Decision modelling insights in cognition and adaptive decision making

BY:
ANGELO PIRRONE

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF SHEFFIELD
FACULTY OF SCIENCE
DEPARTMENT OF PSYCHOLOGY
OCTOBER 2016

Decision modelling insights in cognition and adaptive decision making

Abstract

The goal of this research is using computational models of decision making, in particular two models, the Drift Diffusion Model (DDM; Ratcliff & McKoon, 2008) and Pais et al. (2013) model, to provide insights in cognition and adaptive decision making.

In the first part of this dissertation, we applied the Drift Diffusion Model to three domains: cognition in Autism Spectrum Disorder (ASD), Task-Irrelevant Perceptual Learning (TIPL) and Semantic Congruity Effect Research. Regarding ASD research, we show that differences in reaction times and accuracy in two-alternative forced-choice tasks between ASD subjects and controls, previously interpreted as enhancements or impairments, are instead due to different decision criteria and longer time to execute the motor response for ASD subjects. This result has important consequences for clinical research in which differences in response conservativeness and motor response have been interpreted as differences in information processing. In the third chapter, by applying the DDM, we show that TIPL, learning to better discriminate a stimulus that is irrelevant to a task, does not monolithically affect the sensitivity to the stimulus, but also affects the decision criterion of subjects. Our results show that an analysis only based on accuracy - that is the standard in the literature - could be potentially misleading in the interpretation of learning data, since learning affects different components of decision making, which have different effects on accuracy or reaction times. In the fourth chapter, we perform a DDM decomposition of the semantic congruity effect, the result that subjects are (i) faster in judging the bigger of two big stimuli or the smaller of two small stimuli - as opposed to the bigger of two small stimuli or the smaller of two big stimuli (ii) faster in determining whether a target stimulus is bigger or smaller than a standard stimulus when the size of the two stimuli coincides. Our DDM decomposition allows us to test different verbal theories that have been proposed for the explanation of this phenomenon and to show that this phenomenon arises as an increase in the rate at which subjects accumulate evidence in case of congruency between the magnitude of the standard stimulus and the magnitude of the target stimulus. In sum, in the first part of this dissertation, our work shows the benefits of isolating the different cognitive processes that are involved in decision making and the benefits of testing theories and generating conclusions from data by applying computational models of choice.

In the second part of this dissertation, inspired by a model that describes decision making in honeybees (Pais et al., 2013), we investigate a feature of decision making that arises from this model and that cannot be accounted for by a whole family of computational models of choice, DDM included. The DDM, as many other models of choice, disregards the information regarding the overall magnitude of the alternatives, since it only focuses on the differences between alternatives. In the fifth chapter, we argue from an evolutionary perspective why we should expect decision making to take under consideration the magnitude of alternatives and why this poses a challenge to some decision making models, which are instead insensitive to such information. In the sixth chapter, we provide evidence for different species (humans and monkeys) and different domains (perceptual decision making and reward-based decision making) for the existence of magnitude sensitivity in decision making. In sum, in the second part of this dissertation, a mechanism of a computational model of decision making in honeybees, has led us to generate hypotheses in adaptive decision making and to the understanding of the limitations of some computational models of choice.

Collectively, our work show the benefits of computational models of choice in the analysis of data and in the generation of hypotheses.

Contents

List of figures

List of Tables

Collaborators and Contributions

The chapters of this dissertation correspond to five papers. Copyright permission from the respective Copyright holders has been granted, both for the papers already published and even for those under review, meaning that if and when they are going to be accepted it is fine that they appear in this dissertation. Permission of reproduction has been granted from co-authors as well.

Chapter 2: Pirrone, A., Dickinson, A., Gomez, R., Stafford, T., & Milne, E. (in press). Understanding perceptual judgement in autism using the drift diffusion model. *Neuropsychology*.

Contributions: Elizabeth Milne, Abigail Dickinson, Angelo Pirrone and Tom Stafford conceived of the paper; All authors discussed the material; Rosanna Gomez and Abigail Dickinson run the experiment; Angelo Pirrone conducted all model fitting analyses and drafted the paper for the part regarding the model fitting and the limitations, Tom Stafford conducted all behavioural analyses and drafted the paper for the part regarding behavioural analyses, Abigail Dickinson drafted the paper for the material and methods part, Elizabeth Milne drafted the paper for the introduction and the conclusion and all authors discussed and approved its content.

Chapter 3: Pirrone, A., & Stafford, T. (under review). A Drift Diffusion Model decomposition of task-irrelevant perceptual learning.

Contributions: Angelo Pirrone conceived of the paper; All authors discussed the material; Angelo Pirrone run the experiment; Angelo Pirrone conducted all analyses; All authors drafted the paper and all authors approved its content.

Chapter 4: Pirrone, A., Marshall, J. A., & Stafford, T. (in preparation). A Drift Diffusion Model account of the semantic congruity effect.

Contributions: Angelo Pirrone conceived of the paper; All authors discussed the material; Angelo Pirrone run the experiment; Angelo Pirrone conducted all analyses; All authors drafted the paper and all authors approved its content.

Chapter 5: Pirrone, A., Stafford, T., & Marshall, J. A. (2014). When natural selection should optimize speed-accuracy trade-offs. *Frontiers in neuroscience*, 8.

All Frontiers articles are Open Access and distributed under the terms of the Creative Commons Attribution License, which permits the use, distribution and reproduction of material from published articles.

Contributions: James A. R. Marshall conceived of the paper; James A. R. Marshall, Angelo Pirrone, and Tom Stafford discussed the material; James A. R. Marshall developed the formal argument; Angelo Pirrone and James A. R. Marshall drafted the paper and all authors approved its content.

Chapter 6: Pirrