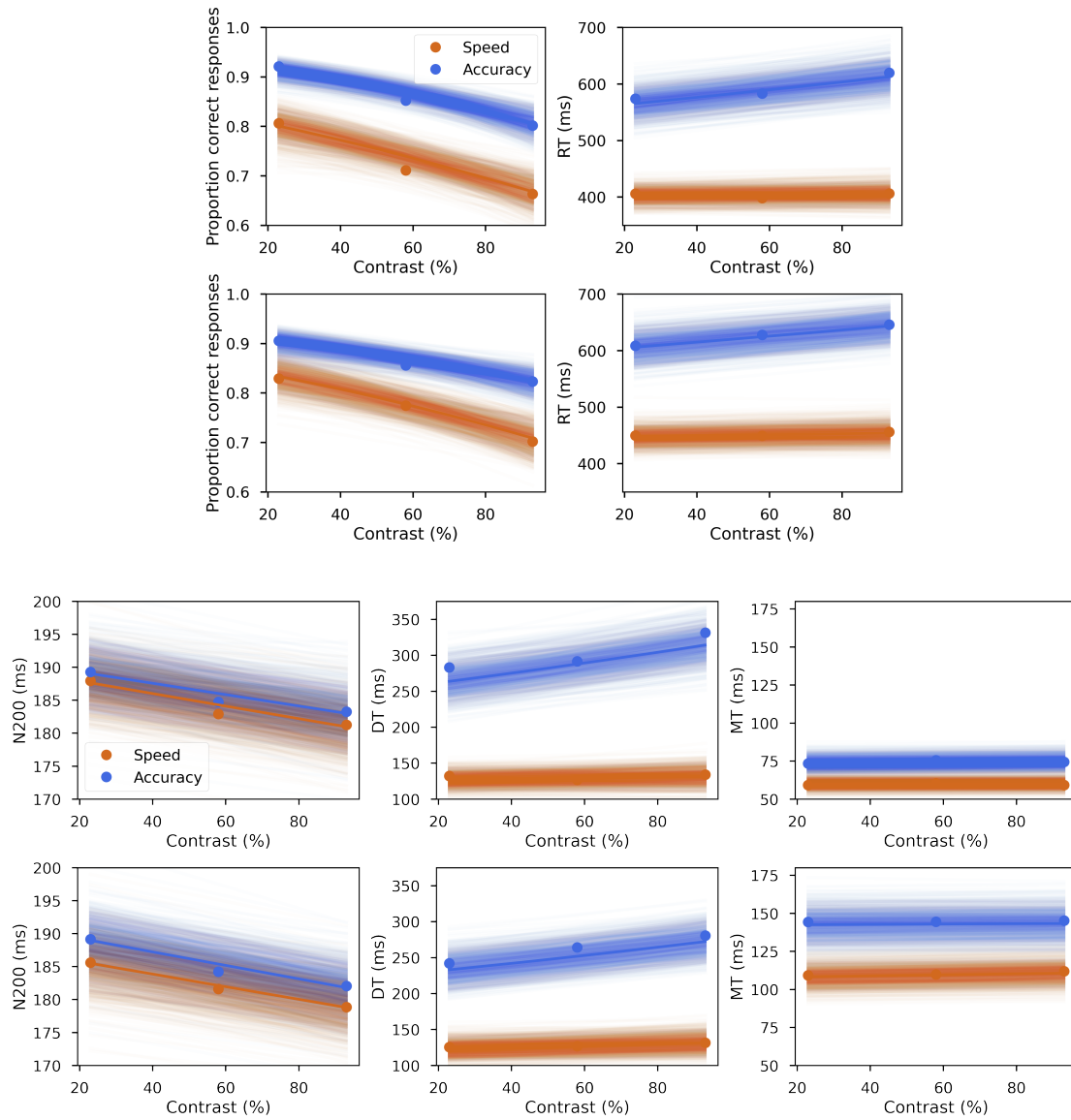
Figure 5.2.: Top : Average values for proportion correct (left) and *RT* (right) plotted for each Force condition (weak force on the top row and high force on the bottom row), broken down by contrast levels and by SAT condition. The lines represent 1000 random draws from the joint posteriors of the combined MCMC chains of the corresponding G/LMM fits. The thick line represent the predicted regression line with all parameters set at their maximum a posteriori value.
Bottom : Same figure for the *RT*s decomposed in *N200*, *DT* and *MT*.
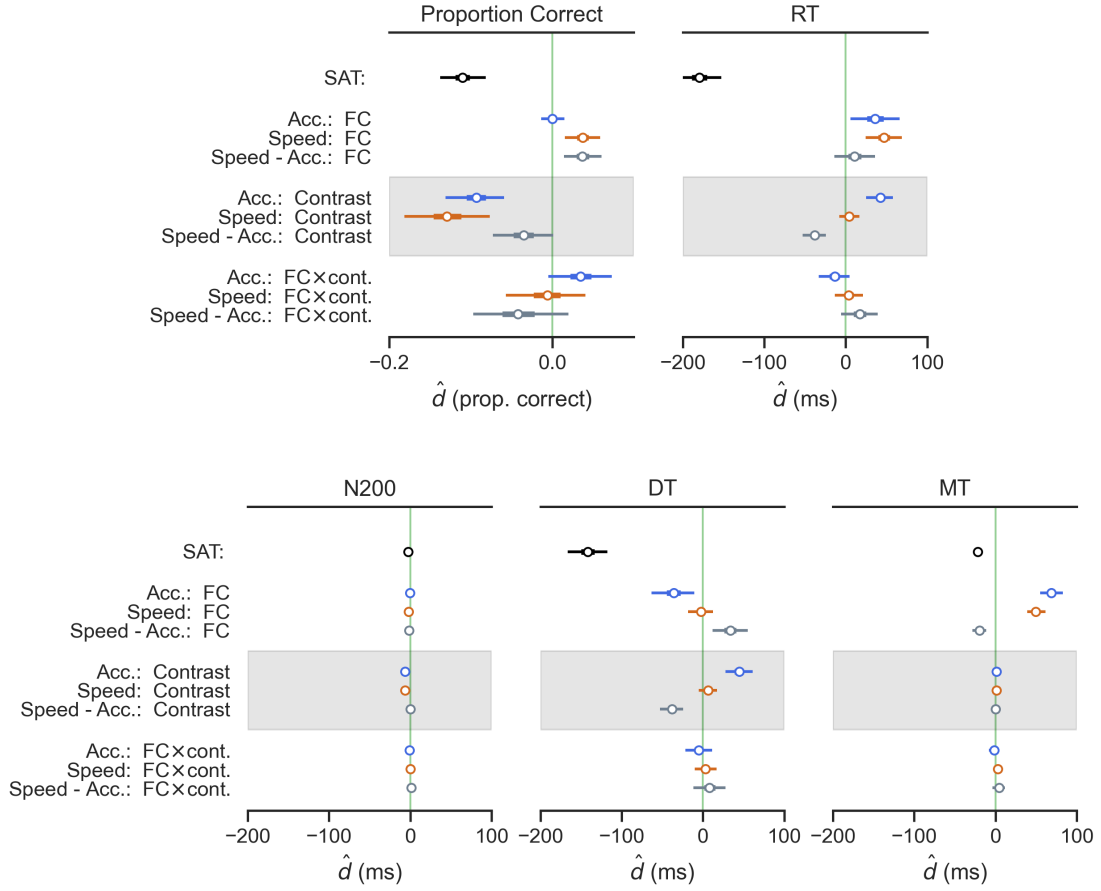
Figure 5.3.: Top : Estimated differences ($\hat{d}$) for proportion correct (left) and $RT$ (right) between condition levels for SAT, Force (FC), Contrast and their interactions on the raw scale of the data in the Accuracy condition (blue), the Speed condition (chocolate), and the difference between both conditions (i.e. interaction; grey). Dots represent the maximum a posteriori, and bars the 2.5% and 97.5 % HPD of the corresponding marginal posterior distributions.

Bottom : Same figure for the $RT$s decomposed in $N200$, $DT$ and $MT$.

## 5.4.1. Behavioral results (response times and error rates)

Proportion correct was lower when participants were asked to speed up $\hat{d}_{SAT} = -0.11$, $CrI = [-0.14, -0.08]$. Force and SAT conditions interacted, $\hat{d}_{FC:Speed-Acc.} = 0.04$, $CrI = [0.01, 0.06]$, showing that a higher force was associated with a higher proportion of correct responses in speed $\hat{d}_{Force:Speed} = 0.04$, $CrI = [0.02, 0.06]$ but not in accuracy $\hat{d}_{Force:Acc.} = 0.00$, $CrI = [-0.01, 0.02]$. An increase in contrast was associated with a decrease in correct responses in Accuracy, $\hat{d}_{Contrast:Acc.} = -0.09$, $CrI = [-0.13, -0.06]$

and in Speed, $\hat{d}_{Contrast:Speed} = -0.13$, $CrI = [-0.18, -0.08]$ (difference between SAT levels : $\hat{d}_{contrast:Speed-Acc.} = -0.04$, $CrI = [-0.07, 0.00]$). Force and Contrast factor did not seem to interact neither in Accuracy, $\hat{d}_{Force \times Contrast:Acc} = 0.04$, $CrI = [-0.01, 0.07]$, nor in Speed, $\hat{d}_{Force \times Contrast:Speed} = -0.01$, $CrI = [-0.06, 0.04]$ (difference between SAT levels : $\hat{d}_{contrast:Speed-Acc.} = -0.04$, $CrI = [-0.10, 0.02]$).

Congruent with proportion correct, the $RT$s proved to be shorter when participants were asked to speed up, $\hat{d}_{SAT} = -180$, $CrI = [-207, -153]$. The increase in Force requirements translated into an increase in the $RT$ both in Accuracy, $\hat{d}_{Force:Acc.} = 36$, $CrI = [6, 66]$, and in speed, $\hat{d}_{Force:Speed} = 47$, $CrI = [24, 69]$ (difference between SAT levels : $\hat{d}_{contrast:Speed-Acc.} = 11$, $CrI = [-14, 36]$). Contrary to proportion correct, SAT and contrast interacted on RT, $\hat{d}_{Contrast:Speed-Acc.} = -38$, $CrI = [-53, -24]$, suggesting a high increase in RT with the increase in contrast in Accuracy, $\hat{d}_{Contrast:Acc.} = 43$, $CrI = [25, 58]$, but not in Speed, $\hat{d}_{Contrast:Speed} = 4$, $CrI = [-8, 17]$. The interaction between contrast and Force appeared to be weak and including 0 in the CrI both in Accuracy, $\hat{d}_{Force \times Contrast:Acc.} = -13$, $CrI = [-33, 5]$, and Speed condition, $\hat{d}_{Force \times Contrast:Speed} = 4$, $CrI = [-8, 21]$ (difference between SAT levels : $\hat{d}_{FC \times contrast:Speed-Acc.} = 17$, $CrI = [-6, 39]$).

## 5.4.2. Results on physiological single trial RT stages

### 5.4.2.1. N200

The $N200$ proved to be sensitive to SAT in a weak but robust way, $\hat{d}_{SAT} = -2.5$, $CrI = [-3.9, -1.2]$. Relative to the effect of SAT, contrast had a high effect both in Accuracy, $\hat{d}_{Contrast:Acc.} = -6.6$, $CrI = [-9.6, -3.6]$, and Speed, $\hat{d}_{Contrast:Speed} = -6.7$, $CrI = [-9.9, -3.8]$ (difference between SAT levels : $\hat{d}_{contrast:Speed-Acc.} = -0.1$, $CrI = [-2.0, 1.7]$)[8]. The effect of contrast on $N200$ can also be appreciated on the traditional ERPs displayed in Figure 5.4. The force factor as well as its interaction with contrast was restricted to low effect sizes and included 0 in the CrI (all absolute maximum a posteriori of the posterior distributions $< 2.1$).

### 5.4.2.2. DT

The $DT$ mainly mirrored the effects of RT with a large effect of SAT, $\hat{d}_{SAT} = -142$, $CrI = [-166, -117]$, and a high effect of contrast in Accuracy, $\hat{d}_{Contrast:Acc.} = 45$, $CrI = [28, 61]$, but not in Speed, $\hat{d}_{Contrast:Speed} = 6$, $CrI = [-5, 18]$ (difference, $\hat{d}_{Contrast:Speed-Acc.} = -38$, $CrI = [-53, -24]$). The effect of Force in Accuracy proved to be comparable in size with the one on RT but of opposite sign, $\hat{d}_{Force:Accuracy} = -36$, $CrI = [-63, -11]$. Force and SAT interacted, $\hat{d}_{FC:Speed-Acc.} = 34$, $CrI = [12, 55]$, showing an effect of force centered near 0 in Speed, $\hat{d}_{Force:Speed} = -2$, $CrI = [-18, 13]$. Contrast and Force did

---

[8]Note that the individual estimates of contrast effect derived from the LMM on $N200$ showed a highly reliable effect across participants, see Figure .10 in Appendix D.2
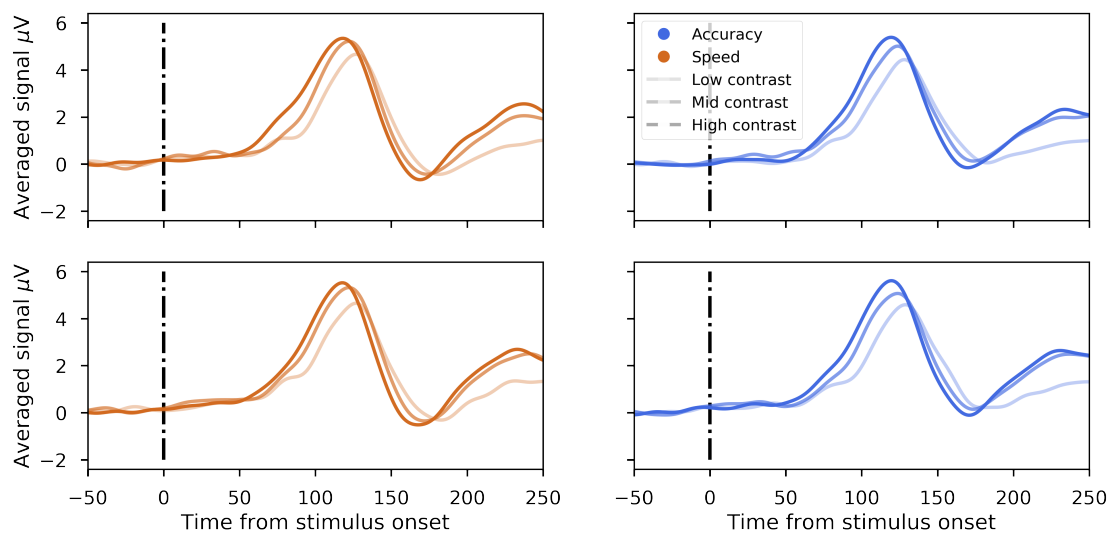
Figure 5.4.: Representation of the event related potential, the grand average value of electrodes PO7 and P08 time-locked to the stimulus onset. The event related potential is broken down across SAT conditions, contrast levels and Force conditions (weak upper row, strong bottom row). We see that the event related potentials mirrors the single trial analysis with a high effect of contrast relative to the effects of SAT and Force. Overall the difference in contrast is replicated across the experimental cells of SAT and Force.

not interact in either SAT condition, $\hat{d}_{Force \times Contrast:Acc} = -5$, $CrI = [-21, 12]$, and, $\hat{d}_{Force \times Contrast:Speed} = 3$, $CrI = [-10, 16]$.

### 5.4.2.3. MT

$MT$s proved to be shorter under speed instructions, $\hat{d}_{SAT} = -22$, $CrI = [-27, -16]$. The force Factor had a high effect on $MT$ compared to its effect on $RT$ both in Accuracy, $\hat{d}_{Force:Accuracy} = 69$, $CrI = [55, 83]$ and Speed, $\hat{d}_{Force:Speed} = 50$, $CrI = [39, 61]$. The difference of Force effect between both SAT instructions proved to be robust, $\hat{d}_{FC:Speed-Acc.} = -19$, $CrI = [-28, -11]$. Contrast as well as its interaction with Force and SAT did not appear to have an effect on $MT$ (all absolute maximum a posteriori of the posterior distributions $< 5$).

### 5.4.2.4. Wagenmakers-Brown law
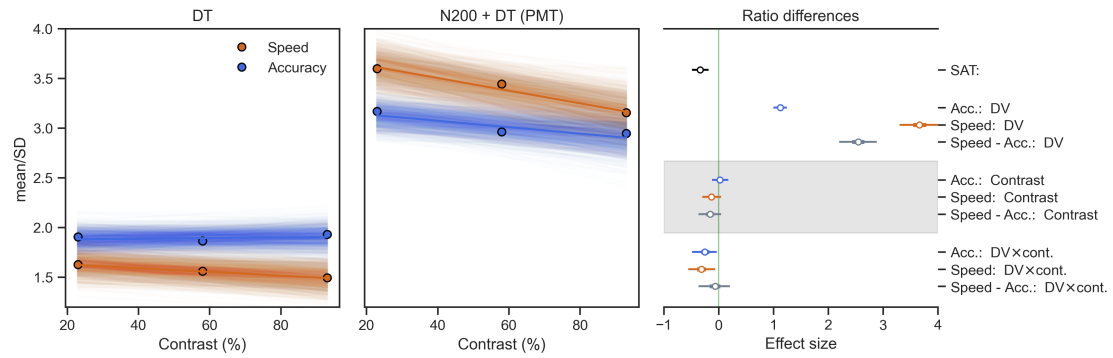


Figure 5.5.: Average values for the ratio between mean and standard deviation for the $DT$ alone (left-most panel) *vs.* the $PMT$ (*i.e.* $N200 + DT$, center panel) broken down by SAT conditions. The lines represent 1000 random draws from the joint posteriors of the combined MCMC chains of the corresponding G/LMM fits. The thick line represent the predicted regression line with all parameters set at their maximum a posteriori value. The right-most panel shows the predicted difference in ratios for the modeled dependent variable (DV factor, with $DT$ coded as 0 and $PMT$ coded as 1), contrast and the interaction between DV and contrast.

The ratio between mean and standard deviation on $DT$ proved to be lower when SAT instruction indicated to speed up the responses $\hat{d}_{SAT} = -0.34$, $CrI = [-0.49, -0.18]$. The ratio proved to be higher when computed on $PMT$ both in Accuracy, $\hat{d}_{DV:Acc.} = 1.12$, $CrI = [1.00, 1.25]$ and even higher in Speed, $\hat{d}_{DV:Speed} = 3.67$, $CrI = [3.30, 4.04]$ (Difference, $\hat{d}_{DV:Speed-Acc.} = -2.54$, $CrI = [-2.88, -2.20]$). The effect of contrast as expected was found to be centered around 0 (see Figure 5.5 left-most panel) both in Accuracy, $\hat{d}_{contrast:Acc.} = 0.02$, $CrI = [-0.12, 0.18]$, and in Speed, $\hat{d}_{contrast:Speed} = -0.13$, $CrI = [-0.3, 0.04]$ (difference, $\hat{d}_{contrast:Speed-Acc.} = 0.16$, $CrI = [-0.05, 0.37]$. But this

was found to be false when the ratio was computed on $PMT$ (see Figure 5.5 middle panel) both in Accuracy, $\hat{d}_{DV \times contrast:Acc.} = -0.25$, $CrI = [-0.49, -0.04]$, and in Speed, $\hat{d}_{DV \times contrast:Speed.} = -0.32$, $CrI = [-0.56, -0.07]$ (difference, $\hat{d}_{DV \times contrast:Speed-Acc.} = 0.07$, $CrI = [-0.21, 0.37]$).

In order to test the robustness of the present result we replicated the analysis of the effect of contrast on the ratio of mean and SD with the data from the previous chapter but on $PMT$ only (as we do not have a measure of $DT$ in the previous chapter). This result again shows that the contrast slope is negative on $PMT$ in Accuracy, $\hat{d}_{PMT:contrast:Acc.} = -0.35$, $CrI = [-0.64, -0.04]$, and in Speed although CrI barely contained 0, $\hat{d}_{PMT:contrast:Speed} = -0.50$, $CrI = [-1.09, 0.05]$ (difference, $\hat{d}_{PMT:contrast:Speed-Acc.} = 0.16$, $CrI = [-0.4, 0.7]$).

### 5.4.3. Midway discussion on behavioral and electrophysiological results

The results on the behavioral variables RT and proportion correct proved to be consistent with previous chapter although not exactly a replication. We see that the proportion of correct responses is affected by the force manipulation in speed only. This was not observed in the previous chapter although the effect estimated was of the same sign and of close magnitude (see Figure 4.2). Aside from this observation the pattern of effects on RT and proportion correct is relatively similar to the previous chapter.

About the decomposition, $MT$ results proved to be well replicated between both experiments although the force factor had a lower effect. This observation does not seem to be related to a difference in the maximum voluntary force applied by the participants at the beginning of the experiment as the ones measured in the present chapter (mean = 6.8 kg, SD = 1.8) are close to those in the previous chapter (mean = 7.0 kg, SD = 2.1). This difference between both experiment in the force effect on $MT$ could be due to the difference in the frequency of force setting alternation (see method).

The new measure of the $N200$ was, as expected negatively affected by the contrast manipulation, in a range close to the one suggested by the V1 data discussed in Section 4.6.2. In addition to the contrast effect we observe a small yet reliable effect of SAT showing that encoding processes are fastened when participants are instructed to respond faster. The force manipulation did however not change the $N200$.

The other new measure $DT$ proved to be similar to the $PMT$ in the previous chapter. SAT had a high effect. Contrast and Force also had a high effect except that they were found only in the Accuracy condition. Contrary to analysis of $PMT$ in the previous chapter we did not observe an effect of Force on $DT$ when speed was emphasized. This result cannot be attributed to the removal of the $N200$ as no effect of Force was found on this measure [9].

---

[9]Additionally an analysis of the $PMT$ replicated the absence of effect of Force in speed.

Overall this section did show that the estimation of the encoding time and decision time through the extraction of single trial *N200* proved to be congruent with our predictions. Contrast had a negative impact on *N200* and the effect of contrast on *DT* proved to verify the Wagenmakers Brown law.

## 5.4.4. Comparison of DDM and physiological estimates of RT stages

### 5.4.4.1. Factor effect on DDM parameters

See Figure 5.6 for a representation of the effect of SAT, Force and contrast factors on the estimated parameters $T_{er}$ and boundaries.
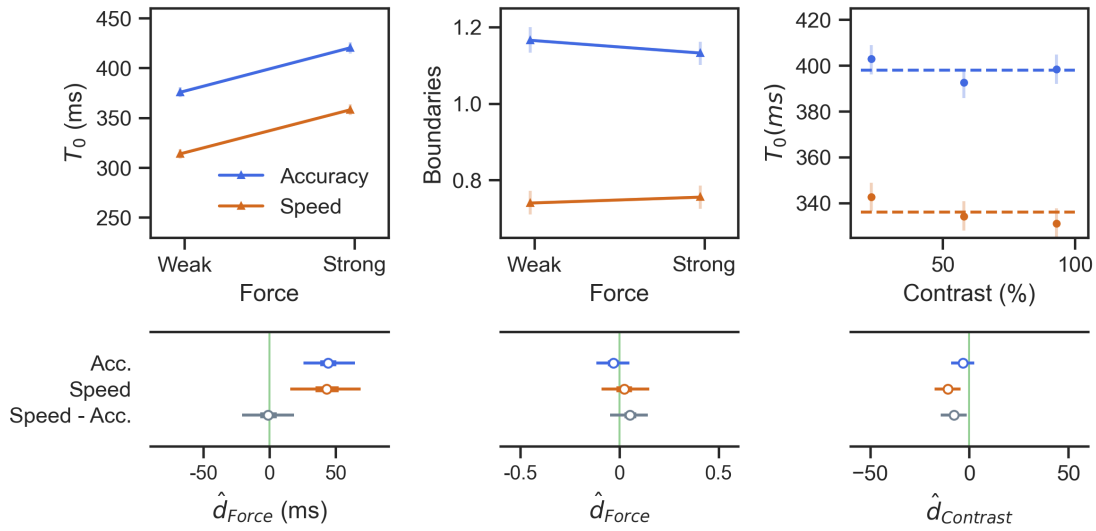


Figure 5.6.: *Effect of Force on the estimated parameter $T_{er}$ (left-most panel), boundaries (center panel) and the effect of contrast on $T_{er}$ (right-most panel).*
*Upper row: estimated mean values and one SD as shaded lines.*
*Bottom row: posterior distribution of the DDM parameter for the effect of Force on $T_{er}$ (left-most panel), Force on boundaries (center panel) and contrast on $T_{er}$ (right-most panel) at each level of SAT, and the difference in the effects between both levels.*

**Effects on** $T_{er}$  The $T_{er}$ was faster when speed was emphasized $\hat{d}_{SAT} = -67.7$, $CrI = [-83.3, -49.2]$. Force had a high effect both in Accuracy, $\hat{d}_{FC:Acc.} = 44.4$, $CrI = [25.7, 64.6]$ and Speed, $\hat{d}_{FC:Speed} = 43.5$, $CrI = [15.9, 68.8]$ (difference, $\hat{d}_{FC:Speed-Acc.} = -0.9$, $CrI = [-20.8, 18.7]$). Contrast had a smaller effect and CrI contained 0 in Accuracy, $\hat{d}_{contrast:Acc.} = -3.1$, $CrI = [-9.0, 2.6]$, but not in Speed, $\hat{d}_{contrast:Speed} = -10.8$,

$CrI = [-17.6, -4.4]$ (difference, $\hat{d}_{contrast:Speed-Acc.} = -7.7, CrI = [-14.5, -1.4]$). Contrast and Force did not seem to interact in Accuracy, $\hat{d}_{FC:contrast:Acc.} = 0.9, CrI = [-7.2, 9.9]$, nor in Speed, $\hat{d}_{FC:contrast:Speed} = 6.3, CrI = [-5.2, 17.3]$, (difference, $\hat{d}_{FC:contrast:Speed-Acc.} = 5.4, CrI = [-8.5, 18.6]$).

**Effects on boundaries**    Drift rate was affected by contrast $\hat{d}_{contrast} = -1.01, CrI = [-1.34, -0.64]$. The effect of Force on the starting point was not found to be robust, $\hat{d}_{FC} = 0.003, CrI = [-0.001, 0.007]$. The boundaries proved to be robustly sensitive to SAT, $\hat{d}_{SAT} = -0.42, CrI = [-0.49, -0.34]$. Force had however no impact on boundaries neither in Accuracy $\hat{d}_{FC:Acc.} = -0.03, CrI = [-0.12, 0.05]$ nor in Speed $\hat{d}_{FC:Speed} = 0.02$, $CrI = [-0.09, 0.15]$ (difference, $\hat{d}_{FC:Speed-Acc.} = 0.05, CrI = [-0.05, 0.14]$).

### 5.4.4.2.  Correlation between decompositions

Figure 5.7 sums up the correlations obtained on mean values and factor differences on $DT$s and $NDT$s estimated with the DDM *vs.* electrophysiology.

**Estimates**    As for the estimation of $DT$s we see that model and electrophysiological data correlates well both in Accuracy, $r_{Acc.} = 0.67, CrI = [0.34, 0.87]$ and in Speed, $r_{Speed} = 0.65, CrI = [0.32, 0.86]$. For the $NDT$s however this correlation is weaker and CrI contains 0 in Accuracy, $r_{Acc.} = 0.37, CrI = [-0.06, 0.69]$ but not in Speed, $r_{Speed} = 0.45, CrI = [0.03, 0.74]$.

**SAT differences**    The differences of SAT levels correlates on the $DT$s, $r = 0.83, CrI = [0.62, 0.94]$ but not robustly in the $NDT$s as CrI contain 0, $r = 0.29, CrI = [-0.16, 0.64]$.

**Force differences**    The differences of Force levels on $DT$s always co-varies both in Accuracy, $r_{Acc.} = 0.66, CrI = [0.32, 0.87]$, and Speed, $r_{Speed} = 0.58, CrI = [0.19, 0.82]$. However the differences of force levels on $NDT$s does not robustly correlate in Accuracy, $r_{Acc.} = 0.28, CrI = [-0.16, 0.64]$ but does so in Speed, $r_{Speed} = 0.58, CrI = [0.46, 0.91]$.

**Contrast differences**    The differences of contrast levels on $DT$s correlates in Accuracy, $r_{Acc.} = 0.58, CrI = [0.19, 0.83]$ but CrI include 0 in Speed, $r_{Speed} = 0.35, CrI = [-0.10, 0.70]$. The correlation of the differences of contrast levels on $NDT$s always included 0 both in Accuracy, $r_{Acc.} = -0.27, CrI = [-0.68, 0.22]$ and Speed, $r_{Speed} = -0.12, CrI = [-0.56, 0.35]$.

## 5.4.5.  Discussion on DDM results

The estimation of $DT$ and $NDT$ by the DDM seems globally related to the decomposition made using electrophysiology as pointed out by the correlations between
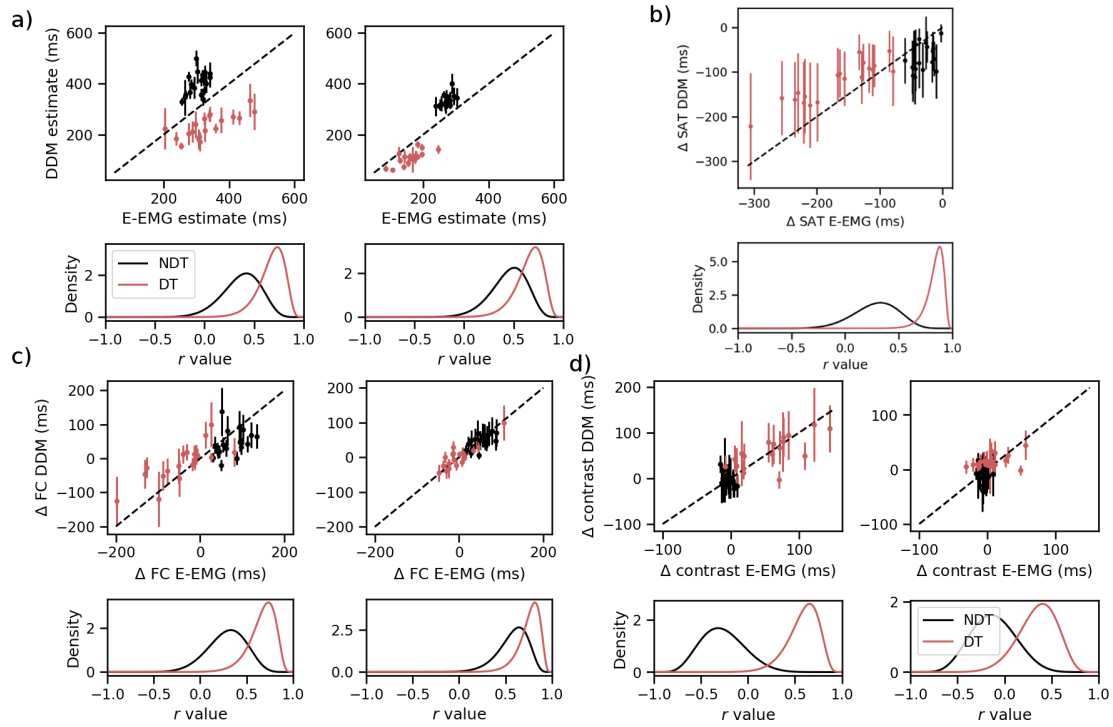
Figure 5.7.: *Correlation between mean values and factor differences on DTs and NDTs estimated with the DDM and electrophysiology (E-EMG).* All plotted values of the DDM are represented with the mean (dot) and 95% CrI (bars) of the corresponding posterior distribution. Inside subplots SAT conditions are separated with accuracy condition in the left panel and speed condition in the right panel. For each subplot the top row represents the estimated values and the bottom row the plausible value distribution.
a) Correlation of estimated mean values of DT and NDT.
b) Correlation of the SAT differences ($\Delta$ SAT).
c) Correlation of the Force differences ($\Delta$ FC).
d) Correlation of the contrast differences ($\Delta$ contrast).

estimated values. However we see that the DDM overestimates the $NDT$ (therefore underestimating the $DT$) compared to the electrophysiological single trial decomposition. In a similar way the effect of SAT on $NDT$ is overestimated by the DDM compared to the effect of SAT on electrophysiological $NDT$, the correlation of these effects is weak and non-robust.

The force effect proved to be challenging to the DDM. Force effects on $NDT$s of the DDM and electrophysiology are off in terms of magnitude and of inter-individual co-variance in the accuracy condition. In this condition DDMs $NDT$ underestimates the effect of Force and conversely overestimates the force effect on $DT$. This overestimation is linked to the fact that electrophysiological $DT$ is negatively impacted by Force manipulation in the accuracy condition while the DDMs $DT$ is found to be unaffected by the force manipulation as suggested by the effects on boundaries and starting point. When electrophysiology $DT$ is not affected by Force, as in the speed condition, both decomposition converge. However in the speed condition, we observe that the DDM does not account for the increased accuracy when a high force is required (resulting in a misfit, see the goodness of fit as illustrated in Figure .11 of Appendix D).

About the contrast effect, we replicate the findings of the previous chapter. DDMs $NDT$ is affected by contrast in a comparable way to the N200 in speed condition only. Inter-individual co-variance of the contrast effect between DDM and electrophysiology is however always weak and non-robust on both $DT$s and $NDT$s. This can point to a mis-estimation in either decomposition or that the rather weak interindividual differences are linked to measurement error more than to the targeted encoding processes.

## 5.5. General Discussion

### 5.5.1. On the non-selective nature of a manipulation of response execution

Contrary to previous studies manipulating response execution conditions (e.g. Gomez, Ratcliff, & Childers, 2015; Ho, Brown, & Serences, 2009; A. Voss, Rothermund, & Voss, 2004) the present study, as well as the previous chapter, have the advantage of manipulating response output quantitatively (changing force threshold) rather than qualitatively (different effector muscles). This probably reduces the processes affected by the response execution changes (e.g. nerve conduction time and response selection should be the same between response execution conditions). But even in this setting, we observe that the force manipulation is not selective to the motor processes. Based on the previous chapter and the present one, it appears clearly that participants engage motor processes earlier when a high force is required, translating in a shorter decision time as measured with electrophysiology. However in the present study this result is not true in the speed condition. Simultaneously specifically in

the speed condition, the proportion of correct responses is lower when force is low. This again can only be explained by the fact that participants probably continue to sample evidence while the motor response is ongoing, therefore increasing accuracy when the response execution time is longer. When instructions are to favor accuracy participants probably achieve the same level of correct responses in low force by delaying the triggering of the effector muscles to avoid errors for the weak force condition.

This strategical adjustment between SAT conditions is problematic for the DDM. In the accuracy condition, it mis-estimates the effect of force on decision and non-decision time but correctly captures the performance of the participants in terms of response proportion. In the speed condition it interprets the force effects as electrophysiology but misses the performance of the participants (shown both by the absence of force effect on boundaries in Figure 5.6 and the misfit in Figure .11). It appears that the interpretation should be that the DDM is not geared to account when non-decision processes are delayed and the proportion of correct responses increases at the same time, i.e. when decision and response execution stages are not serial.

## 5.5.2. On the recovery of encoding processes

The *N*200 proved to be a good candidate for the encoding processes as suggested by the literature (Loughnane, Newman, Bellgrove, et al., 2016; Martin, Huxlin, & Kavcic, 2010; Nunez, Gosai, Vandekerckhove, & Srinivasan, 2019). In our study, we see that the contrast has the expected negative effect on the single trial N200 (and on the ERPs, see Figure 5.4). Additionally the Wagenmakers-Brown law holds for the contrast effect only when the single trial *N*200 were removed. This again suggests that we removed the non-decision processes sensitive to the contrast factor leaving only the decision processes to be sensitive to contrast (therefore predicting a linear increase of mean and SD with contrast).

The analysis of the DDM parameters suggests the same negative effect on the $NDT$; however, it does show that this relationship is not found in the accuracy condition. This interaction between contrast and SAT instructions is inconsistent with the effect of contrast on N200 as it was found to be independent of the SAT condition. Therefore, assuming that the $NDT$ estimated by the DDM should contain the measured N200 effect, we cannot explain why the contrast effect is not found, in two independent samples of participants, by the DDM in the accuracy condition.

This results questions the nature of the failure. Two *ad-hoc* hypothesis are at hand, 1) the SAT instruction to emphasize accuracy adds a non-decision process which latency breaks the relationship between estimated $NDT$ and contrast (congruent with the proposition of Ratcliff & McKoon, 2008, when a difference is found in $T_{er}$) or 2) the DDM does not achieve a satisfactory decomposition between decision and non-decision processes. The former hypothesis would break the parametric modulation, usually through the boundaries, assumed by most modellers to explain SAT, and such hypothesis if verified should undergo a critical and thorough assessment. The latter

hypothesis is backed-up by the fact that the DDM already proved to fail to decompose force effect related to the $DT$ *vs.*the $NDT$. Such failure could occur in the presence of opposing effects between $DT$ and $NDT$ processes, e.g a high opposite effect on the $DT$ could hinder the recovery of the smaller effect on $NDT$ therefore explaining the failure in the accuracy condition where contrast has a higher and opposite effect on DT. This explanation calls simulation studies inspired by the present results to show whether such trade-offs are predicted when simulated from a known set of parameters.

### 5.5.3. Final note on the effect of SAT on non-decisional processes

On the comparison between model and physiological effects, one last comment is at stake. The N200 proved to be reliably shorter when the instructions where to emphasize speed. The MT is also shorter with speed stress as seen in this study and the three previous chapter. However both SAT effects are small (respectively -2.5 and -22ms) compared to the SAT effect estimated on the $T_{er}$ (-67 ms). In addition to this overestimation, model and electrophysiological effect on non-decision times do not seem to correlate. Again we are left with two hypothesis, 1) some additional non-decision process is present accounting for 1/4 of the SAT effect on RT or 2) the model does not achieve a good decomposition between decision and non-decision times and mis-attributes the SAT effect. The first hypothesis is unlikely, by measuring N200 and MT, we removed almost 1/2 of the mean RT (see Figure .12 in Appendix D.4) it would therefore be surprising that an additional non-decisional process with such variance between SAT conditions remains in the $DT$ extracted from electrophysiology. We are hence left with the assumption that the DDM as applied in these studies does not achieve a satisfactory decomposition between decision and non-decision latencies.

## 5.6. Conclusion

In conclusion, in this chapter we have shown, based both on the behavior and electrophysiological recordings, that a decision making model such as the DDM does not accurately separate decision and motor execution. Based on our analysis, this failure is probably due to the fact that participants continue to sample evidence during motor execution, making it difficult for the model to estimate motor execution and decision time. The estimation of encoding processes also failed in one of the, arguably most canonical, SAT level. The reason for this mis-decomposition remains to be found. Along the way we definitely show that SAT manipulation as applied in our studies cannot be explained by a parametric modulation of the decision threshold.

# 6. Discussion

> Since all models are wrong the scientist must be alert to what is importantly wrong. It
> is inappropriate to be concerned about mice when there are tigers abroad.
>
> — Box, 1976

Throughout four chapters and five experiments, we manipulated five time-honored experimental factors, each targeting specific latent cognitive processes involved in perceptual decision tasks. We believe that the experimental procedure used in this thesis lies close to most studies applying cognitive models such as EAMs and, therefore, that the conclusions reached in this thesis will generalize to a large array of decision making experiments. In this final discussion, we will summarize what the collected evidence reveals about motor processes in decision making. We will then move to a more general perspective on how to articulate decision and non-decision processes, and on what our findings imply for EAMs such as the DDM (Ratcliff, 1978) and for the inferences that are drawn from them.

## 6.1. Assumption *vs.* measurement of motor times in decision making

In this first section we review common assumptions about motor processes in $RT$ measurements, especially in 2AFC tasks. First, we inspect the distributional properties we can derive from EMG measured motor time ($MT$), then we move to describe the effect of various experimental factors on the $MT$.

### 6.1.1. Distributional properties of motor times

A common assumption is that execution time in settings such as ours is quite small compared to the mean $RT$ (e.g. see estimates in a simple RT task provided by Smith, 1995, discussed in Section 1.4.3). To provide a sense of the range of values and proportions of $RT$ that $MT$ can take, we looked at the group average mean and standard deviation of $MT$. To that end, we use the $MT$s collected in the experiment of Chapter 4, which provided both the lowest and the highest group average $MT$ in all experiments of the thesis (in the 'low force' and 'speed' vs. 'high force' and 'accuracy' conditions, respectively). In our data, the group averaged mean $MT$ ranged from approximately 65 to 183 ms, accounting for 15 and 28% of the mean $RT$, respectively, in the corresponding conditions.

A second common assumption is that the variance of non-decision processes such as the response execution time is negligible compared to that of the total $RT$ (Ratcliff, 2006). We observed that, in the same conditions used for the mean $MT$ above, the group averaged standard deviation of $MT$ takes values between 22 and 64 ms, accounting for 21 and 41% of the standard deviation of the RT. The lower bound is close to the assumption usually made about the variability of non-decision processes (e.g. 1/4 of the standard deviation of the decision processes in Ratcliff & McKoon, 2008). However, the higher bound shows that this contribution may be substantially larger, depending notably on the execution settings.

Throughout this thesis, in order to fit the linear models on the single-trial pre-motor time ($PMT$, or $DT$ in Chapter 5) and $MT$ we needed to transform the data by taking the log of the latency. Not doing so would have resulted in heavy deviations from normality for the residuals of the linear models as well as heteroscedasticity. Hence, while the question of the distribution of decision times ($DT$) and non-decision times ($NDT$) based on EMG decomposition definitely needs more advanced distributional analysis, we show that both latencies present a skew towards slow times (see Figure .13 in Appendix D.4 for a representation of the distributions and, Allain, Carbonnell, Burle, Hasbroucq, & Vidal, 2004; Possamaı, Burle, Osman, & Hasbroucq, 2002, for a report on the distribution of $MT$s).These observations contrast with the once popular hypothesis of using the Ex-Gaussian distribution to fit $RT$ data (see section 1.1.2.1), with the additional assumption that $DT$ and $NDT$ have separate distributions, one exponential and one normal (over the years, the respective attributions to $DT$ and $NDT$ changed, see Hohle, 1965).

## 6.1.2. "Decision-related" effects and motor times

Each chapter contributed to determine the impact on $MT$ of factors and observations traditionally considered to be linked to decision processes such as speed-accuracy trade-off (SAT), stimulus strength, response correctness, and cue validity.

### 6.1.2.1. Effect of SAT

SAT was manipulated in all of our experiments. All of them revealed that $MT$ was shorter when speed was emphasized, with effect sizes ranging from -10 to -24 ms. This evidence couldn't be more consistent and was obtained using standard 2AFC settings and SAT implementation through verbal instructions (see Heitz, 2014, for other means of inducing a SAT). Therefore, when designing an experiment with a SAT manipulation, researchers should assume that SAT will have an effect on response execution process. In weaker response execution settings (e.g. Experiment 2 in Chapter 2, or low force condition in Chapters 4 and 5) the effect size of SAT was rather small (around 10 ms). Hence, it is likely that ignoring these effects will not lead to dramatically ill inferences but this will in part depend on the scope of the research question and on the experimental set-up.

At the theoretical level, in contrast, this finding has important implications. It shows that motor execution is not a mere constant added to the $RT$ and, most importantly, that the participants' adaptation to the instructions includes the recruitment of motor processes. The analysis of the EMG traces in Chapter 3 revealed that participants achieve a faster motor execution time by an increased slope of the EMG signal (as already reported by Spieser, Servant, Hasbroucq, & Burle, 2017) but also by an increase in the background EMG over the trial.

### 6.1.2.2. Effect of perceptual difficulty

In Chapter 2 we observed that decreasing perceptual difficulty of the stimulus, by increasing the contrast difference between Gabor patches, shortened the $MT$ when the response was correct. About the effect of stimulus intensity on motor processes, Sanders, 1983 suggest the immediate arousal hypothesis where a strong stimulation will enhance response readiness. This theory is compatible with data in 2 as the response associated with the most contrasted (correct) patch was executed faster and that this effect increased with the increase in contrast of the correct patch. However this theory does not hold for Chapters 4 and 5. In these chapters we manipulated the overall contrast of both patches while keeping the difference between both constant. This manipulation was associated with even higher contrast values, i.e. physical stimulation, than the manipulation of Chapter 2. But we did not observe a reliable effect on $MT$ in these chapters. Therefore it seems that the effect of the contrast difference of both patches on $MT$ is linked to the perceptual difficulty rather than the physical stimulation as this perceptual difficulty was manipulated in a broader range in Chapter 2 (e.g. for a quick comparison see quantile probability plot of Chapter 2 in Figure 2.4 *vs.* the same quantile probability plot in Figure .7 of Appendix C).

### 6.1.2.3. Effect of correctness

The effect of correctness is defined as the time difference between correct and incorrect responses. As described in the introduction, accounting for this effect is a hallmark of $RT$ modelling, as it has conducted to refine or abandon models of decision making that couldn't account for it. As observed in numerous studies, in this thesis, the relative speed of error response compared to correct responses on $RT$ is mediated by SAT. Errors are faster than correct responses when speed is emphasized, while a slightly reversed pattern is observed when accuracy is emphasized. Interestingly, across the three experiments of Chapters 2 and 3, we systematically observe a longer $MT$ for errors, independent of the SAT condition or force response execution settings (compare coefficients between both experiments of Chapter 2 in Figure 2.2).

In Chapter 2 the speed difference in $MT$ between correct *vs.*error responses could be explained by the immediate arousal hypothesis by Sanders, 1983 spelled out above as the correct patch had the highest physical stimulation. But physical stimulation was not manipulated in Chapter 3 and we nevertheless observed an effect of correctness on $MT$. Allain, Carbonnell, Burle, Hasbroucq, and Vidal, 2004; Śmigasiewicz, Ambrosi,

Blaye, and Burle, 2020 interpreted the slower motor execution of errors as evidence for an on-line executive control, whereby participants try to stop the ongoing erroneous response. In Chapter 2 this correctness effect interacted with the perceptual difficulty manipulation. Easier stimuli yielded both longer error $MT$s and shorter correct response $MT$s, compared to harder stimuli. The online executive control hypothesis cannot explain how correct response yield faster MT with increasing stimulus ease. We suggest a joint account for the effect of perceptual difficulty and correctness effect below.

### 6.1.2.4. A tentative account on the origin of effects observed on motor times

A parallel project of the thesis (lead with other collaborators and reported in Gajdos, Fleming, Saez Garcia, Weindel, & Davranche, 2019) showed that the motor system is involved in the evaluation of the confidence in a decision. Additionally, studies such as Balsdon, Wyart, and Mamassian, 2020 or Dotan, Meyniel, and Dehaene, 2018 prove that confidence in a decision results from a process happening during the decision. Hence, given that the motor system is linked to confidence evaluation and that such confidence evaluation is evolving during the decision, we can suggest that the confidence evaluation might explain the effects we observe on motor times. For example, the study by Dotan, Meyniel, and Dehaene, 2018 shows that participants speed their movement according to the momentary degree of confidence in the ongoing response. Therefore, the effect of correctness and its interaction with stimulus strength could be linked to such speed variations induced by confidence. In our data, a correct response on easy trials is probably prompting a higher confidence than errors. This could also be an explanation of the effect of cue validity on $MT$ found in chapter 3. When the cue was invalid and the response correct, response execution was faster than when the cue was valid. We can imagine, although this has to be tested given the *ad-hoc* character of the hypothesis, that a greater confidence is needed to beat the odds given by the cue, therefore translating in a faster execution time.

As as summary we see that effects classically assumed to be decisional do also affect the response execution time $MT$. We make the hypothesis that these effects (to the exception of SAT) could be linked to the confidence in the ongoing response. Note that if that interpretation turned out to be true, the effects on motor times couldn't be recovered by a model that would not account for how confidence arises during the decision.

## 6.2. On the relations between decision and non-decision times

While the effects found on $MT$ proved to be counter-intuitive and prone to generate interesting hypothesis regarding their origin(s), we do see that these factors had rather

small effect sizes compared to their effect on the pre-motor time $PMT$ derived from EMG. However, factors known to highly impact the $MT$ such as the changes in the force required for responding also notably translated in effects on $PMT$.

In Chapter 2, we noticed that the estimation of non-decision times by the model did not change across the experiments despite having reduced the force needed to produce a response. This result casts doubts on whether genuine non-decision effects where actually recovered by the estimated non-decision processes. In addition, Chapter 3 showed that the DDM fitted in the Chapter 2 did confound a motor bias with a decisional bias. A difference in the speed of response execution between left and right hands was attributed by the DDM as a change in the starting point of the evidence accumulation process. Using the modelling strategy proposed by A. Voss, Voss, and Klauer, 2010 did align model and EMG inferences that a response execution bias was present while the starting point was unbiased. In short, while factors reputed to be decisional might not highly impact the motor times, it could however be that a genuine effect on non-decisional processes might be interpreted by the DDM as originating from the decision processes.

Chapter 4 and Chapter 5 were geared to test whether manipulations targeted to impact encoding and motor stages could be recovered as affecting the non-decision processes estimated by a sequential decision making model. These chapters provide mixed evidence in favor of the decomposition of $RT$ assumed by these models.

## 6.2.1. Encoding and decision processes

In Chapter 5, based on previous studies we extracted a time we thought to be the best candidate for what we defined as the encoding time in our task. This measure displayed a relatively low variance compared to the $RT$ (group averaged SD of 25ms, see also .13 in Appendix D.4). This measure varied as expected with stimulus strength and removing it from the $RT$s proved to verify the Wagenmakers-Brown law for the effect of stimulus strength on the extracted post-encoding time. Additionally, such measured encoding time presented a small reduction in the speed condition of the SAT manipulation and no effect when response execution force was manipulated.

When directly trying to measure the correlation between measure and estimate of encoding time we did not observe that the inter-individual variability of the measure predicted the inter-individual variability of the estimate. However, this result, or lack thereof, could be due to the fact that the latent encoding time has low inter-individual variability making the inter-individual variance linked to the measurement or estimation noise rather than to the underlying process.

But one problem appeared when comparing the effect of contrast computed on the measured *vs.* the estimated encoding time. Both sources indicated a monotone negative relationship with the contrast factor when SAT instructions emphasized speed, however only the measure showed this expected relationship when SAT instructions emphasized accuracy.

## 6.2.2. Motor and decision times are occurring in parallel

Chapter 4 first showed that decreasing the force required to execute a response increased the $PMT$, however this increase in what we initially assumed to be the time containing decision time did not translate into a difference in the proportion of correct responses. The fact that participants achieve the same accuracy while spending more time prior to the onset of muscle activity can be explained by the fact that they continue to decide while they engage in the response execution. Despite being unable to account for this overlap between motor and decision processes, the DDM provided an indirect confirmation. When fitted on the $PMT$ the DDM predicted a lower accuracy than the one observed. This shows that the DDM interprets the missing portion of the motor time as a time during which a part of the decision was taking place. Hence this chapter shows that the time we record as non-decision time is not interpreted as a non-decision time by the model. Therefore it appears that, in order to account for the accuracy data, the model underestimates the time spent on executing the response. As models such as DDMs are usually used to make inferences on decision processes this can be a desirable property although it precludes from interpreting non-decision time as motor execution time (but see section 6.3.2 for a more detailed discussion on that topic).

Chapter 5 strengthens the interpretation of the overlap and adds on the question of the consequences for a serial model such as the DDM. In this chapter, when speed was emphasized, participants made more correct responses when a high force was required compared to a low force. As only $MT$s were longer, this undoubtedly leads again to the conclusion that participants where deciding while their motor response was ongoing. This pattern of change in accuracy between force conditions was however missed by the DDM. This shows that, in this condition, the DDM was unable to attribute decision time to the execution time period as measured with EMG.

### 6.2.2.1. Is motor time merely decision time ?

Based on these findings it could be tempting to conclude that the motor time derived by EMG is simply a subset of the decision time. This is by no means guaranteed and in fact non-congruent with other findings from the thesis. If pre-motor and motor times were to contain shares of the decision time (supposedly the part of $RT$ that has the most variance) they should positively correlate which is definitely not found in all trials as reported in Chapter 2. Additionally, the experimental factors affecting the decision processes should all have an effect in the same direction on $PMT$ and $MT$ which is again not found as stated in the previous section. Hence, despite the fact that participants can still change their minds while the response is ongoing, this observation is not incompatible with the fact that the motor time has its own variance and sensitivity to experimental factors.

We did observe, though, that the motor time has a negative correlation with $PMT$ on fast trials. We cannot exclude that this observed negative correlation is due to properties of activation dynamics in the motor system once a stimulus is presented

(e.g. progressive increase in cortical excitability following stimulus onset Tandonnet, Davranche, Meynier, et al., 2012). We can however see that the assumption on the link between confidence and $MT$ does fit with the observed negative correlation. If we make the reasonable assumption that fast trials yield less confidence in the chosen response (i.e. guesses or partial guesses), then the slowdown of $MT$ compared to $PMT$ can again be explained by the confidence hypothesis which presumes longer $MT$s when confidence is low.

### 6.2.2.2. A note on the presence of multiples EMG activities trials

The interpretation that participants can change their minds while the response is ongoing leads to the prediction that they can interrupt the ongoing motor plan. This is congruent with the fact that we observed, in all experiments performed, trials with multiple EMG activities. While these trials have an obvious interpretation in tasks where response conflict is present as in the study of Servant, White, Montagnini, and Burle, 2016, the genuine nature of these trials in the present tasks is not warranted for two reasons. First, because some non-negligible part might originate from a failure to record the button press as a response by the recording device or perhaps a failure of the participant to produce the necessary force. In that case these trials should not be seen as valid trials. Second, even if the multiple activities were genuine changes-of-mind, these could also represent trials where the participant were in a different responding state (e.g. mind wandering, fatigue). Hence while a thorough evaluation of these trials is needed, the present thesis could not address how they are generated.

### 6.2.2.3. Confidence and parallel decision and motor times

We can link the hypothesis we made on the source of the effects on motor times being related to confidence and the proposal of parallel occurring motor and decision times. Balsdon, Wyart, and Mamassian, 2020 have shown that, in the decision tasks where evidence comes at sequential steps, the confidence dictates the threshold of the decision. The generalization is not easy as their task is different from the one used here, especially because all evidence is presented simultaneously in our tasks. We can still suggest that their interpretation could accommodate with the temporal relationship between motor and decision time. If the response execution is linked to the ongoing confidence and as decision occurs during response execution time, then mechanistically, a higher confidence dictates a shorter decision time through the speed up of motor processes and hence a lower decision threshold.

## 6.3. Consequences for the inferences made from decision making models

The results and interpretations we outlined so far have consequences for the way we interpret serial decision making models such as the DDM.

### 6.3.1. The fuzzy boundaries between thresholds and response execution time

Implicitly at least, researchers often assume that in order to commit to a decision, it is first needed to reach a necessary level of evidence before producing an action. But the present data show that even when the action is a simple isometric button press this view is incomplete. While there is probably a criterion at which the uncertainty is low enough to start committing to an action, this does not seem to be the criterion that EAMs like the DDM actually model. This might be seen as a conceptual problem requiring simply to redefine what we interpret when estimating thresholds and non-decision time but it seems to translate to a computational problem as the DDM failed to accommodate the accuracy data in Chapter 5.

While less conclusive probably because of a lower effect size, the same interpretation can be made about experiment 1 in Chapter 2 and the reanalysis in Section 3.2. We observed at the same time a faster $MT$ for right hand responses and an increased accuracy (3%) for left responses (see Appendix A.3). The DDM interpreted these results as a difference in the starting point. Mathematically for a model with two thresholds and a starting point such as the DDM a bias in starting point is equivalent to having non-symmetric thresholds. Hence the estimated bias could be interpreted as a higher threshold in the left hand. In the original experiment the DDM missed the effect on $MT$ but captured the difference in accuracy. In the Section 3.2 accounting for a difference in response execution translated in correctly estimating the difference in $MT$ although the pattern of accuracy was then left as unaccounted. We therefore think, as for Chapter 4 and 5, that this shows that a difference in response execution is actually a difference in the available decision time.

Altogether this strengthens the suggestion that motor execution and decision are two different processes that overlap in time as it appears that the DDM can either account for an effect on response execution or on decision (i.e. accuracy) but not on both.

### 6.3.2. On the interpretation of the non-decision time parameter

The previous interpretation shows that the inferences one can make about motor processes by analyzing the non-decision time will likely only reflect part of the actual motoric effect. The fact that the co-variation between DDM estimates of non-decision

time and the single trial motor times is only moderate points to the ability of the model to account for motor processes contained in $MT$ but probably not all of it.

As seen when comparing measured *vs.* estimated encoding time, we also raised a problem in the estimation of a encoding-related non-decision process. When we manipulated the encoding processes by varying contrast, we did not observe the expected effect on the non-decision time parameter (whether we accounted for the motor processes or not) in the accuracy condition. We have no straightforward interpretation for this miss, although it could be, as suggested in Chapter 5, that an opposite effect between decision and non-decision times might be missed by the DDM. In all cases, we interpret the unexpected high coherence between estimate of encoding processes and the data from visual cortex area V1 as measured in awake monkeys by Reynaud, Masson, and Chavane, 2012 as an impressive performance of the DDM.

### 6.3.3. SAT as a parametric modulation

One of the cornerstones of the DDM is arguably its ability to explain the pattern of effects on accuracy and $RT$ by a SAT manipulation selectively through a threshold variation. Throughout five experiments we have shown that the $MT$ is affected by SAT manipulations. Based on the previous interpretations it could be that the speeding of motor processes might actually contribute to the decrease in the thresholds observed when participants are asked to speed up. Indeed we do observe that the effect of SAT on thresholds is slightly reduced when the DDM is fitted on $PMT$ in Chapter 4. Although this difference is too weak to be reliably interpreted it is worth noting that it is coherent with the previous interpretation.

However, we also observe that the SAT manipulation impacts the estimated non-decision time as in previous studies (e.g. Palmer, Huk, & Shadlen, 2005b; Ratcliff, 2006; A. Voss, Rothermund, & Voss, 2004). If the above mentioned explanation of the SAT effect of $MT$ holds and, given the weak SAT effect we observed on the measured encoding time, this means that we do not know which part of the non-decision processes are sped-up. But even if the effect of SAT on $MT$ is genuinely interpreted as a non-decisional effect, the DDM still largely overestimates that effect compared to our electrophysiological measurement. Ratcliff and McKoon, 2008 suggested that a difference in the non-decision time shows that a non-decision process is added between the conditions. This could be linked to the fact that we falsified the prediction of contrast on the estimated encoding time. It could then be suggested, based on this rationale, that a non-decision process is present in accuracy condition that breaks the relation between contrast and estimated encoding time. However this would definitely break the selective threshold explanation of SAT, question the functional purpose of the putative added non-decision process, and more generally, would represent a very surprising *ad-hoc* assumption.

### 6.3.4. Non-decision time is diagnostically important

In their collaborative study on selective influence consensus, Dutilh, Annis, Brown, et al., 2019 (discussed in Section 1.3.6) found that most modelers agreed on the inferences from the blind manipulations only when inferences on the non-decision time were ignored. Smith and Lilburn, 2020a argued that this result amongst others points to an inappropriate modeling strategy as visual properties of the task are ignored. This does show that while non-decision processes (or more generally residual parameters) might not be at the core interest of researchers, their estimation is diagnostically important for the modeling strategy used.

The present thesis shows that 1) decision is parallel to the motor execution when response execution settings are high, 2) measurement *vs.* estimation of non-decision processes are congruent only when speed is emphasized and 3) that accuracy instructions induced an additional time which origin and functional purpose is unknown. We conclude that non-decision times shows that the modeling strategy adopted in the present thesis is only proven to be valid when speed is emphasized and response execution low. While it is possible that other conditions also estimates genuine non-decision processes that can explain the pattern of results this claim can however not be falsified by the data at hand and appears rather unlikely.

## 6.4. General conclusion

Based on the results of the present thesis, we end up with two novel claims about the way human participants perform perceptual decisions. First, the motor execution time captured by EMG is a process with its own variability and sensitivity to experimental factors, even when the response is a simple isometric button press. We provide some clue on how we think these factors impact the execution time. Second, based on the behavior, the EMG decomposition, and the modeling through a decision making model such as the DDM, we see that this measured execution time is occurring in parallel to the decision time.

For models of decision making, and therefore for the theories they are built on, this is more than a simple conceptual clarification. First, accounting for the effects on response execution processes can only be achieved by a mathematical cognitive model and theory therein, that account for the origin of these effects. It is up to the researchers to choose whether such variations can be ignored for the sake of parsimony but the response execution settings will determine the impact of motor processes on RT. Second, while the question of whether the incongruences between estimation and measure of cognitive processes raised in the present thesis are tigers or mices (Box, 1976) for decision making models remains open, we do conclude that the assumptions made by evidence accumulation models should be tested in the design they are applied to before being used for inferences on psychological processes. While cognitive mathematical models such as EAMs of decision making can provide valuable information on the processes underlying performance, their generalizability

to a large array of conditions and tasks is not warranted until such conditions have been thoroughly tested.

# Bibliography

Allain, S., Carbonnell, L., Burle, B., Hasbroucq, T., & Vidal, F. (2004). On-line executive control: An electromyographic study. *Psychophysiology*, *41*(1), 113–116. https://doi.org/10.1111/j.1469-8986.2003.00136.x (cit. on pp. 48, 52, 55, 56, 72, 99, 154, 155, 225)

Anders, R., Hinault, T., & Lemaire, P. (2018). Heuristics versus direct calculation, and age-related differences in multiplication: An evidence accumulation account of plausibility decisions in arithmetic. *Journal of Cognitive Psychology*, *30*(1), 18–34 (cit. on p. 32).

Anders, R., Alario, F., Van Maanen, L., Et al. (2016). The shifted wald distribution for response time data analysis. *Psychological methods*, *21*(3), 309 (cit. on p. 19).

Anders, R., Riès, S., van Maanen, L., & Alario, F. X. (2017). Lesions to the left lateral prefrontal cortex impair decision threshold adjustment for lexical selection. *Cognitive Neuropsychology*, *34*(1-2), 1–20. https://doi.org/10.1080/02643294.2017.1282447 (cit. on p. 32)

Ando, T. (2007). Bayesian predictive information criterion for the evaluation of hierarchical bayesian and empirical bayes models. *Biometrika*, *94*(2), 443–458 (cit. on pp. 63, 112).

Baddeley, A. D., & Hitch, G. (1974). Working memory, In *Psychology of learning and motivation*. Elsevier. (Cit. on p. 21).

Balsdon, T., Wyart, V., & Mamassian, P. (2020). Confidence controls perceptual evidence accumulation. *Nature communications*, *11*(1), 1–11 (cit. on pp. 156, 159, 225, 227).

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*(June), 3–4. https://doi.org/10.3389/fpsyg.2013.00328 (cit. on pp. 90, 184)

Barthélemy, F. V., Fleuriet, J., & Masson, G. S. (2010). Temporal dynamics of 2d motion integration for ocular following in macaque monkeys. *Journal of Neurophysiology*, *103*(3), 1275–1282 (cit. on p. 195).

Bartlett, N. R. (1963). A Comparison of Manual Reaction Times as Measured by Three Sensitive Indices. *Psychological Record*, *13*, 51–56 (cit. on p. 56).

Bastien, C., & Bastien-Toniazzo, M. (2003). Le décalage entre performance et processus. *Psychologie et psychométrie*, *24*(4), 7–24 (cit. on pp. 21, 221).

Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., Krypotos, A. M., Lerche, V., Logan, G. D., Palmeri, T. J., van Ravenzwaaij, D., Servant, M., Singmann, H., Starns, J. J., Voss, A., Wiecki, T. V., Matzke, D., & Wagenmakers, E. J. (2018). Estimating across-trial variability parameters of the Diffusion

Decision Model: Expert advice and recommendations. *Journal of Mathematical Psychology, 87*, 46–75. https://doi.org/10.1016/j.jmp.2018.09.004 (cit. on pp. 63, 79, 111)

Boehm, U., Marsman, M., Matzke, D., & Wagenmakers, E.-J. (2018). On the importance of avoiding shortcuts in applying cognitive models to hierarchical data. *Behavior research methods, 50*(4), 1614–1631 (cit. on p. 112).

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review, 113*(4), 700 (cit. on pp. 25, 222).

Botwinick, J., & Thompson, L. W. (1966). Premotor and motor components of reaction time. *Journal of Experimental Psychology, 71*(1), 9–15. https://doi.org/10.1037/h0022634 (cit. on pp. 38, 39)

Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association, 71*(356), 791–799 (cit. on pp. 153, 162).

Brody, C. D., & Hanks, T. D. (2016). Neural underpinnings of the evidence accumulator. *Current opinion in neurobiology, 37*, 149–157 (cit. on p. 38).

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology, 57*(3), 153–178. https://doi.org/10.1016/j.cogpsych.2007.12.002 (cit. on p. 25)

Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review, 112*(1), 117–128. https://doi.org/10.1037/0033-295X.112.1.117 (cit. on p. 103)

Buc Calderon, C., Verguts, T., & Gevers, W. (2015). Losing the boundary: Cognition biases action well after action selection. *Journal of Experimental Psychology: General, 144*(4), 737 (cit. on pp. 71, 128, 227).

Burle, B., Roger, C., Vidal, F., & Hasbroucq, T. (2008). Spatio-temporal dynamics of information processing in the brain: Recent advances, current limitations and future challenges. *International Journal of Bioelectromagnetism, 10*, 17–21 (cit. on p. 38).

Burle, B., Possamaï, C.-A., Vidal, F., Bonnet, M., & Hasbroucq, T. (2002). Executive control in the Simon effect: an electromyographic and distributional analysis. *Psychological Research, 66*(4), 324–36. https://doi.org/10.1007/s00426-002-0105-6 (cit. on pp. 38, 39, 57, 72, 105, 123)

Calderon, C. B., Van Opstal, F., Peigneux, P., Verguts, T., & Gevers, W. (2018). Task-relevant information modulates primary motor cortex activity before movement onset. *Frontiers in Human Neuroscience, 12*(March), 1–12. https://doi.org/10.3389/fnhum.2018.00093 (cit. on p. 83)

Callaway, E., Halliday, R., Naylor, H., & Thouvenin, D. (1984). The latency of the average is not the average of the latencies. *Psychophysiology, 21*, 571 (cit. on p. 38).

Carpenter, R. H. (1981). Oculomotor procrastination. *movement: Cognition and Visual Perception* (cit. on p. 19).

Chatrian, G.-E., Lettich, E., & Nelson, P. L. (1988). Modified nomenclature for the "10%" electrode system1. *Journal of Clinical Neurophysiology, 5*(2), 183–186 (cit. on p. 134).

Christie, L. S., & Luce, R. D. (1956). Decision structure and time relations in simple choice behavior. *The bulletin of mathematical biophysics, 18*(2), 89–112 (cit. on p. 19).

Coles, M. G. (1988). Modern Mind-Brain Reading : Psychophysiology, physiology and cognition. (Cit. on p. 17).

Coles, M. G. (1989). Modern mind-brain reading: Psychophysiology, physiology, and cognition. *Psychophysiology, 26*(3), 251–269 (cit. on p. 70).

Coles, M. G., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. *Journal of Experimental Psychology: Human Perception and Performance, 11*(5), 529 (cit. on p. 38).

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*(1), arXiv arXiv:1011.1669v3, 42–45. https://doi.org/10.20982/tqmp.01.1.p042 (cit. on p. 51)

de Lafuente, V., & Romo, R. (2006). Neural correlate of subjective sensory experience gradually builds up across cortical areas. *Proceedings of the National Academy of Sciences, 103*(39), 14266–14271. https://doi.org/10.1073/pnas.0605826103 (cit. on p. 38)

Dirk, J., Kratzsch, G. K., Prindle, J. P., Kröhne, U., Goldhammer, F., & Schmiedek, F. (2017). Paper-Based Assessment of the Effects of Aging on Response Time: A Diffusion Model Analysis. *Journal of Intelligence, 5*(2), 12. https://doi.org/10.3390/jintelligence5020012 (cit. on p. 32)

Donders, F. C. (1868). Die schnelligkeit psychischer processe: Erster artikel. *Archiv für Anatomie, Physiologie und wissenschaftliche Medicin,* 657–681 (cit. on pp. 16, 17).

Donkin, C., Brown, S. D., & Heathcote, A. (2009). The overconstraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin & Review, 16*(6), 1129–1135 (cit. on p. 27).

Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E. J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin and Review, 18*(1), 61–69. https://doi.org/10.3758/s13423-010-0022-4 (cit. on p. 33)

Dotan, D., Meyniel, F., & Dehaene, S. (2018). On-line confidence monitoring during decision making. *Cognition, 171,* 112–121 (cit. on pp. 71, 128, 156, 225, 227).

Dubarry, A.-S., Llorens, A., Trébuchon, A., Carron, R., Liégeois-Chauvel, C., Bénar, C.-G., & Alario, F.-X. (2017). Estimating Parallel Processing in a Language Task Using Single-Trial Intracerebral Electroencephalography. *Psychological Science, 28*(4), 414–426. https://doi.org/10.1177/0956797616681296 (cit. on pp. 38, 70)

Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P., Hawkins, G. E., Heathcote, A., Holmes, W. R., Krypotos, A. M., Kupitz, C. N., Leite, F. P., Lerche, V., Lin, Y. S., Logan, G. D., Palmeri, T. J., Starns, J. J., Trueblood, J. S., van Maanen, L., ... Donkin, C. (2019). The Quality of Response Time Data Inference: A Blinded, Collaborative Assessment of the Validity of Cognitive Models. *Psychonomic Bulletin and Review*, *26*(4), 1051–1069. https://doi.org/10.3758/s13423-017-1417-2 (cit. on pp. 35, 36, 130, 162)

Dutilh, G., Wagenmakers, E.-J., Visser, I., & van der Maas, H. L. J. (2011). A phase transition model for the speed-accuracy trade-off in response time experiments. *Cognitive Science*, *35*, 211–250. https://doi.org/10.1111/j.1551-6709.2010.01147.x (cit. on p. 75)

Fessard, A.-B., Fessard, A., Kowarski, D., Laugier, H., & Monnin, J. (1936). Rapidité et variabilité des temps de réaction chez les enfants d'âge scolaire. *Biotypologie*, *4*(2) (cit. on p. 18).

Fluchère, F., Burle, B., Vidal, F., Van den Wildenberg, W., Witjas, T., Eusebio, A., Azulay, J.-P., & Hasbroucq, T. (2018). Subthalamic nucleus stimulation, dopaminergic treatment and impulsivity in parkinson's disease. *Neuropsychologia*, *117*, 167–177 (cit. on p. 72).

Forstmann, B. U., & Wagenmakers, E. J. (2015). Model-based cognitive neuroscience: A conceptual introduction. *An Introduction to Model-Based Cognitive Neuroscience*, 139–156. https://doi.org/10.1007/978-1-4939-2236-9_7 (cit. on p. 37)

Gajdos, T., Fleming, S. M., Saez Garcia, M., Weindel, G., & Davranche, K. (2019). Revealing subthreshold motor contributions to perceptual confidence [niz001]. *Neuroscience of Consciousness*, *2019*(1), https://academic.oup.com/nc/article-pdf/2019/1/niz001/27857234/niz001.pdf. https://doi.org/10.1093/nc/niz001 (cit. on pp. 156, 225)

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472 (cit. on pp. 48, 62, 111).

Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, *36*(2), 299–308. https://doi.org/10.1016/S0896-6273(02)00971-6 (cit. on pp. 21, 222)

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual review of neuroscience*, *30* (cit. on p. 34).

Gomez, P., & Perea, M. (2014). Decomposing encoding and decisional components in visual-word recognition: A diffusion model analysis. *The Quarterly Journal of Experimental Psychology*, *67*(12), 2455–2466 (cit. on p. 29).

Gomez, P., Perea, M., & Ratcliff, R. (2013). A diffusion model account of masked versus unmasked priming: are they qualitatively different? *Journal of Experimental Psychology: Human Perception and Performance*, *39*(6), 1731–1740. https://doi.org/10.1037/a0032333 (cit. on p. 29)

Gomez, P., Ratcliff, R., & Childers, R. (2015). Pointing, looking at, and pressing keys: A diffusion model account of response modality. *Journal of Experimental Psy-*

*chology: Human Perception and Performance, 41*(6), 1515–1523. https://doi.org/10.1037/a0039653 (cit. on pp. 29, 103–105, 112, 150)

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, (7 DEC), arXiv [1] A. Gramfort, âĂIJMEG and EEG data analysis with MNE-Python,âĂİ Front. Neurosci., vol. 7, no. December, pp. 1âĂŞ13, 2013. https://doi.org/10.3389/fnins.2013.00267 (cit. on pp. 46, 108, 135, 183)

Grayson, D. (1983). *The Role of the Response Stage in Stochastic Models of Simple Reaction Time* (Doctoral dissertation). University of Sydney. (Cit. on pp. 39, 44, 56).

Grice, G., & Spiker, V. (1979). Speed-accuracy tradeoff in choice reaction time: Within conditions, between conditions, and between subjects. *Attention, Perception & Psychophysics, 26*(2), 118–126. https://doi.org/10.3758/bf03208305 (cit. on p. 48)

Guest, O., & Martin, A. E. (2020). How computational modeling can force theory building in psychological science (cit. on pp. 21, 32).

Haaf, J. M., & Rouder, J. N. (2017). Developing Constraint in Bayesian Mixed Models. *Psychological Methods, 22*(4), 779–798. https://doi.org/10.1037/met0000156 (cit. on p. 100)

Haith, A. M., Huberdeau, D. M., & Krakauer, J. W. (2015). Hedging your bets: Intermediate movements as optimal behavior in the context of an incomplete decision. *PLoS Comput Biol, 11*(3), e1004171 (cit. on p. 128).

Harwerth, R. S., & Levi, D. M. (1978). Reaction time as a measure of suprathreshold grating detection. *Vision research, 18*(11), 1579–1586 (cit. on p. 105).

Hasbroucq, T., Mouret, I., Seal, J., & Akamatsu, M. (1995). Finger pairings in two-choice reaction time tasks: Does the between-hands advantage reflect response preparation? *Journal of Motor Behavior, 27*(3), 251–262 (cit. on p. 39).

Hasbroucq, T., Osman, A., Possamaı, C.-A., Burle, B., Carron, S., Dépy, D., Latour, S., & Mouret, I. (1999). Cortico-spinal inhibition reflects time but not event preparation: Neural mechanisms of preparation dissociated by transcranial magnetic stimulation. *Acta Psychologica, 101*(2-3), 243–266 (cit. on p. 83).

Hasbroucq, T., Possamaı, C.-A., Bonnet, M., & Vidal, F. (1999). Effect of the irrelevant location of the response signal on choice reaction time: An electromyographic study in humans. *Psychophysiology, 36*(4), 522–526 (cit. on p. 72).

Heathcote, A., Brown, S. D., & Wagenmakers, E.-J. (2015). An introduction to good practices in cognitive modeling, In *An introduction to model-based cognitive neuroscience*. Springer. (Cit. on p. 35).

Heathcote, A., Lin, Y.-S., Reynolds, A., Strickland, L., Gretton, M., & Matzke, D. (2019). Dynamic models of choice. *Behavior Research Methods, 51*(2), 961–985 (cit. on p. 64).

Heathcote, A., & Love, J. (2012). Linear deterministic accumulator models of simple choice. *Frontiers in Psychology, 3*(AUG), 1–19. https://doi.org/10.3389/fpsyg.2012.00292 (cit. on pp. 25, 29, 103)

Heathcote, A., Popiel, S. J., & Mewhort, D. (1991). Analysis of response time distributions: An example using the stroop task. *Psychological bulletin, 109*(2), 340 (cit. on p. 19).

Heathcote, A., Wagenmakers, E.-J., & Brown, S. D. (2014). The falsifiability of actual decision-making models. (cit. on p. 34).

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience, 8*(June), 1–19. https://doi.org/10.3389/fnins.2014.00150 (cit. on pp. 20, 75, 125, 154)

Henmon, V. A. C. (1911). The relation of the time of a judgment to its accuracy. *Psychological Review, 18*(3), 186–201. https://doi.org/10.1037/h0074579 (cit. on p. 19)

Herz, D. M., Bogacz, R., & Brown, P. (2016). Neuroscience: Impaired decision-making in parkinson's disease. *Current Biology : CB, 26*, R671–R673. https://doi.org/10.1016/j.cub.2016.05.075 (cit. on p. 32)

Ho, T. C., Brown, S., & Serences, J. T. (2009). Domain general mechanisms of perceptual decision making in human cortex. *Journal of Neuroscience, 29*(27), 8675–8687 (cit. on pp. 29, 104, 112, 150).

Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology, 69*(4), 382 (cit. on pp. 19, 154).

JASP Team. (2020). JASP (Version 0.14)[Computer software]. https://jasp-stats.org/. (Cit. on pp. 79, 81)

Jepma, M., Wagenmakers, E.-J., Band, G. P. H., & Nieuwenhuis, S. (2009). The effects of accessory stimuli on information processing: evidence from electrophysiology and a diffusion model analysis. *Journal of Cognitive Neuroscience, 21*(5), 847–864. https://doi.org/10.1162/jocn.2009.21063 (cit. on p. 37)

Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review, 121*(1), 1–32. https://doi.org/10.1037/a0034190 (cit. on p. 34)

Kruschke, J. K. (2010). Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(5), 658–676 (cit. on p. 109).

Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. A. (2019). ArviZ a unified library for exploratory analysis of Bayesian models in Python. *The Journal of Open Source Software.* https://doi.org/10.21105/joss.01143 (cit. on pp. 48, 111)

Latimer, K. W., Yates, J. L., Meister, M. L. R., Huk, A. C., & Pillow, J. W. (2015). Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science, 349*(6244), 184–187. https://doi.org/10.1126/science.aaa4056 (cit. on p. 38)

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course.* https://doi.org/10.1017/CBO9781139087759. (Cit. on p. 80)

Leite, F. P., & Ratcliff, R. (2011). What cognitive processes drive response biases? A diffusion model analysis. *Judgment and Decision Making, 6*(7), 651–687 (cit. on pp. 82, 101).

Lerche, V., & Voss, A. (2018). Speed–accuracy manipulations and diffusion modeling: Lack of discriminant validity of the manipulation or of the parameter estimates? *Behavior Research Methods, 50*(6), 2568–2585 (cit. on p. 72).

Lerche, V., & Voss, A. (2019). Experimental validation of the diffusion model based on a slow response time paradigm. *Psychological Research, 83*(6), 1194–1209. https://doi.org/10.1007/s00426-017-0945-8 (cit. on p. 35)

Lerche, V., Voss, A., & Nagler, M. (2017). How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behavior Research Methods, 49*(2), 513–537. https://doi.org/10.3758/s13428-016-0740-2 (cit. on pp. 63, 79, 112)

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis, 100*(9), 1989–2001 (cit. on p. 184).

Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: Principles and practice.* SAGE publications. (Cit. on p. 21).

Liu, J., & Liu, Q. (2016). Use of the integrated profile for voluntary muscle activity detection using EMG signals with spurious background spikes: A study with incomplete spinal cord injury. *Biomedical Signal Processing and Control, 24*, 19–24. https://doi.org/10.1016/j.bspc.2015.09.004 (cit. on pp. 46, 108)

Loughnane, G. M., Newman, D. P., Bellgrove, M. A., Lalor, E. C., Kelly, S. P., & O'Connell, R. G. (2016). Target selection signals influence perceptual decisions by modulating the onset and rate of evidence accumulation. *Current Biology, 26*(4), 496–502 (cit. on pp. 132, 151).

Luce, R. (1986). Response Times: Their Role in Inferring Elementary Mental Organization. *Oxford University Press New York*, (3), 562. https://doi.org/10.1093/acprof:oso/9780195070019.001.0001 (cit. on pp. 15, 18, 19, 28, 40, 70, 129)

Ly, A., Boehm, U., Heathcote, A., Turner, B. M., Forstmann, B., Marsman, M., & Matzke, D. (2017). A flexible and efficient hierarchical bayesian approach to the exploration of individual differences in cognitive-model-based neuroscience. *Computational Models of Brain and Behavior*, 467–480 (cit. on pp. 63, 64, 139).

Martin, T., Huxlin, K. R., & Kavcic, V. (2010). Motion-onset visual evoked potentials predict performance during a global direction discrimination task. *Neuropsychologia, 48*(12), 3563–3572 (cit. on pp. 132, 151).

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-gaussian and shifted wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review, 16*(5), 798–817 (cit. on pp. 19, 62, 111, 185).

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86*, 287–330 (cit. on pp. 70, 130).

McClelland, J. L. (2009). The Place of Modeling in Cognitive Science. *Topics in Cognitive Science, 1*(1), 11–38. https://doi.org/10.1111/j.1756-8765.2008.01003.x (cit. on pp. 20, 21)

Meckler, C., Allain, S., Carbonnell, L., Hasbroucq, T., Burle, B., & Vidal, F. (2010). Motor inhibition and response expectancy: A laplacian erp study. *Biological Psychology, 85*(3), 386–392 (cit. on p. 83).

Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology, 26*(1-3), 3–67 (cit. on pp. 15, 36, 219).

Miller, J. (1988). Discrete and continuous models of human information processing: Theoretical distinctions and empirical results. *Acta Psychologica, 67*, 191–257 (cit. on p. 70).

Miller, J., Ulrich, R., & Rinkenauer, G. (1999). Effects of stimulus intensity on the lateralized readiness potential. *Journal of Experimental Psychology: Human Perception and Performance, 25*(5), 1454–1471. https://doi.org/10.1037/0096-1523.25.5.1454 (cit. on p. 56)

Mollon, J., & Perkins, A. (1996). Errors of judgment and Greenwich in 1796. *Nature,* (380), 101–102 (cit. on pp. 15, 219).

Moustafa, A. A., Kéri, S., Somlai, Z., Balsdon, T., Frydecka, D., Misiak, B., & White, C. (2015). Drift diffusion model of reward and punishment learning in schizophrenia: Modeling and experimental data. *Behavioural Brain Research, 291*, 147–154. https://doi.org/10.1016/j.bbr.2015.05.024 (cit. on p. 32)

Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the Brain: A Diffusion Model Analysis of Prior Probability and Potential Payoff. *Journal of Neuroscience, 32*(7), 2335–2343. https://doi.org/10.1523/JNEUROSCI.4156-11.2012 (cit. on pp. 82, 101)

Myung, J. I., & Pitt, M. A. (2016). Model Comparison in Psychology. *Journal of Mathematical Psychology, 5*, 1–40 (cit. on pp. 33, 35).

Nicenboim, B., Vasishth, S., Engelmann, F., & Suckow, K. (2018). Exploratory and confirmatory analyses in sentence processing: A case study of number interference in german. *Cognitive Science, 42*, 1075–1100 (cit. on pp. 48, 111).

Noorani, I., & Carpenter, R. (2013). Antisaccades as decisions: Later model predicts latency distributions and error responses. *European Journal of Neuroscience, 37*(2), 330–338 (cit. on pp. 18, 19).

Nunez, M. D., Gosai, A., Vandekerckhove, J., & Srinivasan, R. (2019). The latency of a visual evoked potential tracks the onset of decision making. *Neuroimage, 197*, 93–108 (cit. on pp. 74, 132, 135, 151).

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience, 15*(12), 1729–1735. https://doi.org/10.1038/nn.3248 (cit. on p. 34)

Oganian, Y., Froehlich, E., Schlickeiser, U., Hofmann, M. J., Heekeren, H. R., & Jacobs, A. M. (2016). Slower perception followed by faster lexical decision in longer words: A diffusion model analysis. *Frontiers in Psychology, 6*, 1958 (cit. on p. 29).

Oliphant, T. E. (2007). SciPy: Open source scientific tools for Python. *Computing in Science and Engineering, 9*, 10–20. https://doi.org/10.1109/MCSE.2007.58 (cit. on pp. 46, 86, 89, 108)

Ollman, R. (1966). Fast guesses in choice reaction time. *Psychonomic Science, 6*(4), 155–156 (cit. on p. 75).

Osman, A., Lou, L., Muller-Gethmann, H., Rinkenauer, G., Mattes, S., & Ulrich, R. (2000). Mechanisms of speed-accuracy tradeoff: Evidence from covert motor processes. *Biological Psychology, 51*(2-3), 173–199. https://doi.org/10.1016/S0301-0511(99)00045-9 (cit. on p. 44)

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005a). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision, 5*(5), 1. https://doi.org/10.1167/5.5.1 (cit. on pp. 16, 27, 85, 86)

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005b). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision, 5*(5), 1. https://doi.org/10.1167/5.5.1 (cit. on pp. 35, 56, 62, 86, 103, 104, 112, 161)

Palminteri, S., Wyart, V., & Koechlin, E. (2017). The importance of falsification in computational cognitive modeling. *Trends in Cognitive Sciences, 21*(6), 425–433 (cit. on p. 21).

Pe, M. L., Vandekerckhove, J., & Kuppens, P. (2013). A diffusion model account of the relationship between the emotional flanker task and rumination and depression. *Emotion, 13*(4), arXiv arXiv:1011.1669v3, 739–747. https://doi.org/10.1037/a0031628 (cit. on p. 32)

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017 (cit. on pp. 45, 84, 85, 106, 107)

Perfors, A. (2020). *Generating theories is hard, and none of our theories are good enough* [Virtual MathPsych 2020]. Virtual MathPsych 2020. https://mathpsych.org/presentation/109. (Cit. on p. 34)

Piéron, H. (1913). II. Recherches sur les lois de variation des temps de latence sensorielle en fonction des intensités excitatrices. *L'Année Psychologique, 20*(1), 17–96. https://doi.org/10.3406/psy.1913.4294 (cit. on p. 16)

Piéron, H. (1914). Iii. henri piéron. une oeuvre psychologique de guerre. l'examen des aviateurs. *L'Année Psychologique, 21*(1), 237–252 (cit. on p. 16).

Pieters, J. P. (1983). Sternberg's additive factor method and underlying psychological processes: Some theoretical considerations. *Psychological Bulletin, 93*, 411–426 (cit. on p. 70).

Poincaré, H. (1908). *La science et l'hypothèse*. E. Flammarion. (Cit. on p. 6).

Posner, M. I. (1978). Chronometric Explorations of Mind. Lawrence Erlbaum. http://books.google.com/books?id=tQwHAAAACAAJ%7B%5C&%7Dprintsec=frontcover%7B%5C%%7D5Cnpapers2://publication/uuid/775A7210-14E1-4951-AB23-C49663FEDD23. (Cit. on pp. 15, 219)

Possamaı, C.-A., Burle, B., Osman, A., & Hasbroucq, T. (2002). Partial advance information, number of alternatives, and motor processes: An electromyographic study. *Acta Psychologica, 111*(1), 125–139 (cit. on pp. 39, 84, 154).

Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review, 117*(4), 1113–43. https://doi.org/10.1037/a0020311 (cit. on p. 34)

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The Hare and the Tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1226–1243. https://doi.org/10.1037/a0036801 (cit. on pp. 35, 63, 112, 187)

Rafiei, F., & Rahnev, D. (2020). Does the diffusion model account for the effects of speed-accuracy tradeoff on response times? (Cit. on p. 125).

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological bulletin, 86*(3), 446–461. https://doi.org/10.1037/0033-2909.86.3.446 (cit. on p. 192)

Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception & Psychophysics, 65*(4), 523–535. https://doi.org/10.3758/BF03194580 (cit. on p. 32)

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59 (cit. on pp. 24–26, 29, 44, 103, 129, 153, 222)

Ratcliff, R. (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin & Review, 9*(2), 278–291 (cit. on p. 33).

Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive Psychology, 53*(3), 195–237 (cit. on pp. 35, 62, 104, 112, 154, 161).

Ratcliff, R. (2013). Parameter variability and distributional assumptions in the diffusion model. *Psychological Review, 120*(1), arXiv NIHMS150003, 281–292. https://doi.org/10.1037/a0030775 (cit. on pp. 30, 34)

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation, 20*(4), arXiv NIHMS150003, 873–922. https://doi.org/10.1162/neco.2008.12-06-420 (cit. on pp. 28, 29, 31, 44, 61, 65, 73, 82, 88, 101, 111, 151, 154, 161, 197, 198)

Ratcliff, R., Perea, M., Colangelo, A., & Buchanan, L. (2004). A diffusion model account of normal and impaired readers. *Brain and Cognition, 55*(2), 374–382. https://doi.org/10.1016/j.bandc.2004.02.051 (cit. on p. 32)

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*(5), 347–356 (cit. on pp. 30, 73, 103).

Ratcliff, R., & Smith, P. L. (2010). Perceptual Discrimination in Static and Dynamic Noise: The Temporal Relation Between Perceptual Encoding and Decision Making. *Journal of Experimental Psychology: General, 139*(1), 70–94. https://doi.org/10.1037/a0018128 (cit. on p. 130)

Ratcliff, R., Smith, P. L., Brown, S. D., & Mckoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007 (cit. on pp. 32, 129)

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A Diffusion Model Analysis of the Effects of Aging in the Lexical-Decision Task. *Psychology and Aging, 19*(2), arXiv NIHMS150003, 278–289. https://doi.org/10.1037/0882-7974.19.2.278 (cit. on p. 32)

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging, 16*(2), 323–341. https://doi.org/10.1037//0882-7974.16.2.323 (cit. on p. 32)

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review, 9*(3), arXiv NIHMS150003, 438–481. https://doi.org/10.3758/BF03196302 (cit. on pp. 29, 30, 55, 79, 104)

Ratcliff, R., Voskuilen, C., & Teodorescu, A. (2018). Modeling 2-alternative forced-choice tasks: Accounting for both magnitude and difference effects. *Cognitive Psychology, 103*, 1–22 (cit. on pp. 34, 105).

Requin, J., Riehle, A., & Seal, J. (1988). Neuronal activity and information processing in motor control: From stages to continuous flow. *Biological Psychology, 26*, 179–198 (cit. on p. 70).

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature, 461*(7261), 263–266. https://doi.org/10.1038/nature08275 (cit. on pp. 71, 128, 227)

Reynaud, A., Masson, G. S., & Chavane, F. (2012). Dynamics of local input normalization result from balanced short-and long-range intracortical interactions in area v1. *Journal of Neuroscience, 32*(36), 12558–12569 (cit. on pp. 124, 161, 195).

Rinkenauer, G., Osman, A., Ulrich, R., Muller-Gethmann, H., & Mattes, S. (2004). On the locus of speed-accuracy trade-off in reaction time: inferences from the lateralized readiness potential. *Journal of Experimental Psychology: General, 133*(2), 261–82. https://doi.org/10.1037/0096-3445.133.2.261 (cit. on p. 44)

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107*(2), 358–367. https://doi.org/10.1037//0033-295X.107.2.358 (cit. on p. 33)

Rochet, N., Spieser, L., Casini, L., Hasbroucq, T., & Burle, B. (2014). Detecting and correcting partial errors: Evidence for efficient control without conscious access. *Cognitive, Affective & Behavioral Neuroscience, 14*(3), 970–82. https://doi.org/10.3758/s13415-013-0232-0 (cit. on pp. 48, 55, 56, 72, 89, 99)

Roger, C., Núñez Castellar, E., Pourtois, G., & Fias, W. (2014). Changing your mind before it is too late: The electrophysiological correlates of online error correction during response selection. *Psychophysiology, 51*, 746–760. https://doi.org/10.1111/psyp.12224 (cit. on p. 55)

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of*

*Neuroscience, 22*(21), 9475–9489. https://doi.org/10.1016/S0377-2217(02) 00363-6 (cit. on pp. 74, 131)

Sanders, A. (1983). Towards a model of stress and human performance. *Acta Psychologica, 53*(1), 61–97. https://doi.org/https://doi.org/10.1016/0001-6918(83)90016-1 (cit. on pp. 56, 72, 155, 225)

Santello, M., & Mcdonagh, M. J. (1998). The control of timing and amplitude of EMG activity in landing movements in humans. *Experimental Physiology, 83*(6), 857–874. https://doi.org/10.1113/expphysiol.1998.sp004165 (cit. on pp. 46, 108)

Schad, D. J., Betancourt, M., & Vasishth, S. (in press). Toward a principled bayesian workflow in cognitive science. *Psychological Methods* (cit. on pp. 91, 184, 191).

Schall, J. D. (2004). On Building a Bridge Between Brain and Behavior. *Annual Review of Psychology, 55*(1), 23–50. https://doi.org/10.1146/annurev.psych.55.090902. 141907 (cit. on pp. 37, 70)

Schall, J. D. (2019). Accumulators, neurons, and response time. *Trends in Neurosciences, 42*(12), 848–860 (cit. on pp. 17, 37, 70, 220).

Schmidgen, H. (2002). Of frogs and men: The origins of psychophysiological time experiments, 1850–1865. *Endeavour, 26*(4), 142–148. https://doi.org/https://doi.org/10.1016/S0160-9327(02)01466-7 (cit. on pp. 16, 220)

Schmied, A., Vedel, J. P., & Pagni, S. (1994). Human spinal lateralization assessed from motoneurone synchronization: dependence on handedness and motor unit type. *The Journal of Physiology, 480*, 369–87. https://doi.org/10.1113/jphysiol. 1994.sp020367 (cit. on pp. 48, 57, 61)

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python, In *Proceedings of the 9th python in science conference*. Scipy. (Cit. on p. 186).

Selen, L. P. J., Shadlen, M. N., & Wolpert, D. M. (2012). Deliberation in the Motor System: Reflex Gains Track Evolving Evidence Leading to a Decision. *Journal of Neuroscience, 32*(7), 2276–2286. https://doi.org/10.1523/JNEUROSCI.5273-11.2012 (cit. on pp. 128, 227)

Servant, M., Montagnini, A., & Burle, B. (2014). Conflict tasks and the diffusion framework: Insight in model constraints based on psychological laws. *Cognitive Psychology, 72*, 162–195. https://doi.org/10.1016/j.cogpsych.2014.03.002 (cit. on p. 16)

Servant, M., White, C., Montagnini, A., & Burle, B. (2015). Using Covert Response Activation to Test Latent Assumptions of Formal Decision-Making Models in Humans. *Journal of Neuroscience, 35*(28), 10371–10385. https://doi.org/10. 1523/JNEUROSCI.0078-15.2015 (cit. on pp. 40, 46, 56)

Servant, M., White, C., Montagnini, A., & Burle, B. (2016). Linking Theoretical Decision-making Mechanisms in the Simon Task with Electrophysiological Data: A Model-based Neuroscience Study in Humans. *Journal of Cognitive Neuroscience, 28*(10), arXiv 1511.04103, 1501–1521. https://doi.org/10.1162/ jocna00989 (cit. on pp. 37, 39, 44, 46, 104, 159)

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology, 86*(4), 1916–36. http://www.ncbi.nlm.nih.gov/pubmed/11600651 (cit. on p. 34)

Sigman, M., & Dehaene, S. (2005). Parsing a cognitive task: A characterization of the mind's bottleneck. *PLoS biology, 3*(2), e37 (cit. on p. 19).

Singmann, H., Brown, S., Gretton, M., Heathcote, A., Voss, A., Voss, J., Terry, A., & Singmann, M. H. (2016). Package 'rtdists' (cit. on p. 139).

Śmigasiewicz, K., Ambrosi, S., Blaye, A., & Burle, B. (2020). Inhibiting errors while they are produced: Direct evidence for error monitoring and inhibitory control in children. *Developmental Cognitive Neuroscience, 41*, 100742 (cit. on pp. 48, 52, 56, 72, 99, 155, 225).

Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review, 102*(3), 567–593. https://doi.org/10.1037/0033-295X.102.3.567 (cit. on pp. 40, 56, 153)

Smith, P. L., & Lilburn, S. D. (2020a). Vision for the blind: visual psychophysics and blinded inference for decision models. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-020-01742-7 (cit. on p. 162)

Smith, P. L., & Lilburn, S. D. (2020b). Vision for the blind: Visual psychophysics and blinded inference for decision models. *Psychonomic Bulletin & Review* (cit. on pp. 36, 72, 130).

Smith, P. L., & Ratcliff, R. (2015). *An introduction to the diffusion model of decision making*. https://doi.org/10.1007/978-1-4939-2236-9_3. (Cit. on p. 24)

Smith, P. L., Ratcliff, R., & McKoon, G. (2014). The diffusion model is not a deterministic growth model: Comment on jones and dzhafarov (2014). (cit. on p. 34).

Spieser, L., Servant, M., Hasbroucq, T., & Burle, B. (2017). Beyond decision! Motor contribution to speed–accuracy trade-off in decision-making. *Psychonomic Bulletin & Review, 24*(3), 950–956. https://doi.org/10.3758/s13423-016-1172-9 (cit. on pp. 39, 44, 55, 71, 72, 101, 105, 155)

Stan Development Team. (2018, January 16). *Pystan: The python interface to stan* (Version 2.17.1.0). http://mc-stan.org. (Cit. on pp. 48, 111)

Starns, J. J., & Ma, Q. (2018). Response biases in simple decision making: Faster decision making, faster response execution, or both? *Psychonomic Bulletin and Review, 25*(4), 1535–1541. https://doi.org/10.3758/s13423-017-1358-9 (cit. on pp. 82–84, 99)

Starns, J. J., & Ratcliff, R. (2014). Validating the unequal-variance assumption in recognition memory using response time distributions instead of roc functions: A diffusion model analysis. *Journal of Memory and Language, 70*, 36–52 (cit. on p. 35).

Starns, J. J., Ratcliff, R., & McKoon, G. (2012). Evaluating the unequal-variance and dual-process explanations of zROC slopes with response time data and the diffusion model. *Cognitive Psychology, 64*(1-2), 1–34 (cit. on p. 187).

Steinemann, N. A., O'Connell, R. G., & Kelly, S. P. (2018). Decisions are expedited through multiple neural adjustments spanning the sensorimotor hierarchy. *Nature Communications, 9*(1), 3627 (cit. on pp. 44, 55, 71, 105, 133).

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, 30*, 276–315. https://doi.org/10.1016/0001-6918(69)90055-9 (cit. on p. 70)

Stine, G. M., Zylberberg, A., Ditterich, J., & Shadlen, M. N. (2020). Differentiating between integration and non-integration strategies in perceptual decision making (V. Wyart, M. J. Frank, V. Wyart, & M. Usher, Eds.). *eLife, 9*, e55365. https://doi.org/10.7554/eLife.55365 (cit. on pp. 22, 34)

Stone, M. (1960). Models for choice-reaction time. *Psychometrika, 25*(3), 251–260. https://doi.org/10.1007/BF02289729 (cit. on pp. 22–26, 28, 29, 52, 70, 103)

Strijkers, K., Costa, A., & Thierry, G. (2009). Tracking Lexical Access in Speech Production: Electrophysiological Correlates of Word Frequency and Cognate Effects. *Cerebral Cortex, 20*(4), https://academic.oup.com/cercor/article-pdf/20/4/912/869382/bhp153.p 912–928. https://doi.org/10.1093/cercor/bhp153 (cit. on p. 131)

Tandonnet, C., Burle, B., Vidal, F., & Hasbroucq, T. (2003). The influence of time preparation on motor processes assessed by surface laplacian estimation. *Clinical Neurophysiology, 114*(12), 2376–2384 (cit. on p. 39).

Tandonnet, C., Burle, B., Vidal, F., & Hasbroucq, T. (2006). Knowing when to respond and the efficiency of the cortical motor command: A laplacian erp study. *Brain Research, 1109*(1), 158–163. https://doi.org/10.1016/j.brainres.2006.06.052 (cit. on p. 39)

Tandonnet, C., Davranche, K., Meynier, C., Burle, B., Vidal, F., & Hasbroucq, T. (2012). How does temporal preparation speed up response implementation in choice tasks? evidence for an early cortical activation. *Psychophysiology, 49*(2), https://onlinelibrary.wiley 8986.2011.01301.x, 252–260. https://doi.org/https://doi.org/10.1111/j.1469-8986.2011.01301.x (cit. on p. 159)

Teller, D. Y. (1984). Linking propositions. *Vision Research, 24*(10), 1233–1246. https://doi.org/10.1016/0042-6989(84)90178-0 (cit. on pp. 37, 70)

Todd, A. R., Johnson, D. J., Lassetter, B., Neel, R., Simpson, A. J., & Cesario, J. (2020). Category Salience and Racial Bias in Weapon Identification: A Diffusion Modeling Approach. *Journal of Personality and Social Psychology*, (July). https://doi.org/10.1037/pspi0000279 (cit. on p. 32)

Towal, R. B., Mormann, M., & Koch, C. (2013). Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. *Proceedings of the National Academy of Sciences, 110*(40), E3858–E3867 (cit. on p. 32).

Turner, B. M., Rodriguez, C. A., Norcia, T. M., McClure, S. M., & Steyvers, M. (2016). Why more is better: Simultaneous modeling of EEG, fMRI, and behavioral data. *NeuroImage, 128*, 96–115. https://doi.org/10.1016/j.neuroimage.2015.12.030 (cit. on p. 37)

Turner, B. M., van Maanen, L., & Forstmann, B. U. (2015). Informing cognitive abstractions through neuroimaging: the neural drift diffusion model. *Psychological Review, 122*(2), 312–336. https://doi.org/10.1037/a0038894 (cit. on p. 70)

Ulrich, R., & Stapf, K. H. (1984). A double-response paradigm to study stimulus intensity effects upon the motor system in simple reaction time experiments. *Perception & Psychophysics, 36*(6), 545–558. https://doi.org/10.3758/BF03207515 (cit. on p. 56)

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model., *108*(3), 550–592. https://doi.org/10.1037/0033-295X.108.3.550 (cit. on pp. 25, 103)

Valéry, P. (1942). *Mauvaises pensées et autres.* Gallimard Paris, France. (Cit. on p. 6).

Van Den Brink, R. L., Murphy, P. R., Desender, K., De Ru, N., & Nieuwenhuis, S. (2020). Temporal expectation hastens sensory encoding but does not affect evidence quality. *BioRxiv* (cit. on p. 37).

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review, 7*(3), 424–465. https://doi.org/10.3758/BF03214357 (cit. on p. 18)

van Maanen, L., Grasman, R. P. P. P., Forstmann, B. U., & Wagenmakers, E. J. (2012). Piéron's law and optimal behavior in perceptual decision-making. *Frontiers in Neuroscience, 5*(JAN), 1–15. https://doi.org/10.3389/fnins.2011.00143 (cit. on p. 16)

van Maanen, L., van der Mijn, R., van Beurden, M. H., Roijendijk, L. M., Kingma, B. R., Miletić, S., & van Rijn, H. (2019). Core body temperature speeds up temporal processing and choice behavior under deadlines. *Scientific Reports, 9*(1), 1–12 (cit. on p. 32).

van Ravenzwaaij, D., Provost, A., & Brown, S. D. (2017). A confirmatory approach for integrating neural and behavioral data into a single model. *Journal of Mathematical Psychology, 76*, 131–141 (cit. on p. 70).

Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review, 14*(6), 1011–1026 (cit. on pp. 49, 54, 75, 113).

Verdonck, S., Loossens, T., & Philiastides, M. G. (2020). The leaky integrating threshold and its impact on evidence accumulation models of choice rt (cit. on p. 125).

Verdonck, S., & Tuerlinckx, F. (2016). Factoring out nondecision time in choice reaction time data: Theory and implications. *Psychological Review, 123*(2), 208–218. https://doi.org/10.1037/rev0000019 (cit. on p. 34)

Vidal, F., Burle, B., Grapperon, J., & Hasbroucq, T. (2011). An erp study of cognitive architecture and the insertion of mental processes: Donders revisited. *Psychophysiology, 48*(9), 1242–1251 (cit. on p. 17).

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1436 (cit. on p. 131).

Voss, A., & Voss, J. (2007). Fast-dm: a free program for efficient diffusion model analysis. *Behavior Research Methods, 39*(4), 767–775. https://doi.org/10.3758/BF03192967 (cit. on pp. 31, 79, 88)

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: an empirical validation. *Memory & cognition, 32*(7), 1206–1220. https://doi.org/10.3758/BF03196893 (cit. on pp. 29, 35, 62, 82, 101, 104, 105, 112, 150, 161)

Voss, A., Voss, J., & Klauer, K. C. (2010). Separating response-execution bias from decision bias: Arguments for an additional parameter in Ratcliff's diffusion model. *British Journal of Mathematical and Statistical Psychology, 63*(3), 539–555. https://doi.org/10.1348/000711009X477581 (cit. on pp. 78, 80–83, 101, 157, 226)

Voss, A., Voss, J., & Lerche, V. (2015). Assessing cognitive processes with diffusion model analyses: A tutorial based on fast-dm-30. *Frontiers in Psychology, 6*, 336 (cit. on p. 79).

Wagenmakers, E. J., Grasman, R. P., & Molenaar, P. C. (2005). On the relation between the mean and the variance of a diffusion model response time distribution. *Journal of Mathematical Psychology, 49*(3), 195–204. https://doi.org/10.1016/j.jmp.2005.02.003 (cit. on pp. 19, 119, 138)

Wagenmakers, E.-J. (2009). Methodological and empirical developments for the ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology, 21*(5), 641–671 (cit. on p. 129).

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review, 114*(3), 830–841. https://doi.org/10.1037/0033-295X.114.3.830 (cit. on pp. 18, 19)

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language, 58*(1), 140–159. https://doi.org/https://doi.org/10.1016/j.jml.2007.04.006 (cit. on p. 29)

Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin and Review, 14*(1), 3–22. https://doi.org/10.3758/BF03194023 (cit. on pp. 29, 132)

Wald, A. (1947). Foundations of a General Theory of Sequential Decision Functions. *Econometrica, 15*(4), 279. https://doi.org/10.2307/1905331 (cit. on pp. 22, 222)

Weinberg, I. C. (2017). *Expectation in motor planning and execution* (Doctoral dissertation August). (Cit. on pp. 83, 99).

Weindel, G., Anders, R., Alario, F.-X., & Burle, B. (2020). Assessing model-based inferences in decision making with single-trial response time decomposition. OSF. osf.io/frhj9

White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Anxiety enhances threat processing without competition among multiple inputs: A diffusion model

analysis. *Emotion, 10*(5), 662–677. https://doi.org/10.1037/a0019474 (cit. on p. 32)

Wiecki, T. V., Sofer, I., & Frank, M. J. (2016). Hddm 0.6. 0 documentation. (Cit. on pp. 63, 111).

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). Hddm: Hierarchical bayesian estimation of the drift-diffusion model in python. *Frontiers in Neuroinformatics, 7*, 14 (cit. on pp. 44, 62, 111, 113, 185).

Wundt, W. M. (1874). *Grundzüge der physiologischen psychologie* (Vol. 1). W. Engelman. (Cit. on p. 16).

Wyart, V., & Tallon-Baudry, C. (2009). How ongoing fluctuations in human visual cortex predict perceptual awareness: Baseline shift versus decision bias. *Journal of Neuroscience, 29*(27), 8715–8725 (cit. on p. 36).

Zhang, Q., Walsh, M. M., & Anderson, J. R. (2017). The effects of probe similarity on retrieval and comparison processes in associative recognition. *Journal of Cognitive Neuroscience, 29*(2), 352–367 (cit. on p. 132).

# Appendices

# A. Appendix of chapter 2

## A.1. EMG traces

Figure .1 provides examples of the analyzed EMG signal for the first trial of 10 participants.
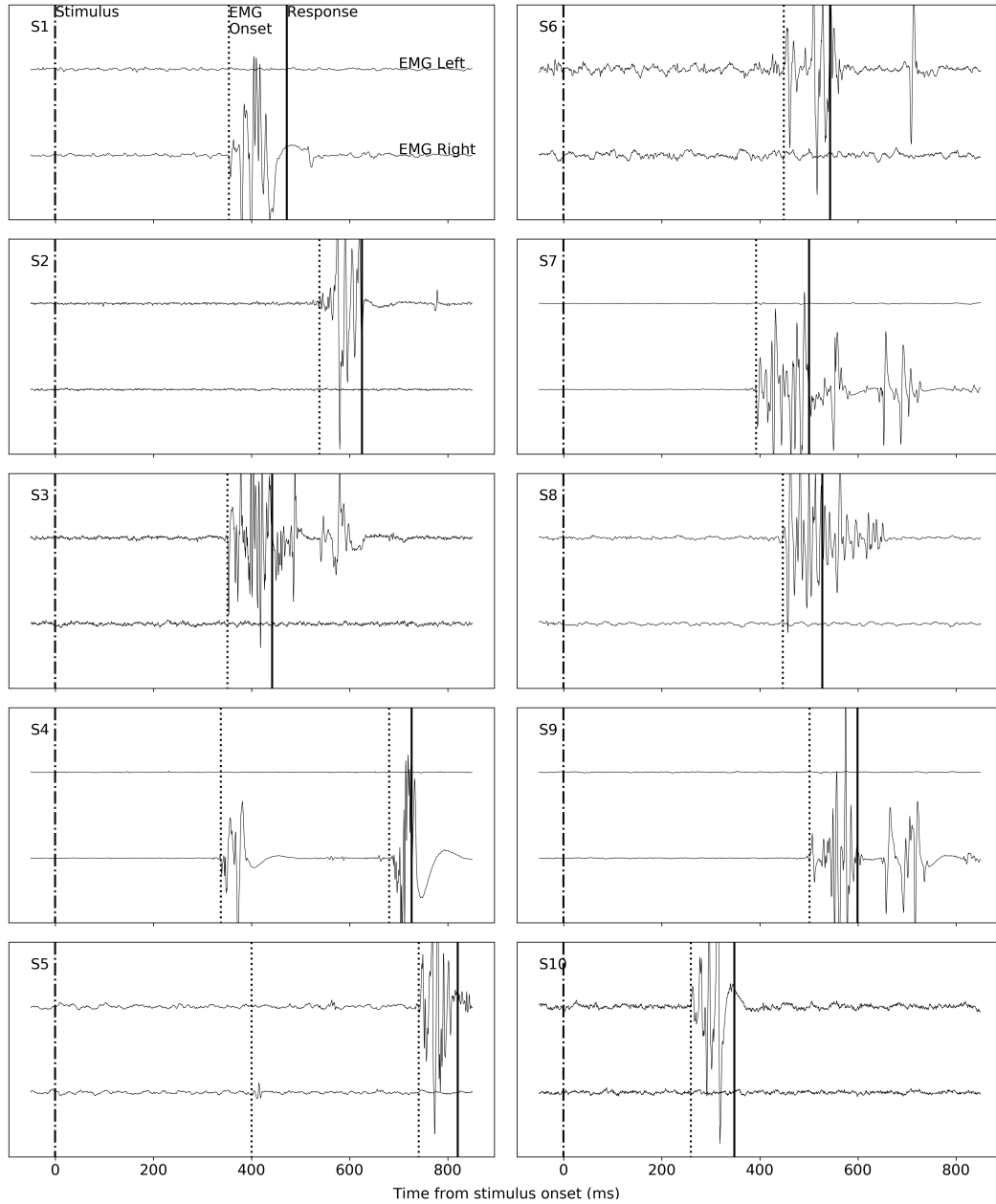
Figure .1.: Graphic representation of the recorded EMG signal for the first trial of 10 participants in Experiment 1. Time is centered at stimulus onset (dot-dashed line), EMG onset(s) and recorded response are shown respectively with dotted line(s) and plain line. Left and right EMG signals are displayed respectively on the upper-and lower half of each panel. For graphical purpose values are z-scores of the amplitude (usually displayed in $\mu$V), the data was read in Python using the MNE module (Gramfort, Luessi, Larson, et al., 2013).

## A.2. Choice of priors

**Linear Mixed Models on chronometric data**   In agreement with the suggestion made by Schad, Betancourt, and Vasishth, in press for each applied LMM we selected reasonable priors and estimated the predicted values from the specified priors and whether these values are in the range of expected values (*e.g.* most likely values for mean $RT$ are sub second given our design, hence a factor effect of 100 ms on $RT$ is more likely than a factor effect of 1000 ms). It should be noted that given the large number of observations in both experiments, the chosen priors have little influence as long as they do not exclude plausible values. All priors for the chronometric variables are expressed on the log scale.

We defined the following priors for the populations intercepts ($\mu_\alpha$) of the LMMs on $RT$, $PMT$ and $MT$ :

$$\mu_{\alpha_{RT}} \sim \mathcal{N}(\mu = 6.15, \sigma = 0.3)$$

$$\mu_{\alpha_{PMT}} \sim \mathcal{N}(\mu = 5.87, \sigma = 0.3)$$

$$\mu_{\alpha_{MT}} \sim \mathcal{N}(\mu = 4.67, \sigma = 0.3)$$

The population slope parameters ($\mu_{\beta_x}$), as well as the random effects on the intercept ($\sigma_\alpha$) and the slopes ($\sigma_{\beta_x}$), and the residual standard deviation ($\sigma_r$), were all given the same prior irrespective of the chronometric data fitted :

$$\mu_{\beta_x}, \sigma_\alpha, \sigma_{\beta_x}, \sigma_r \sim \mathcal{N}(\mu = 0, \sigma = 0.4)$$

Allowing for a wide range of effect sizes, random effects and residual standard deviation with values close to 0 more likely than large values.

Finally, as Schad, Betancourt, and Vasishth, in press we chose an LKJ prior, so-called because it was first described by Lewandowski, Kurowicka, and Joe, 2009, with a value of 2 for the correlation matrix of the by-participant parameter adjustments, firstly because we do not expect strong correlations among participant specific parameter adjustments and because a weakly informative priors on the correlation matrix helps estimating full random-effect structure as advocated by Barr, 2013.

**Linear Mixed Models on $PMT$-$MT$ correlation values**   The parameters of the LMM on the correlation values between $PMT$ and $MT$ were all given the same prior :

$$\mu_\alpha, \sigma_\alpha, \mu_{\beta_{SAT}}, \sigma_{\beta_{SAT}}, \sigma_r \sim \mathcal{N}(\mu = 0, \sigma = 0.3)$$

Therefore allowing all possible values of intercept ($\mu_\alpha$), slopes of SAT ($\mu_{\beta_{SAT}}$), random effect of the intercept ($\sigma_\alpha$) and the slope ($\sigma_{\beta_{SAT}}$), and residual standard deviation ($\sigma_r$) given that correlation values are contained between -1 and 1. In agreement with the defined normal distribution, values close to 0 are more likely than extreme values coherent with what we expect on the correlation value between $PMT$ and $MT$.

|  | Exp. 1 | | | | Exp. 2 | | | |
|---|---|---|---|---|---|---|---|---|
|  | mean | SD | 2.5% | 97.5% | mean | SD | 2.5% | 97.5% |
| intercept | 0.84 | 0.02 | 0.80 | 0.87 | 0.75 | 0.06 | 0.63 | 0.86 |
| SAT | 0.06 | 0.01 | 0.03 | 0.08 | 0.12 | 0.03 | 0.07 | 0.18 |
| Contrast | 0.15 | 0.02 | 0.11 | 0.19 | 0.18 | 0.03 | 0.13 | 0.23 |
| Resp. Side | 0.03 | 0.01 | 0.01 | 0.06 | -0.01 | 0.03 | -0.07 | 0.05 |
| SAT × Contrast | 0.08 | 0.02 | 0.04 | 0.12 | 0.17 | 0.04 | 0.10 | 0.25 |

Table .1.: Results of the generalized LMM model on Accuracy in Experiment 1 (left column) and in Experiment 2 (right column). As for the chronometric LMMs, the fitting was performed on a transformed scale (logit) but parameters where back-transformed at each MCMC iteration (by taking the inverse logit). Mean, SD and 2.5 and 97.5% summarize the posterior distribution of the parameters.

**HDDM**   The informative piors used for the fit of the Hierarchical drift diffusion model are given by Wiecki, Sofer, and Frank, 2013 based on the analysis of range of plausible values done by Matzke and Wagenmakers, 2009.

## A.3. Accuracy analysis

To analyze accuracy we fitted a General LMM with contrast, SAT, response side and the interaction between contrast and SAT as fixed effects. Random effects included the random intercept by participants and random slopes for all predictors. In this analysis for Experiment 1, we see a main effect of contrast, SAT instructions and response side. The interaction between SAT and contrast was also significant. Experiment 2 replicates these effects except the response side effect (see Table .1). The presence of an effect for response side only in Experiment 1 is surprising. This could be linked to the observed effect on $MT$ in Table 2.1, if right responses are faster they might also be less costly to execute, hence inducing a small bias in these responses.

## A.4. Scatter plot of the quantile correlation

To convey a sense of the correlation values contained in Figure 2.3, we present scatter plots for the first five participants of Experiment 1 in Figure .2.
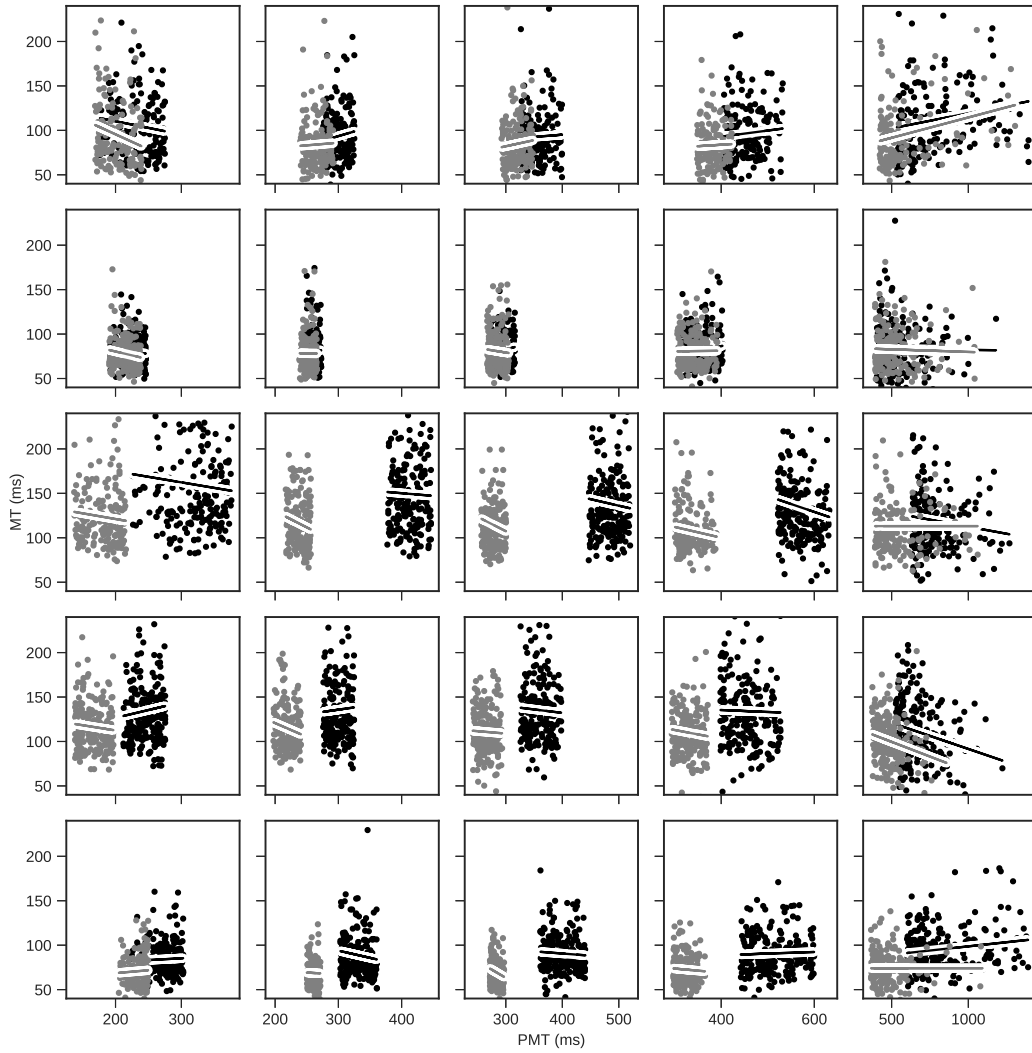
Figure .2.: Scatter plots of *PMT* versus *MT* for the 5 first participants in Experiment 1 (participant 1 to 5 from top to bottom) across the 5 quantiles (.1, .3, .5, .7 and .9 quantiles, from left to right) either with the speed (gray) or accuracy (black) instructions. In order to give a sense of the linear relationship between the pairs of *PMT* and *MT*, each scatter plot is associated with a line drawn from the parameters of a linear regression of *MT* over *PMT* estimated using the ordinary least squares method as implemented in the `statsmodel` python package (Seabold & Perktold, 2010).

## A.5. Mean table of the chronometric variables

In Table .2 we report the mean values for correct response, for each chronometric variable, in the experimental cells combining SAT and contrast levels.

| Exp. | Data | SAT | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|
| 1 | RT | Acc. | 636 | 621 | 571 | 505 | 470 |
| | | Speed | 457 | 451 | 437 | 410 | 388 |
| | PMT | Acc. | 512 | 500 | 448 | 387 | 355 |
| | | Speed | 355 | 347 | 335 | 311 | 290 |
| | MT | Acc. | 124 | 121 | 122 | 118 | 116 |
| | | Speed | 102 | 103 | 101 | 99 | 99 |
| 2 | RT | Acc. | 628 | | 551 | 487 | |
| | | Speed | 407 | | 393 | 366 | |
| | PMT | Acc. | 545 | | 470 | 408 | |
| | | Speed | 334 | | 322 | 299 | |
| | MT | Acc. | 82 | | 81 | 79 | |
| | | Speed | 73 | | 71 | 68 | |

Table .2.: Means (in ms) by experiment, Speed-Accuracy Trade-off and contrast condition for each chronometric data for correct responses only.

## A.6. Model Selection Results

The models where a modulation of the starting point and a response side effect on $T_{er}$ were simultaneously estimated did not reach convergence, despite a high amount of MCMC iterations. This can be due to the difficulty to separate a "cognitive" bias effect from a motor bias (i.e. right hand is preferred or right stimulus is preferred) when response side and stimulus location are not counterbalanced.

For the remaining models, including a modulation of the starting point always lead to lower criteria, whichever was chosen, for both experiments. The models allowing the drift rate to vary with SAT conditions were systematically selected when considering the DIC criterion, coherent with Rae, Heathcote, Donkin, Averell, and Brown, 2014 (but see Starns, Ratcliff, and McKoon, 2012 for an interpretation of SAT effects on drift rate). However, when considering the BPIC and the fit on data from Experiment 1, the models estimating only one drift rate across the SAT conditions were preferred. We also note that, when considering the models with a fixed starting point only (models 1 to 4 in Table .4), the fit always improved when $T_{er}$ was free to vary between left and right responses, for both experiments. This result is surprising because we do not observe an effect of response side on $MT$ in the second experiment. Given the issue raised above concerning the possible confound between cognitive and motor bias it could be that this result is simply caused by not allowing the starting point to vary, hence forcing the cognitive bias to be captured by the non-decision time.

Overall, only the BPIC suggests a different model for Experiment 1 than for Experiment 2. For the sake of simplicity, we chose to use the same model for both experiments, namely model 6 in Table .4, with starting point estimation and drift free

to vary between stimulus strength level and SAT conditions.

| | $a$ | $v$ | $T_{er}$ | $z$ | $s_z$ | $s_v$ | $s_t$ | $DIC_{Exp.1}$ | $DIC_{Exp.2}$ | $BPIC_{Exp.1}$ | $BPIC_{Exp.2}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | SAT | contrast | SAT | 0 | SAT | 1 | 1 | -21927 | -31552 | -21802 | -31438 |
| 2 | SAT | SAT × contrast | SAT | 0 | SAT | 1 | 1 | -21961 | -31722 | -21782 | -31568 |
| 3 | SAT | contrast | SAT × Response | 0 | SAT | 1 | 1 | -22295 | -31864 | -22143 | -31721 |
| 4 | SAT | SAT × contrast | SAT × Response | 0 | SAT | 1 | 1 | -22329 | -32031 | -22116 | -31842 |
| 5 | SAT | contrast | SAT | 1 | SAT | 1 | 1 | -22898 | -32664 | **-22761** | -32538 |
| 6 | SAT | SAT × contrast | SAT | 1 | SAT | 1 | 1 | **-22931** | **-32844** | -22732 | **-32671** |
| 7 | SAT | contrast | SAT × Response | 1 | SAT | 1 | 1 | – | – | – | – |
| 8 | SAT | SAT × contrast | SAT × Response | 1 | SAT | 1 | 1 | – | – | – | – |

Table .3.: Model comparison for models fitted to data from Experiment 1 and Experiment 2.
$a$: boundary parameter; $v$: drift rate, $T_{er}$: non-decision time; $z$: starting point of the accumulation; $s_z$, $s_v$, and $s_t$ are, respectively, the inter-trial variability of the starting point, of the drift rate and of the non-decision time. 0 and 1 indicates whether a parameter was estimated or not.

# B. Appendix of chapter 3

## B.1. PRDM fits

See Figure .3.
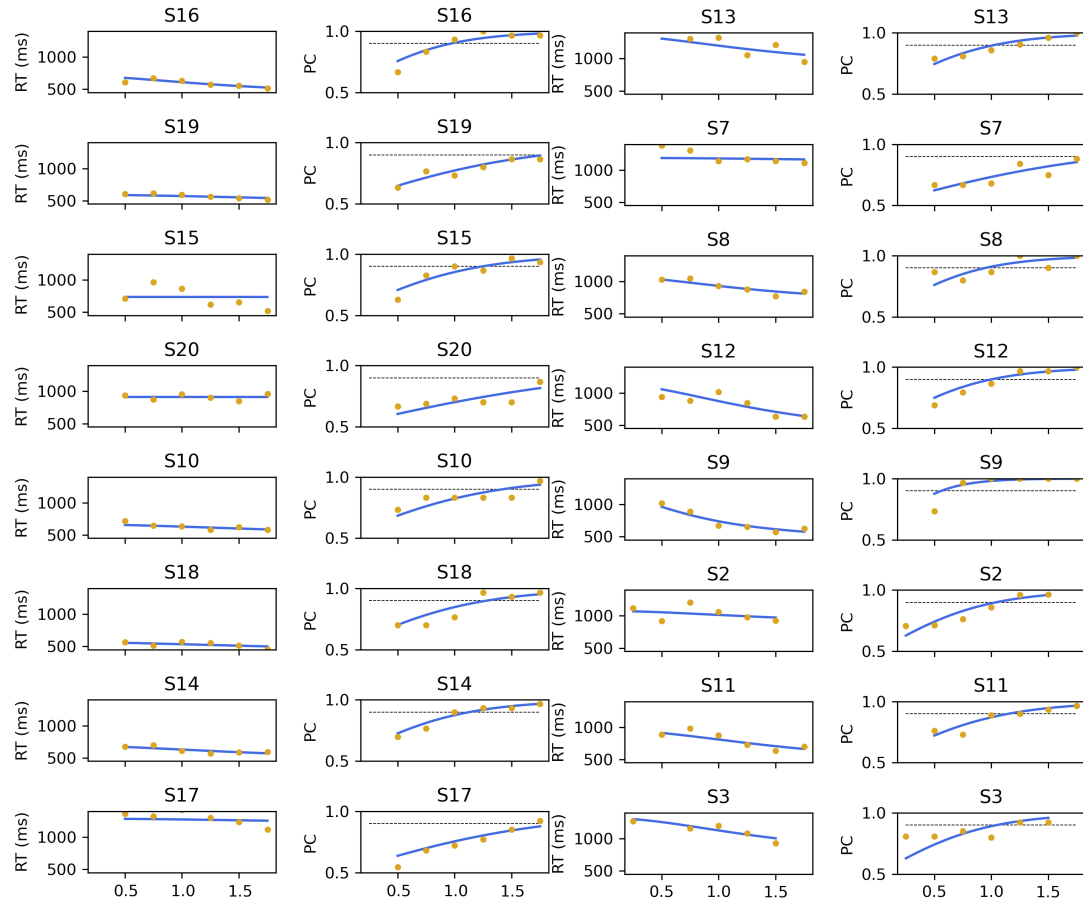


Figure .3.: Individual fit on the joint distribution of mean RT and proportion correct for each deviation angle (x-axis) of the PRDM for each retained participant in the experiment. The dots represent the observed data and the blue line the model fits. Aimed accuracy of 0.90 is shown by the dashed line in the plots showing the proportion of correct responses. For participant S7, S17 and S20 the orientation producing 90% accuracy was probably over the selected deviation levels, out of simplicity we kept these participants and chose to fix the deviation angle at the maximum level tested 1.75. Note that before participant S7 the deviation angle chosen were from 0.25 to 1.5 but this first selection yielded too many rejections of participants.

## B.2. Choice of priors

**Linear Mixed Models on baseline EMG, slopes EMG and starting point** For the priors on the populations intercepts ($\mu_\alpha$) of the LMMs on the baseline EMG, slope of the EMG and starting point (SP) of the DDM we chose normal distributed priors with following means ($\mu$) and standard deviations ($\sigma$) :

$$\mu_{\alpha_{Baseline}} \sim \mathcal{N}(\mu = 2.5, \sigma = 1.5)$$

$$\mu_{\alpha_{Slope}} \sim \mathcal{N}(\mu = 4, \sigma = 2)$$

$$\mu_{\alpha_{SP}} \sim \mathcal{N}(\mu = 0.5, \sigma = 0.25)$$

The population slope parameters ($\mu_{\beta_x}$), as well as the random effects on the intercept ($\sigma_\alpha$) and the slopes ($\sigma_{\beta_x}$), and the residual standard deviation ($\sigma_r$), for the baseline EMG, slope EMG and SP were all normal distribution centered on 0 with the following standard deviations :

$$\text{Baseline} : \mu_{\beta_x}, \sigma_\alpha, \sigma_{\beta_x}, \sigma_r \sim \mathcal{N}(\mu = 0, \sigma = 1.5)$$

$$\text{Slope} : \mu_{\beta_x}, \sigma_\alpha, \sigma_{\beta_x}, \sigma_r \sim \mathcal{N}(\mu = 0, \sigma = 2)$$

$$\text{SP} : \mu_{\beta_x}, \sigma_\alpha, \sigma_{\beta_x}, \sigma_r \sim \mathcal{N}(\mu = 0, \sigma = 0.25)$$

Allowing for a wide range of effect sizes, random effects and residual standard deviation with values close to 0 more likely than large values, see Appendix B.3 for a representation of the priors used for the baseline LMM. See the Appendix A.2 for the description of the priors used for the other variables.

## B.3. Prior predictive check on baseline EMG

To perform a prior predictive check we fitted the designed LMM model by specifying the priors but without providing the data hence directly sampling from the prior distributions (see Schad, Betancourt, & Vasishth, in press, for a principled way to apply prior predictive checks). As, when the amount of available data is high, priors have little influence on the posterior distribution we mainly use prior predictive check to ensure that we do not exclude plausible values. Figure .4 provide an example for the analysis of the baseline EMG where the priors for the population parameters are plotted against the observed data. As in Appendix A.2 the random effects on the intercept and on the slopes had the same prior as the residual standard deviation (sigma in Figure .4).

## B.4. DDM fits

Averaged quantile probability plot of the fits of the DDM to the data from the experiment in Section 3.3. For a description of quantile probability plots see Section
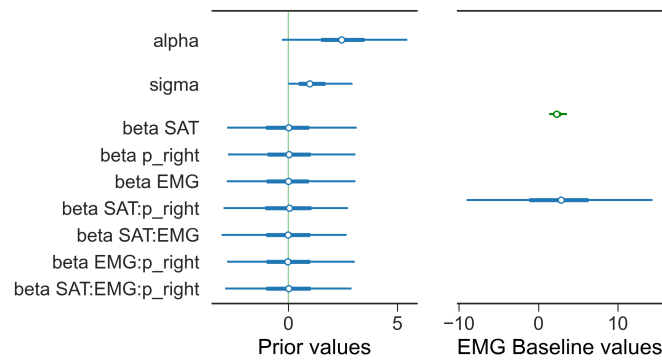
Figure .4.: On the left, distributions (mean and 95% CrI) of the samples for all population parameters from the prior only LMM model used for the baseline EMG analysis. On the right, actual distribution of the data in green (mean and 95% CrI) compared to predicted data computed from the parameters sampled from the prior only LMM in blue (mean and 95% CrI). Comparing the actual distribution of the data to the data compatible with the priors shows that the priors used do not exclude plausible data.

1.2.5.4. Note that, for simplicity, contrary to the quantile probability plot in Chapter 2 this quantile probability plot was constructed based on the group RT distribution, which according to Ratcliff, 1979 provides an adequate unbiased representation of the individual RT distributions.
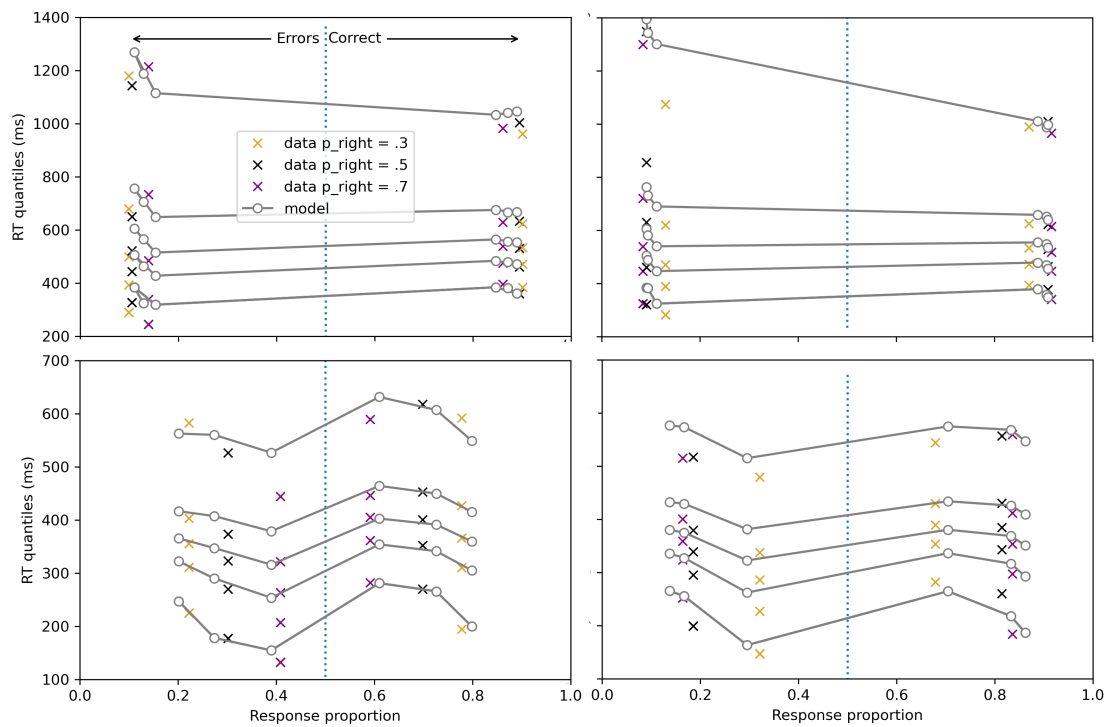
Figure .5.: Quantile-probability plots for a left tilted presented stimulus (left column) and a right tilted stimulus (right column), in the accuracy (upper row) and speed (lower row) conditions. The color code refers to the right probability cue levels.

# C. Appendix of Chapter 4

## C.1. From Linear model parameters to estimated effects

All regression models, including the regression on the DDM parameters followed the same factor coding scheme. The SAT factor was coded as a treatment factor (0 for accuracy and 1 for speed). The force condition was coded as a sum contrast (-0.5 for weak and 0.5 for strong force). The factor contrast was centered on the middle value and scaled so that -0.5 represented the lowest contrast and 0.5 the highest. A summary of this factor coding for the regression models is given in the matrices below, the first row represent the original levels, the second row the values on which the regression were estimated :

$$SAT = \begin{pmatrix} Accuracy & Speed \\ 0 & 1 \end{pmatrix}$$

$$Force = \begin{pmatrix} Weak & Strong \\ -0.5 & 0.5 \end{pmatrix}$$

$$Contrast = \begin{pmatrix} 23\% & 37\% & 51\% & 65\% & 79\% & 93\% \\ -.5 & -.3 & -.1 & .1 & .3 & .5 \end{pmatrix}$$

These coding features were chosen to ease the interpretation of the resultant coefficients. When the binary predictor is sum-contrasted (-0.5 and 0.5), the estimated $\beta$ value can be read as the difference between both conditions. When the binary predictor is treatment-contrasted (0 and 1), the estimated $\beta$ can be read as the difference to add to the intercept (predictor at 0) to obtain the mean of the condition where the predictor is at value 1. Hence, in our analysis, the intercept can be read as the predicted time for the reference condition where the SAT emphasis is on accuracy, and at an intermediate value for the predictors contrast and FC.

Given these coding features and the Bayesian nature of the estimation we can estimate the effect of a factor in a given condition and preserve the uncertainty associated with the effects. E.g. To compute the effect of Force in the speed condition we can add the interaction term $\beta_{SAT \times Force}$ to the estimated $\beta_{Force}$ in the accuracy condition. As coefficients are estimated using a MCMC procedure this addition is done on each MCMC iteration, allowing to keep the uncertainty around the resulting coefficient.

**Units of the g/LMMs parameters**   For the LMMs on proportion correct, $RT$, $PMT$ and $MT$, the data was transformed prior to the modeling (logit for proportion correct and log for the other variables). Using Monte Carlo Markov Chain (MCMC) processes, we back-transformed the predictions of the linear models for the chosen differences at each iteration, with the exponential for log transformed variables (LMM) or the inverse logit for proportion correct (gLMM). This preserves the uncertainty around the parameter values while reverting them to the natural units of the dependent variables.

## C.2. Model Selection

As seen in table .4 the DIC criterion almost always favor the complex models over the simpler one. However two patterns are consistent across the models, allowing the boundaries and the bias to vary with force conditions and drift rate to vary with SAT in addition to the contrast always improves the goodness of fit as assessed by the DIC. However when considering the BPIC criterion, initially intended to correct the complexity bias of the DIC, only allowing the variation of force on boundaries and on the bias seems to improve the goodness of fit. Hence based on BPIC we select the model allowing the boundaries and the bias to vary across force condition in addition to the designed base model (respectively M13 and M1 in Table .4). Importantly the results of the model selection is the same for a fit on PMT. The goodness of fit both on RT and PMT as displayed with quantile probability plot (see Figure .7 and .6) is satisfactory in most conditions but the amount of errors is rather misfitted when considering a high force especially in the fit on PMT (that pattern is common across all tested models).

## C.3. Predictions by V1 neuron activation onset

Reynaud, Masson, and Chavane, 2012 performed a measurement of the temporal activation of V1 neurons in awake monkey using voltage sensitive dye and the variation of this temporal activation with the contrast of stimuli close to the one used in this study. They then fitted the relationship between onset of V1 neurons activity and contrast with an inverted Naka-rushton equation from Barthélemy, Fleuriet, and Masson, 2010 :

$$\tau_c(c) = \tau_{max} + \tau_{shift} \cdot \frac{c_n}{c_n + s_{50}^n}$$

Where $c$ is contrast, $\tau_{max}$ and $\tau_{shift}$ are respectively the minimum latency observed at highest contrast and the maximum decrease in latency. $n$ is the estimated latency shift exponent, $s_{50}$ the estimated half decay contrast value. For the purpose of our analysis we recovered the values of the parameters estimated by Reynaud, Masson, and Chavane, 2012 and draw the predictions associated with the mean contrast levels used in our study.

Figure .8 represent the adjustment between point estimate of $T_0$ and $T_{encoding}$ with the curve predicted by the recovered parameters of Reynaud, Masson, and Chavane, 2012 for the inverted Naka-Rushton equation.

| | Bound. | Drift | $T_0$ | Bias | sBias | sDrift | $sT_0$ | $BPIC_{RT}$ | $DIC_{RT}$ | $BPIC_{PMT}$ | $DIC_{PMT}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| M1 | S | C | C×F×S | 1 | S | 1 | 1 | -12963 | 13447 | -16125 | -16595 |
| M2 | S | C×S | C×F×S | 1 | S | 1 | 1 | -12931 | -13483 | -16072 | -16618 |
| M3 | S | C×F | C×F×S | 1 | S | 1 | 1 | -12866 | -13419 | -16105 | -16645 |
| M4 | S | C×S×F | C×F×S | 1 | S | 1 | 1 | -12800 | -13467 | -16003 | -16668 |
| M5 | S×F | C | C×F×S | 1 | S | 1 | 1 | -13157 | -13661 | -16471 | -16967 |
| M6 | S×F | C×S | C×F×S | 1 | S | 1 | 1 | -13108 | -13689 | -16423 | -16993 |
| M7 | S×F | C×F | C×F×S | 1 | S | 1 | 1 | -13042 | -13614 | -16393 | -16959 |
| M8 | S×F | C×S×F | C×F×S | 1 | S | 1 | 1 | -12946 | -13639 | -16272 | -16964 |
| M9 | S | C | C×F×S | F | S | 1 | 1 | -13022 | 13519 | -16132 | -16617 |
| M10 | S | C×S | C×F×S | F | S | 1 | 1 | -12981 | -13548 | -16075 | -16637 |
| M11 | S | C×F | C×F×S | F | S | 1 | 1 | -12926 | -13493 | -16115 | -16668 |
| M12 | S | C×S×F | C×F×S | F | S | 1 | 1 | -12860 | -13540 | -16006 | -16687 |
| M13 | S×F | C | C×F×S | F | S | 1 | 1 | **-13212** | -13729 | **-16474** | -16987 |
| M14 | S×F | C×S | C×F×S | F | S | 1 | 1 | -13150 | **-13750** | -16418 | **-17007** |
| M15 | S×F | C×F | C×F×S | F | S | 1 | 1 | -13095 | -13654 | -16392 | -16976 |
| M16 | S×F | C×S×F | C×F×S | F | S | 1 | 1 | -13000 | -13728 | -16284 | -16987 |

Table .4.: Summary of the tested models displaying for each model (row) which parameters could vary with experimental conditions (S, F and C respectively for SAT, Force and Contrast, 1 indicates that only 1 estimate was fitted across all conditions). sBias, sDrift and $sT_0$ refer to the inter-trial variability parameters of the corresponding main parameters. The results in terms of BPIC and DIC are presented in the two last columns.
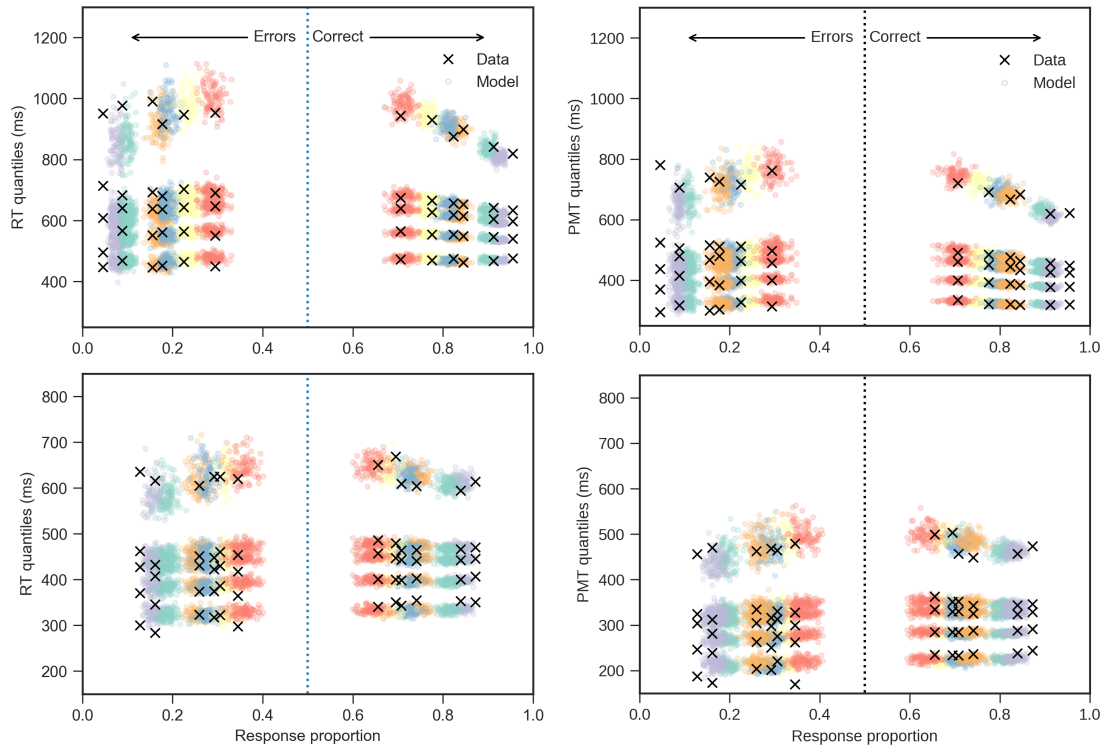
Figure .6.: Quantile-probability plots (Ratcliff & McKoon, 2008) for high force based on a fit on RT (left column) and PMT (right column), in the accuracy (upper row) and speed (lower row) conditions, computed from the best fitting model.

The X-axis displays obtained response proportion across contrast levels (color coded), symmetrically for errors (left side) and correct responses (right side). The Y-axis displays the fitted (dot) and observed (cross) $RT$ binned in 5 quantiles (.1, .3, .5, .7 and .9 quantiles, from bottom to top). Observed response proportion and $RT$ quantiles were computed from values pooled across participants. Model predictions were obtained by drawing 250 parameter values from the joint posterior distribution and computing their associated predicted performance. The misfit of the DDM is particularly apparent in the fit on PMT in accuracy (upper right corner), where the DDM clearly predicts a response proportion lower than the one observed on the data.
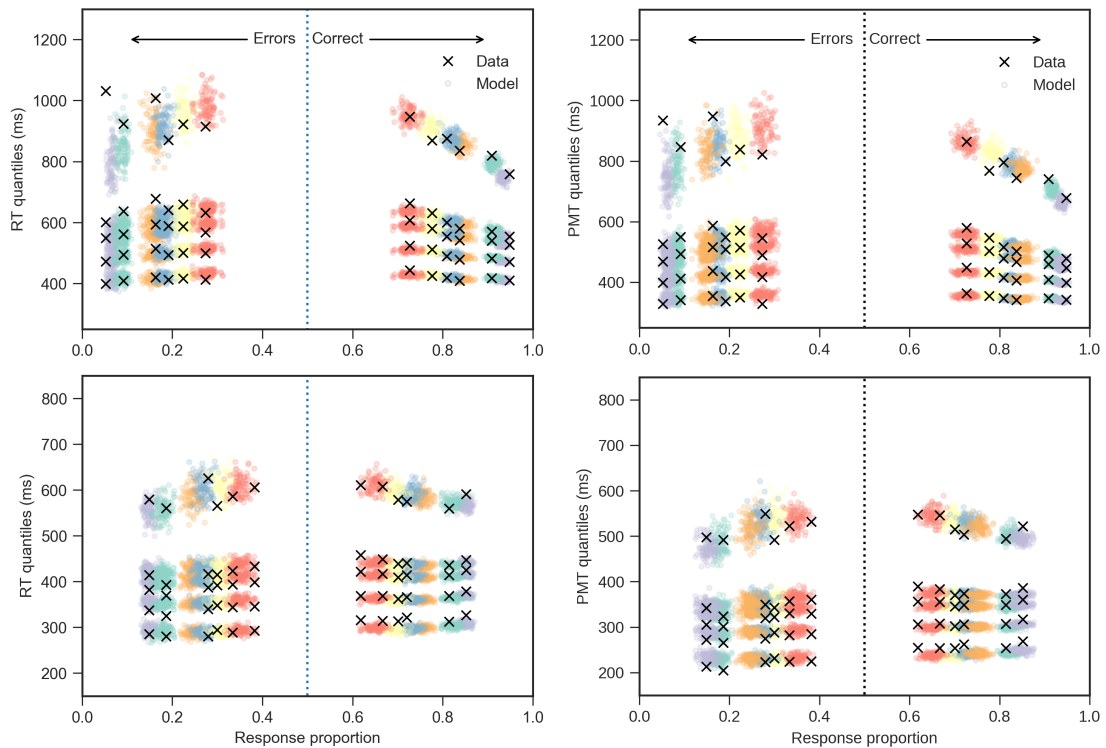
Figure .7.: Quantile-probability plots (Ratcliff & McKoon, 2008) for low force based on a fit on RT (left column) and PMT (right column), in the accuracy (upper row) and speed (lower row) conditions, computed from the best fitting model.

The misfit of the DDM is particularly apparent in the fit on PMT in accuracy (upper right corner), where the DDM clearly predicts a response proportion lower than the one observed on the data.
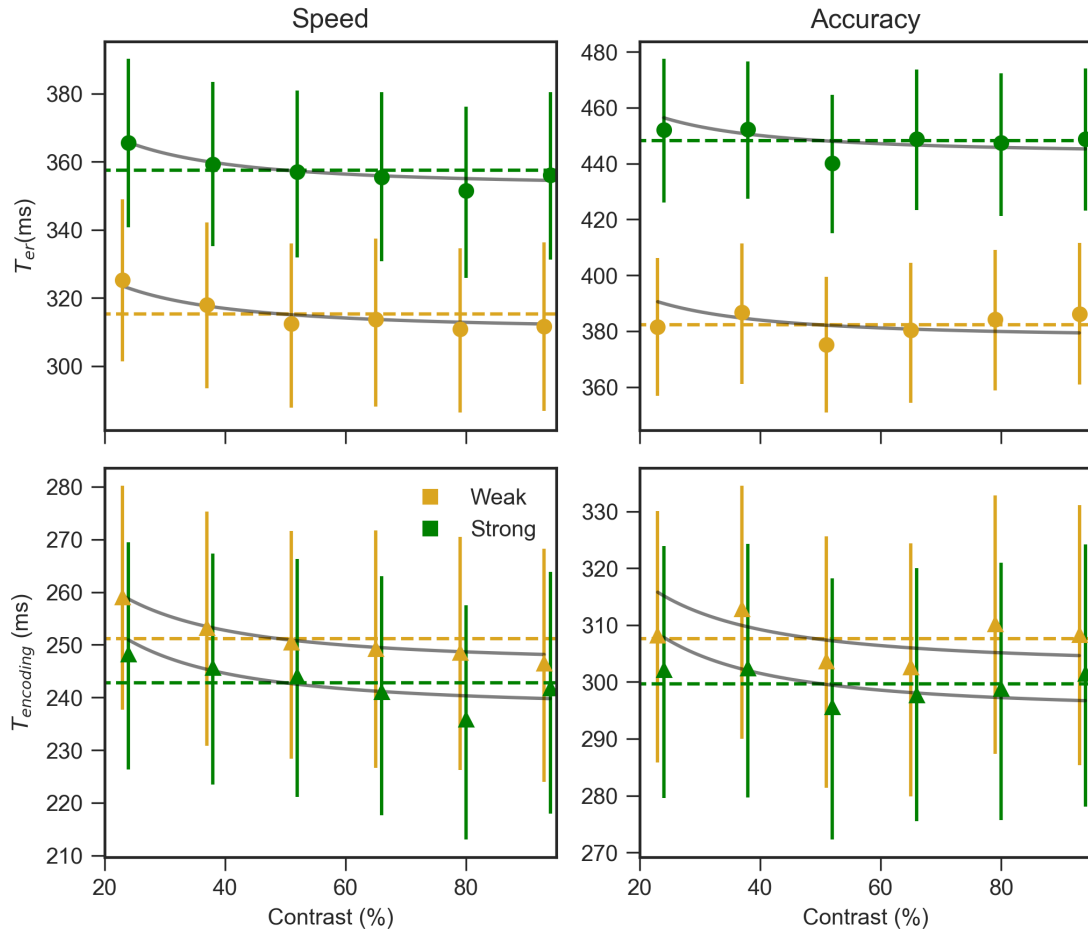
Figure .8.: Estimated $T_{er}$ (obtained from a DDM fit on $RT$) and $T_{encoding}$ (obtained from a fit on $PMT$) across contrast levels and splitted between SAT and Force conditions. Bars around the point estimate represents 65% CrI of the population mean. The colored lines represent the mean of $T_0$ or $T_{encoding}$ for each corresponding sub-cell. The grey lines represents the values predicted by the parameters of the inverted Naka-Rushton recovered from Reynaud, Masson, and Chavane 2012. These prediction have been first centered on 0 by subtracting their mean then rescaled by adding the mean of the corresponding sub-cell. Grey and color lines have therefore the same mean in each sub-cell formed by the combination of SAT and Force levels.

# D. Appendix of Chapter 5

## D.1. Replication of the analysis on the untrimmed data
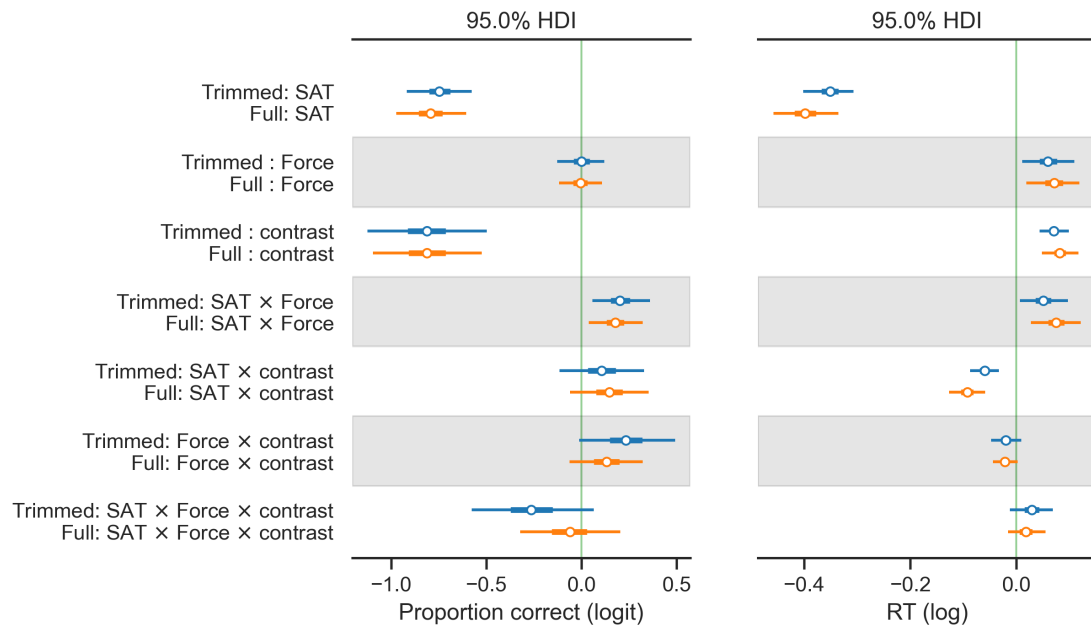


Figure .9.: Comparison of the slope values obtained for a g/LMM fit performed on the censored dataset (blue) *vs.* the full dataset (orange) on proportion correct (left) and $RT$ (right). Estimates are on the fitted scale, logit for proportion correct and log for $RT$.

## D.2. Random effect of contrast on N200

Figure .10 show the estimation of the individual effect of contrast. Note that as LMM are based on a hierarchical constraints the individual estimates are shrunk towards the population estimate.

## D.3. Goodness of Fit

Figure .11 shows the goodness of fit using quantile probability plots. The top right panel (speed and low force) indicates a clear misfit between DDM and RTs on the predicted error and correct response proportion.
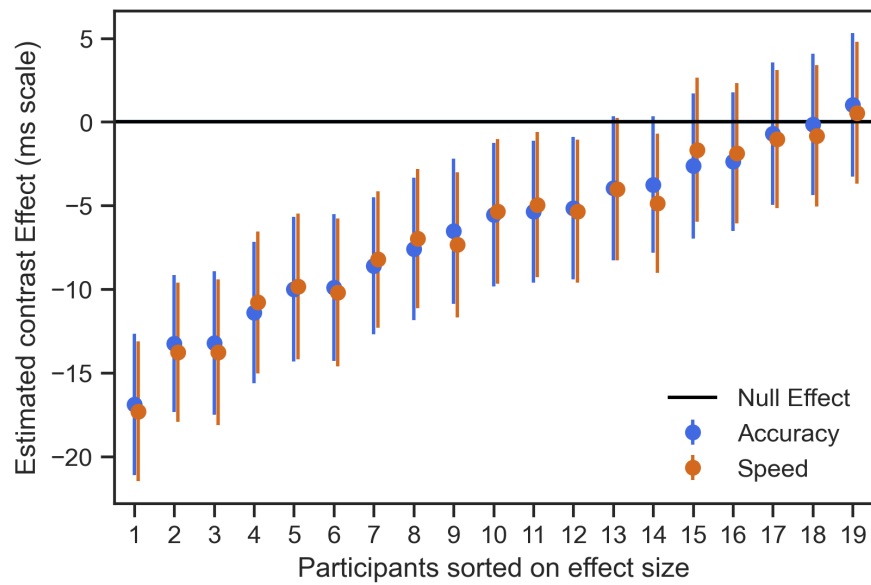
Figure .10.: Means of the marginal posterior distribution of the by-participant pre-
dicted contrast difference (lowest - highest contrast) in ms from the LMMs on the
N200 for both SAT conditions. The participants are sorted on the size of the esti-
mated difference in the Accuracy condition. The black line shows the null effect
and the bars around the markers show the estimated 95% CrI for the individual
differences.
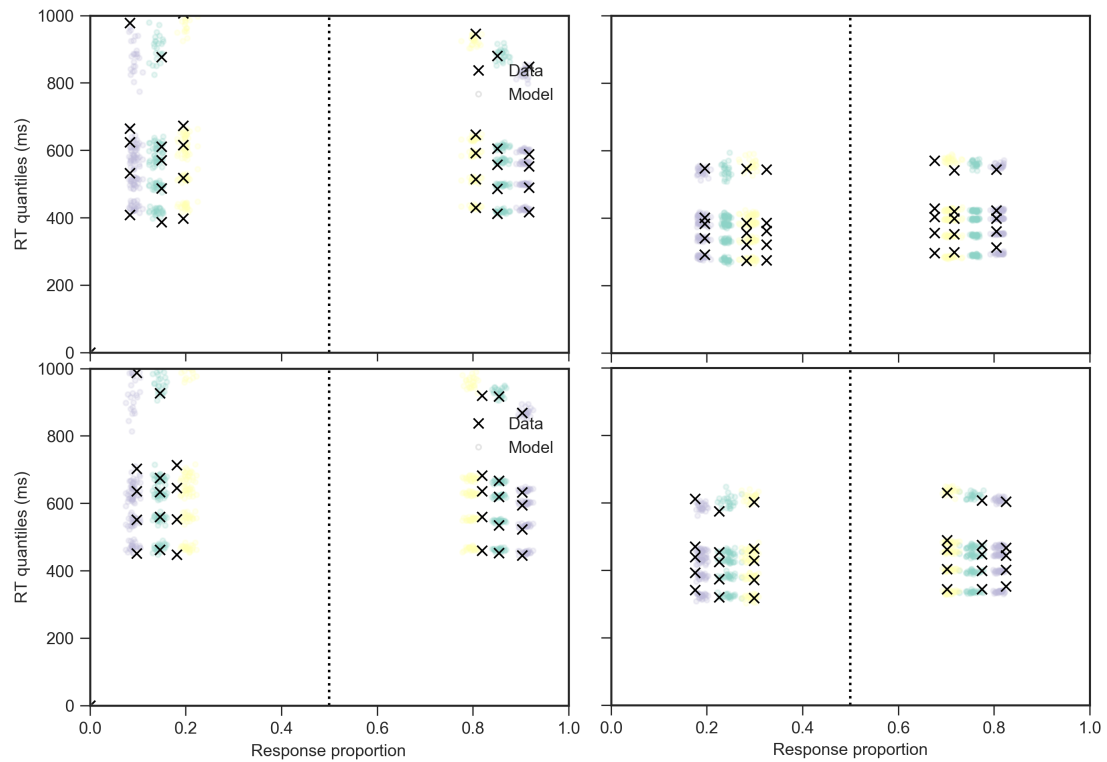
Figure .11.: Quantile probability plot for the DDM fit on RT split across SAT (Accuracy left panel and Speed right panel) and Force conditions (low force top row and high force bottom row). See Figure .6 for a complete description of the figure.

## D.4. Contribution of each latencies to the $RT$

It is to our knowledge the first attempt to decompose each $RT$ into an encoding, decision and response execution time. This decomposition does allow us to provide an answer to questions and assumptions given by mathematical psychologist. One common question is the relative variance of the $DT$ and the $NDT$. While on the mean latencies, $N200$ and $MT$ can sometimes be longer than the $DT$ (see Figure .12), both the $N200$ and the $MT$ contribute only weakly to the variance of the $RT$ (see Figure .13). This can also be appreciated at the individual level see Figure .14 for an example of the single trial decomposition for the first participant in the study. Therefore the common assumption that the variance of $NDT$ processes is small compared to that of the $DT$ seems true assuming that our decomposition is correct.
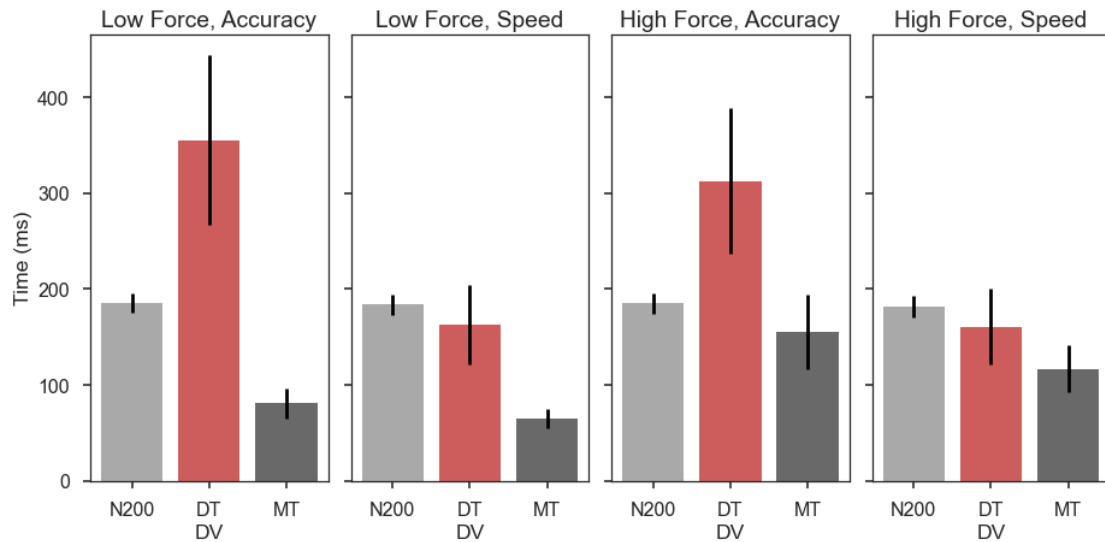


Figure .12.: Grand average of the $N200$, $DT$ and $MT$ across the four experimental cells formed by the combination of SAT and force factors. Bars represent one SD around the mean of the latency in the experimental cell.
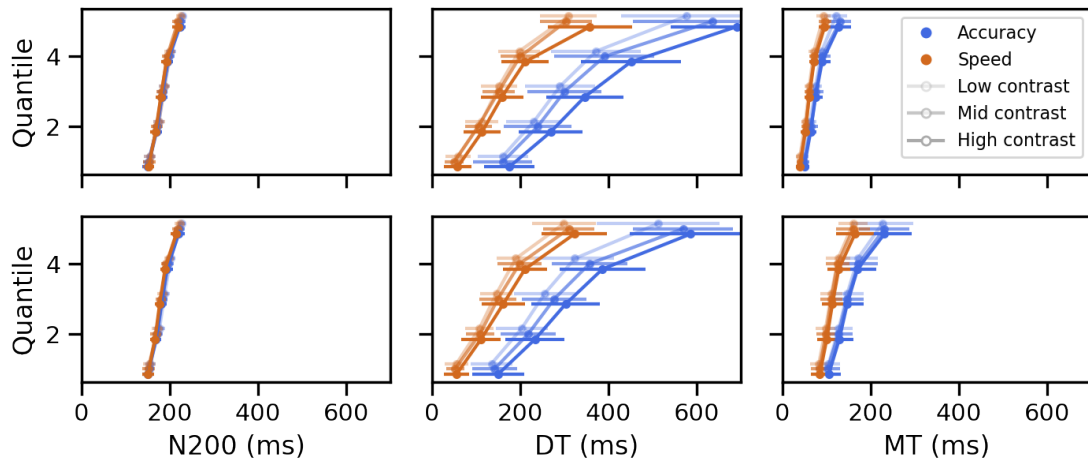
Figure .13.: Grand average of the $N200$, $DT$ and $MT$ separated into five equal-sized quantiles across all experimental cells. Bars represent one SD around the mean of the quantile.
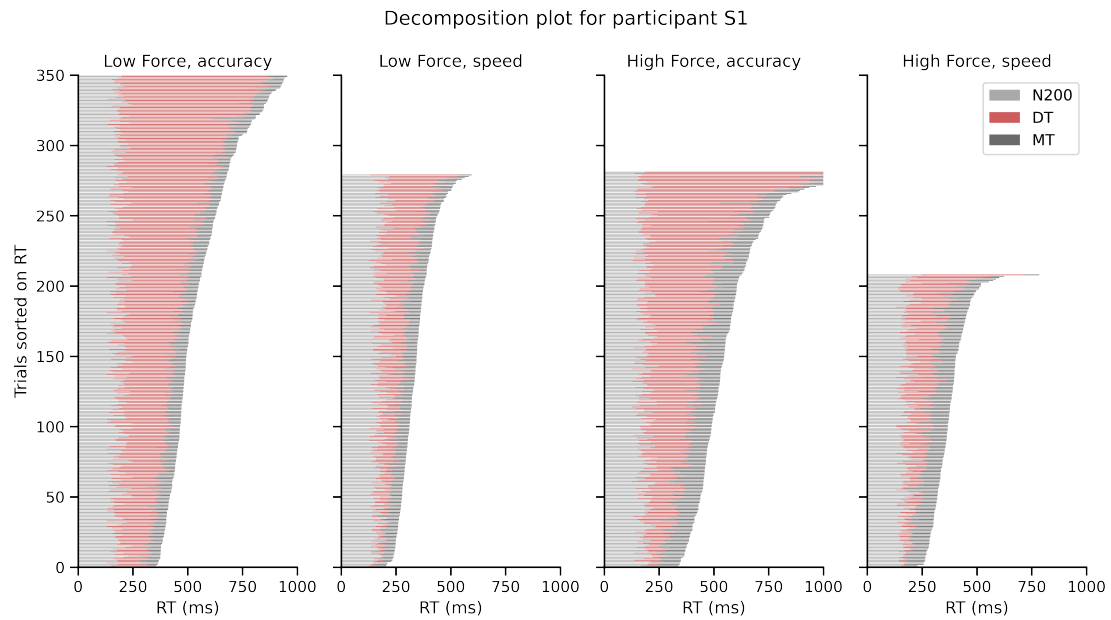


Figure .14.: Cumulated single-trial estimates of $N200$, $DT$ and $MT$ of the trials retained for the analysis for participant "S1" across the four sub-cell formed by SAT and Force factor combination.

# List of Figures

# List of Tables

# List of acronyms

**2AFC**

two alternative forced choice task. 18, 22, 25, 32, 83, 153, 154

**DDM**

Drift Diffusion Model. 25–30, 33–35, 40, 41, 78, 82

**EAM**

evidence accumulation models. 22, 24, 25, 27, 29, 31–40, 128, 130, 153, 160, 162

**EMG**

electro-myography. 38, 40, 41, 78, 82, 83, 104, 153, 154, 162

**LBA**

Linear Ballistic Accumulators. 25, 29, 33, 34

**LCA**

Leaky Competing Accumulators. 25, 34

$MT$

motor time. 38–41, 205

$PMT$

pre-motor time. 38–40, 205

**PRDM**

Proportional Rate Diffusion Model. 27, 85, 86, 99

**SAT**

speed accuracy trade-off. 20, 28, 31, 35, 36, 40, 154, 161

$T_D$

decision time. 22, 23, 25–28

$T_E$

encoding time. 22, 26, 28, 37

$T_{er}$

Encoding and Response execution Time. 28, 29, 39–41, 78, 88

$T_R$

response execution time. 22, 26, 28, 40, 78

# A. Résumé étendu

## I. Chronométrie mentale et décision

La chronométrie mentale est décrite par Posner en 1978 comme "l'étude du déroulement temporel du traitement de l'information dans le système nerveux humain" (Posner, 1978, p.7). Cette approche est apparue comme un élément central de la psychologie et de la neurophysiologie dès le milieu du XIXe siècle. Depuis lors, la chronométrie mentale est devenue l'un des domaines d'étude les plus influents de la psychologie. En effet, les mesures du temps de réaction (TR), c'est-à-dire le temps écoulé entre le début d'un stimulus et la réponse observable correspondante, couvrent une grande partie des travaux effectués en psychologie cognitive.

La psychologie cognitive postule que les fonctions mentales sont sous-tendues par des opérations de traitement de l'information élémentaires (les processus cognitifs). L'exécution de chaque processus cognitif prend un certain temps. Ainsi, le TR est la résultante des temps nécessaires à l'exécution des processus cognitifs impliqués dans une tâche. L'étude de ce TR a été un objet d'investigation privilégié des travaux de psychologie cognitive, qui visent à décrire l'architecture du système cognitif. De son côté, la neurophysiologie offre différents outils de mesure de l'activité cérébrale, dont certains avec une excellente précision temporelle. Ceci suggère une interaction prometteuse entre la psychologie et la neurophysiologie pour comprendre les processus cognitifs, et comment le cerveau met en œuvre, et donc contraint, ces processus.

### I.1. Histoire de la chronométrie mentale

En 1796, l'astronome royal Nevil Maskelyne a licencié son assistant de recherche David Kinnebrook pour avoir rapporté 800 millisecondes (ms) plus tard que lui la course temporelle d'un objet stellaire à travers le télescope de Greenwich (Mollon & Perkins, 1996). De ce licenciement abusif est née la chronométrie mentale et, selon certains historiens, la psychologie expérimentale. Si le conseil de prud'hommes avait existé, le cas de M. Kinnebrook aurait été difficile à défendre, car les philosophes de l'époque, influencés par les travaux de René Descartes, considéraient les pensées comme instantanées, ou du moins non-mesurables. Néanmoins, le besoin de précision des astronomes en matière de mesure a conduit Friedrich Bessel à proposer en 1822 l'"équation personnelle" comme moyen de corriger la variabilité inter-individuelle du temps de transit observé des objets stellaires (Meyer, Osman, Irwin, & Yantis, 1988).

Ce travail est la première preuve connue que les pensées prennent un temps qui peut être mesuré et qui est variable d'un individu à l'autre.

Par la suite, des physiologistes intéressés par la vitesse de conduction nerveuse sont arrivés à des conclusions similaires. Hermann von Helmholtz a rapporté en 1850 que la vitesse de propagation d'un courant électrique dans le nerf sciatique d'une grenouille était de 25,43 mètres par seconde (Schmidgen, 2002). Von Helmholtz s'intéressait en fait à la durée nécessaire après stimulation pour qu'une sensation soit perçue par l'humain, mais il ne pouvait pas appliquer à l'humain le protocole invasif employé chez la grenouille. Il a donc créé ce que nous appelons aujourd'hui une tâche de temps de réaction (TR) simple. Il a appliqué un léger choc électrique à un endroit du corps des participants et leur a demandé de produire une réponse dès qu'ils détectaient la stimulation. Von Helmholtz a supposé que le temps de réaction était défini sur la base de trois étapes additives séquentielles, la conduction de la stimulation entre les fibres nerveuses jusqu'au cerveau, les "processus de perception et de volonté" (ici et après, traduction libre de von Helmholtz tel que cité par Schmidgen, 2002) et la conduction de la commande motrice de réponse émise vers les muscles. En supposant que ces deux derniers stades sont constants, quel que soit le site de stimulation, il a calculé la vitesse de conduction de la sensation par soustraction, par exemple entre une stimulation du sacrum comparé à une stimulation du gros orteil. Au cour de ces mesures, il a de plus noté que "si au moment de la perception du signal les pensées sont occupées par autre chose, [...], cela [la réaction] prend beaucoup plus de temps" (Schmidgen, 2002). Ainsi, à l'idée que les pensées prennent un temps qui varie selon les humains, von Helmholtz a ajouté l'observation que ce temps peut être modifié par des changements dans l'attention du participant.

À la suite des travaux de von Helmholtz, de nombreux psychologues et physiologistes ont mesuré le temps de réaction et étudié les effets de l'attention et de bien d'autres conditions. Cependant, la décomposition fiable du TR en différentes étapes semblait rapidement hors de portée, ce qui a conduit à des déceptions au début du XXe siècle, comme le montre la citation de Robert Sessions Woodworth : "Puisque nous ne pouvons pas décomposer la réaction en actes successifs et obtenir le temps de chaque acte, à quoi sert le temps de réaction ? (traduction libre de la citation donnée par Schall, 2019). Depuis lors, les recherches visant à expliquer comment l'humain peut décider entre différentes alternatives à l'aide d'outils développés en statistique, physique, économie et neuroscience ont conduit à de nouvelles façons de parvenir à la décomposition des TR.

## I.2. Étudier la prise de décision grâce au temps de réaction

Notre vie quotidienne est rythmée par des décisions, des plus courantes comme le choix d'un masque contre le COVID-19 aux plus engageantes comme le choix d'une carrière. Lors de ces choix, nous pouvons mesurer deux dimensions: le TR et le choix effectivement fait. Ces deux indices peuvent servir à retracer les opérations effectuées par le système cognitif entre la stimulation et la réaction, à condition de prendre en

compte leur nature stochastique. Si le même stimulus est présenté sous les mêmes consignes à plusieurs reprises, un participant présentera un TR différent et même des choix différents. Cela implique que pour comparer des participants ou des conditions expérimentales, il faut répéter les essais, de sorte à extraire des tendances centrales, représentées par des indices calculés sur la base de plusieurs mesures de TR (par exemple, la moyenne et l'écart type des TR). On comparera ensuite les indices obtenus dans différentes conditions, en supposant que la différence de temps de traitement entre celles-ci est causée par des différences cognitives plutôt que par la variabilité aléatoire de la mesure. En effectuant plusieurs essais d'une même tâche, donc en mesurant plusieurs décisions d'un participant, on obtient une distribution des TR mesurés et une distribution des choix effectués. Ce sont ces distributions que la modélisation mathématique des prises de décisions cherche à ajuster.

## II. Modélisation cognitive du temps de réaction

Les psychologues et les physiologistes s'intéressent à la façon dont les humains et d'autres espèces traitent l'information en étudiant le TR. Cependant, comme pour beaucoup de mesures comportementales, plusieurs modèles cognitifs théoriques sont susceptibles de générer le comportement mesuré (Bastien & Bastien-Toniazzo, 2003). Par exemple, on observe des TR plus lents lorsque les participants sont tenus de privilégier la précision plutôt que la vitesse. Cela ne montre pas nécessairement que les participants aient été plus prudents : la différence de TR pourrait également s'expliquer par une exécution motrice plus lente des réponses. Autrement dit, les modèles cognitifs proposés sont des hypothèses qu'il s'agit de mettre à l'épreuve pour les étayer ou, au contraire, les rejeter.

La mise à l'épreuve expérimentale de modèles verbaux (p.ex. la "prudence" ou "l'exécution motrice" du précédent paragraphe) rencontre certaines limites. En effet, ces modèles se heurtent aux variabilités inter-individuelles concernant le signifié associé à chaque mot et également, par cette nature plus fluctuante, à des développements collaboratifs moins minutieux et plus à même de générer des interprétations implicites. La modélisation mathématique est un moyen de traduire le modèle verbal d'un comportement en un modèle formel dans lequel chaque processus supposé ainsi que les relations entre l'ensemble des processus impliqués dans la tâche sont représentés mathématiquement. La force de la modélisation mathématique est qu'elle ne peut se faire sans rendre explicite l'ensemble des hypothèse formulées sur la nature des processus impliqués et leurs relations. Le simple développement d'un modèle mathématique peut ainsi faire apparaître des hypothèses implicites aberrantes. Un modèle sera d'autant plus crédible qu'il permettra de bien ajuster les données observées (c-à-d., de produire des prédictions proches des données comportementales mesurées). Les modèles mathématiques permettent enfin de lier directement l'effet d'une condition sur les processus, leur paramètres et leurs relations, plutôt que de se fier uniquement à la mesure observée.

La chronométrie mentale est probablement le domaine de la psychologie ayant la plus longue tradition de modélisation mathématique. Elle bénéficie à ce titre de l'enseignement de décennies de recherches sur différentes fonctions cognitives, notamment celle qui nous intéresse ici, la décision entre différentes alternatives dans une tâche de prise de décision perceptive. Le cadre théorique le plus utilisé pour rendre compte d'une décision perceptive entre deux alternatives possibles est celui de l'accumulation d'évidences[1]. Dans ce cadre, après présentation d'un stimulus sur lequel un choix doit être opéré (p.ex. "CRTAF" est-il un mot fran ais ou pas?), le sujet échantillonne à des intervalles de temps réguliers des évidences en faveur ou défaveur d'une des deux alternatives (mot ou non-mot). A chaque intervalles ces évidences sont accumulées. Une fois que cette accumulation atteint un seuil d'une valeur prédéfinie, la réponse correspondante (p.ex. la réponse correspondant à "non-mot") est exécutée [2]. L'histoire des modèles de décision se basant sur l'accumulation d'évidence remonte aux travaux d'Alan Turing (voir Gold & Shadlen, 2002, pour une revue de l'implication des travaux d'Alan Turing dans les neurosciences de la décision) et du statisticien Abraham Wald, 1947. Ils ont depuis été déclinés en de nombreux modèles différents. Dans la section suivante, nous choisissons de nous limiter à décrire les modèles les plus utilisés lors de la modélisation de la décision, les modèles de diffusion avec dérive (DDM, de l'anglais *Drift Diffusion Model*).

**Modèle de diffusion** Le modèle de diffusion proposé par Ratcliff, 1978 postule l'existence d'une variable de décision $x$ (l'accumulation des évidences au cours du temps) qui présente une variabilité au cours de l'essai de sorte qu'elle prend, une valeur probabiliste donnée par :

$$dx = cdW + vdt, \qquad x(0) = 0 \qquad \qquad (A.1)$$

où $dx$ est l'évolution de la variable de décision sur un petit intervalle de temps $dt$. Cette évolution est une diffusion par construction: le premier terme $cdW$ est un bruit blanc distribué de manière Gaussienne avec une moyenne 0 et un écart-type $c$ appelé le *coefficient de diffusion*. Mais cette diffusion peut dériver vers l'une des alternatives en fonction de la valeur de $v$ appelée le *taux de dérive*. Dans le cadre d'une tâche de choix, $v$ représente l'augmentation moyenne sur $dt$ vers l'alternative favorisée par le stimulus. Ce processus de diffusion s'arrête dès que la valeur absolue de $x$ dépasse le paramètre des *seuils a*. Bogacz, Brown, Moehlis, Holmes, and Cohen, 2006 fournit deux expressions pour montrer comment ces paramètres sont liés à la proportion moyenne de réponses correctes ($P_c$, le nombre de fois ou des stimuli

---

[1]Notez que le terme anglais pour ce cadre théorique est celui d'*evidence accumulation* qui se traduit littéralement par "accumulation de preuves". Le terme d'évidence est un anglicisme que nous adoptons par commodité car le terme de preuves ne correspond pas exactement à la terminologie anglaise.

[2]Les lecteurs intéressés par une description plus exhaustive pourront se référer au texte en anglais à partir de la section 1.2.1

comme "CRTAF" ont été correctement choisis comme des non-mots) et la moyenne du temps de décision ($\overline{T_D}$) de plusieurs décisions :

$$P_c = 1 - \frac{1}{1 + \exp\left(2va/c^2\right)}, \qquad \text{et} \tag{A.2}$$

$$\overline{T_D} = \frac{a}{v}\tanh\left(\frac{va}{c^2}\right) \tag{A.3}$$

Pour obtenir le TR moyen, un paramètre supplémentaire est nécessaire, une constante pour déplacer la distribution du TR vers les valeurs observées. Cette constante ou temps résiduel $T_{er}$ est ajoutée au $\overline{T_D}$ de l'équation A.3 pour calculer le TR moyen prédit.

Les équations A.2 et A.3 supposent que le point de départ de l'accumulation est situé à 0, à égale distance des deux seuils. Une absence de symétrie entre ces seuils est saisie par le paramètre *point de départ z*. Si $z$ prend des valeurs positives, l'accumulation est biaisée vers l'alternative représentée par le seuil positif ($a+$) et inversement à $a-$ si $z$ prend des valeurs négatives[3]

Le formalisme de ces paramètres définit l'interprétation cognitive que nous en avons. Le taux de dérive détermine la vitesse de l'accumulation, on le suppose donc être lié à la vitesse de traitement du stimulus par un sujet. Les seuils de décision représentent la quantité d'information nécessaire avant d'engager une réponse et donc la précaution de réponse du sujet (p.ex. demander à un sujet de privilégier la vitesse conduit à des seuils plus bas que lorsqu'on lui demande de privilégier la précision de sa réponse). Le point de départ représente le biais a priori envers une réponse particulière, si par exemple une catégorie de stimuli est plus présente qu'une autre le participant va juger comme plus probable la réponse associée à cette catégorie. Finalement le temps résiduel est supposé être le temps pris par les processus indépendants de la décision comme le temps d'encodage du stimulus (p.ex. la traduction de la stimulation visuelle en un code neural exploitable) ou le temps d'exécution de la réponse (comme la décomposition de Van Helmoltz), on parle donc du paramètre du temps non-décisionnel.

## III. Objectifs de la thèse

Par leur capacité à bien ajuster les données observées dans de nombreuses tâches de décision, ainsi que grâce à des perspectives optimistes sur l'interprétabilité cognitive des paramètres numériques issus de la décomposition des TR, les modèles de la prise

---

[3]Les paramètres sont relatifs les uns aux autres, c'est-à-dire qu'on peut multiplier chaque paramètre (sauf la constante ajoutée au $T_D$) par la même valeur et obtenir exactement les mêmes prédictions. Il est donc nécessaire de fixer l'un de ces paramètres à une valeur donnée afin de mettre les paramètres à l'échelle. Habituellement, dans les applications du DDM, le *coefficient de diffusion c* est fixé à 1 (ou 0,1 dans certaines applications). Fixer $c$ à une valeur donnée ramène donc le nombre de paramètres libres dans le DDM à 4 paramètres : le taux de dérive, les seuils, le point de départ et le temps résiduel supplémentaire.

de décision connaissent une popularité grandissante. On les utilise pour lire à une échelle plus élémentaire des phénomènes psychologiques historiques jusqu'alors uniquement appréciés par le biais comportemental, pour produire une interprétation fonctionnelle des données neuro-physiologiques ou encore pour déterminer l'origine cognitive de différences entre des groupes d'individus, par exemple différentes populations cliniques. Mais l'intérêt de ces applications est conditionné à l'interprétabilité de ces paramètres, qui, comme nous l'avons vu, offre des perspectives prometteuses mais n'est pas encore établie.

La présente thèse vise a évaluer l'interprétabilité de ces modèles, dont on sait déjà qu'ils offrent un excellent ajustement des données, en s'appuyant sur des données comportementales et électro-physiologiques. Pour ce faire, nous utilisons l'activité électrique du groupe musculaire responsable de l'exécution de la réponse du participant, mesurée à l'aide de l'electro-myographie (EMG), et l'activité électrique cérébrale durant l'ensemble de la tâche mesuré à l'aide d'un électro-encéphalogramme. Ces mesures électro-physiologiques nous ont permis de décomposer les TR mesurés, par exemple en isolant les temps d'exécution motrice ou d'encodage. Nous avons ainsi pu confronter aux décompositions réalisées par les modèles une décomposition électrophysiologique et par là mettre à l'épreuve les hypothèses des modèles sur l'organisation du système cognitif lors d'une prise de décision. Pour ce faire, nous avons manipulé des facteurs expérimentaux supposés être liés aux différents processus cognitifs impliqués. Cette démarche nous a permis de mettre en évidence les résultats résumés et exposés dans la section suivante.

## IV.  Discussion

Nous combinerons ici les preuves recueillies tout au long des chapitres de la thèse pour résumer ce qu'ils révèlent sur les processus cognitifs à l'oeuvre dans la prise de décision. Nous passerons ensuite à une perspective plus générale sur la façon dont s'articulent les processus de décision et de non-décision, et sur ce que nos résultats impliquent pour des modèles comme le modèle de diffusion avec dérive (DDM) et donc pour la théorie sur la prise de décision qu'ils implémentent.

### IV.1.  Effets liés à la décision et temps moteurs

Dans les quatre chapitres de la thèse, cinq expériences étaient associées à une manipulation des instructions de vitesse données aux participants. Elles ont toutes indiqué que le temps d'exécution motrice mesuré à l'EMG (temps moteur, TM) était plus court lorsqu'il était demandé au sujet d'accélérer sa réponse. Ces effets ne pourraient être plus cohérents. Au niveau théorique, cette constatation a des implications importantes. Elle montre que l'exécution motrice n'est pas une simple constante ajoutée au temps de réaction (TR) et, plus important encore, que les participants peuvent adapter le recrutement des processus moteurs en fonction des instructions.

Dans les expériences du Chapitre 2 nous avons également manipulé la difficulté de l'essai en manipulant la différence d'intensité des stimuli. Lorsque la réponse du sujet est correcte, on observe un effet sur le TM : un stimulus facile génère un TM plus rapide qu'un stimulus difficile. Sanders, 1983 suggère que l'effet de l'intensité du stimulus sur TM est lié à l'éveil immédiat produit par l'intensité globale de la stimulation physique. Cependant, lorsque l'on a manipulé cette intensité globale, on n'a plus observé d'effet sur le TM. Cette observation remet en cause l'hypothèse de Sanders, 1983; lorsqu'on rajoute l'observation que la difficulté de l'essai a un effet opposé sur le TM lorsque le sujet commet une erreur, l'ensemble suggère plutôt que la variabilité du TM est lié à la décision en cours.

Un autre effet nous renseigne sur la nature du TM. L'effet de la justesse de réponse est défini comme la différence de TR entre les réponses correctes et incorrectes. Il est intéressant de noter que, dans les trois expériences des chapitres 2 et 3, nous observons systématiquement un TM plus court pour les erreurs. L'exécution plus lente des réponses d'erreur a été également observée par Allain, Carbonnell, Burle, Hasbroucq, and Vidal, 2004; Śmigasiewicz, Ambrosi, Blaye, and Burle, 2020. Ces auteurs ont interprété celle-ci comme une preuve de contrôle exécutif en direct, par lequel les participants tentent de supprimer la réponse erronée en cours. Dans le chapitre 2, cet effet de correction interagit avec la manipulation de la difficulté de l'essai. Les stimuli plus faciles ont produit simultanément des erreurs plus lentes et des réponse correctes plus rapides, au regard du TM, que les essais plus difficiles. L'hypothèse du contrôle exécutif en ligne ne peut pas expliquer cette interaction. A la place de cette interprétation, nous proposons une nouvelle hypothèse, basée sur un projet parallèle à la thèse mené avec d'autres collaborateurs.

Gajdos, Fleming, Saez Garcia, Weindel, and Davranche, 2019 a montré que le système moteur est lié à l'évaluation de la confiance dans une décision (c-à-d., dans quelle mesure un participant pense que sa réponse est correcte). En outre, des études telles que celles menées par Balsdon, Wyart, and Mamassian, 2020; Dotan, Meyniel, and Dehaene, 2018 suggèrent que la confiance dans une décision est un phénomène qui évolue en même temps que la décision. Par conséquent, si le système moteur est lié à l'évaluation de la confiance et que cette évaluation évolue au cours de la décision, nous pouvons suggérer que l'évaluation de la confiance pourrait expliquer les effets que nous observons sur les temps moteurs. Par exemple, l'étude de Dotan, Meyniel, and Dehaene, 2018 montre que les participants accélèrent leur mouvement selon le degré de confiance momentané dans la réponse en cours. Par conséquent, l'effet d'interaction sur le TM observé entre la justesse de réponse et la difficulté de l'essai pourrait être lié à des variations de vitesse dues à la confiance. Sur des essais difficiles, la confiance est faible quelle que soit la justesse de la réponse tandis que pour les essais faciles, le paramètre de confiance pourrait être beaucoup plus contrasté entre les réponses justes et fausses. Ainsi, les variations observées sur le TM pourraient être liées au jugement de confiance lors de la prise de décision. Si cette interprétation se révélait être valable, les effets sur les temps moteurs ne pourraient pas être récupérés par un modèle, et donc une théorie, qui ne tiendrait pas compte de la façon dont la

confiance se manifeste pendant la décision.

## IV.2. Sur les relations entre temps décisionnel et temps non-décisionnel

Les facteurs connus pour avoir un impact important sur le TM, tels que les manipulations de la force nécessaire pour répondre, se traduisent également de façon notable par des effets sur le temps pré-moteur déterminé par l'EMG.

Dans le chapitre 2, nous avons remarqué que l'estimation des temps de non décision par le modèle n'a pas changé d'une expérience à l'autre, malgré la réduction de la force nécessaire pour produire une réponse. Ce résultat jette un doute sur le fait que les effets de non-décision réels aient été correctement reflétés par les processus de non-décision estimés. En outre, le chapitre 3 a montré que le modèle de diffusion avec dérive (DDM) ajusté dans le chapitre 2 confondait effectivement un biais moteur avec un biais décisionnel. Une différence dans la vitesse d'exécution de la réponse entre les mains gauche et droite a été interprété par le DDM comme étant lié à un changement dans le point de départ du processus d'accumulation des évidences. En utilisant la stratégie de modélisation proposée par A. Voss, Voss, and Klauer, 2010, les décompositions du modèle et de l'EMG ont pu être alignées puisque ce modèle estimait qu'un biais d'exécution de la réponse était présent alors que le point de départ n'était pas biaisé. En généralisant, il se peut qu'un effet réel sur les temps moteurs soit interprété par des modèles de décision de type DDM comme provenant des processus décisionnels. Ceci remettrait en cause à minima l'interprétation cognitive que nous avons des paramètres de modèles de décision.

Les chapitres 4 et 5 cherchaient à vérifier si les manipulations visant à modifier le temps d'encodage et le temps moteur seraient identifiées comme telles (c-à-d., non-décisionnelles) par les modèles. Ces chapitres fournissent des preuves mitigées en faveur de la décomposition du TR telle qu'elle est supposée par ces modèles.

Dans le chapitre 5 nous avons extrait un temps supposé d'encodage du stimulus dans notre tâche à l'aide de l'électro-encéphalographie. Cette mesure a varié comme prévu avec le facteur de l'intensité du stimulus, confortant la méthodologie employée pour l'extraction du temps d'encodage sur la base des données électro-physiologiques. Le paramètre de temps non-décisionnel estimé ne rendait pas compte de ces variations dans toutes les conditions. Ce résultat remet en cause le lien entre paramètres du modèle et la mesure du processus non-décisionnel de l'encodage d'un stimulus.

Le chapitre 4 a d'abord montré que l'augmentation de la force nécessaire pour exécuter une réponse diminuait le temps pré-moteur. Cependant, le raccourcissement de ce que le modèle considère être le temps de décision, ne s'est pas traduit par une différence dans la proportion de réponses correctes. Le fait que les participants atteignent la même précision de réponse tout en passant moins de temps avant le début de l'activité musculaire ne peut s'expliquer que par le fait qu'ils continuent à décider pendant qu'ils s'engagent dans l'exécution de la réponse. Bien que le DDM ne puisse tenir compte de ce chevauchement entre les processus moteurs

et décisionnels, il en a fourni une confirmation indirecte. Ajusté uniquement sur le temps pré-moteur, le DDM prédit (à tort) moins de réponses correctes. Cela montre que le DDM interprète la partie soustraite (temps moteur) comme un temps pendant lequel une partie de la décision a eu lieu. Ce chapitre montre donc que le temps que nous enregistrons comme temps de non-décision n'est pas interprété comme un temps de non-décision par le modèle. Il semble donc que, pour ajuster les données de précision, le modèle sous-estime le temps passé à exécuter la réponse. Comme des modèles tels que les DDM sont généralement utilisés pour faire des inférences sur les processus de décision, cela peut être une propriété souhaitable bien que cela empêche d'interpréter le temps de non-décision comme le temps d'exécution motrice.

Le chapitre 5 renforce l'interprétation du chevauchement entre temps moteur et temps de décision et ajoute sur la question des conséquences pour un modèle tel que le DDM. Dans ce chapitre on observe une interaction entre l'effet de la force d'exécution et l'instruction de vitesses sur le nombre de réponse correcte. Lorsqu'on donnait pour consigne aux participants de privilégier la rapidité de réponse, leur précision était accrue lorsqu'une force élevée était requise par rapport à une force faible. Comme seul le temps moteur était allongé dans cette condition de vitesse et exécution élevée, cela conduit à nouveau à la conclusion que les participants décidaient pendant que leur réponse motrice était en cours. Ce pattern de changement de précision entre les conditions de force a cependant été manqué par le DDM. Cela montre que, dans cette condition, le DDM n'a pas été en mesure d'attribuer le temps de décision à la période d'exécution telle que mesurée avec l'EMG.

Nous pouvons relier l'hypothèse que nous avons faite sur la source des effets sur les temps moteurs étant liée à la confiance et la proposition du temps moteur et de décision se produisant en parallèle. Balsdon, Wyart, and Mamassian, 2020 ont montré que, dans les tâches de décision où les évidences se présentent par étapes séquentielles, la confiance dicte le seuil de la décision. La généralisation n'est pas facile car leur tâche est différente de celle utilisée ici, notamment parce que toutes les évidences sont présentées simultanément dans nos tâches. Nous pouvons cependant suggérer que leur interprétation pourrait s'accommoder de la relation temporelle entre le temps moteur et de décision. Si l'exécution de la réponse est liée à la confiance en directe et que la décision se produit pendant le temps d'exécution de la réponse, alors mécaniquement, une confiance plus élevée dicte un temps de décision plus court grâce à l'accélération des processus moteurs et donc un seuil de décision plus bas.

## IV.3. Conséquences pour les inférences faites à partir des modèles de prise de décision

La conclusion d'un temps d'exécution et de décision parallèle est déjà suggérée par d'autres études sur des exécution d'actions plus lentes et plus complexes (Buc Calderon, Verguts, & Gevers, 2015; Dotan, Meyniel, & Dehaene, 2018; Resulaj, Kiani, Wolpert, & Shadlen, 2009; Selen, Shadlen, & Wolpert, 2012). Sa démonstration dans le cadre d'une exécution motrice rapide et en l'absence de mouvement, donc conforme

à la majorité des études sur la prise de décision, est très importante sur la théorie de la décision. Si cette conclusion est valide, alors cela implique que les modèles comme le DDM ne sont pas en mesure de rendre compte à la fois du temps de décision et du temps d'exécution motrice car la théorie les suppose indépendants et sériels, l'étape motrice démarrant nécessairement à la fin de la décision.

Implicitement au moins, les chercheurs supposent souvent que pour s'engager dans une décision, il faut d'abord atteindre un niveau de preuve nécessaire avant de produire une action, ce que nous montrons comme étant invalide. S'il existe probablement un critère pour lequel l'incertitude est suffisamment faible pour commencer à s'engager dans une action, ce n'est pas le critère que l'on retrouve dans les modèles comme le DDM (sans quoi la décomposition opérée par le DDM correspondrait à la décomposition que nous faisons avec l'EMG). Cela pourrait être considéré comme un problème conceptuel impliquant simplement de redéfinir ce que nous interprétons lors de l'estimation des seuils et du temps de non-décision, mais cela semble se traduire par un problème d'estimation. Plus précisément, lorsque les participants n'ont pas compensé la perte de temps d'exécution/de décision de la réponse induite par la réduction de la force d'appui, le DDM a manqué le pattern de la précision des participants. Inversement, lorsque les participants ont compensé la perte de temps d'exécution de la réponse/décision, le DDM a sous-estimé l'effet de la force sur le temps d'exécution de la réponse.

Dans l'ensemble, cela renforce l'idée que l'exécution motrice et la décision sont deux processus différents qui se chevauchent dans le temps, car il semble que le DDM peut soit récupérer la présence d'un effet sur l'exécution de la réponse, soit sur un effet sur la décision (c'est-à-dire le nombre de réponses correctes), mais pas sur les deux.

L'interprétation précédente montre que les conclusions que l'on peut tirer sur les processus moteurs en analysant le temps de non-décision ne refléteront probablement qu'une partie de l'effet moteur réel. Le fait que l'estimation que fait le DDM du temps de non-décision soit corrélée avec le temps moteur, mais pas fortement, indique également la capacité du modèle à prendre en compte les processus moteurs contenus dans les TM, mais probablement pas la totalité.

## V. Conclusion générale

Sur la base des résultats de la présente thèse, nous aboutissons à deux nouvelles affirmations sur la façon dont les humains prennent des décisions perceptives. Premièrement, le temps d'exécution motrice capturé par l'EMG est un processus ayant sa propre variabilité et sensibilité aux facteurs expérimentaux, même lorsque la réponse est une simple pression isométrique sur un bouton. Nous fournissons quelques hypothèses sur la manière dont nous pensons que ces facteurs influencent le temps d'exécution. Ensuite, en se basant sur les résultats issus de l'analyse du comportement, la décomposition de l'EMG et la modélisation, nous voyons que ce temps d'exécution

se produit parallèlement au temps de décision.

Pour les modèles de prise de décision, et donc pour les théories sur lesquelles ils sont construits, il s'agit de plus qu'une simple clarification conceptuelle. Premièrement, la prise en compte des effets sur les processus d'exécution de réponse ne peut être réalisée que par un modèle mathématique cognitif, et donc une théorie, qui rendent compte de l'origine de ces effets. Il appartient aux chercheurs de choisir si ces variations peuvent être ignorées par souci de parcimonie, mais les paramètres d'exécution des réponses détermineront l'impact des processus moteurs sur le TR. Deuxièmement, nous concluons que si les modèles mathématiques cognitifs tels que les modèles d'accumulation d'évidence peuvent fournir des informations précieuses sur les processus sous-jacents aux performances, leur généralisation à un large éventail de conditions et de tâches n'est pas justifiée tant que ces conditions n'ont pas été testées de manière approfondie. Dans nos expériences, la mesure comparée à l'estimation des processus cognitifs est congruente seulement avec des $TR$ accélérés et dans des conditions d'exécution motrice de réponse faibles.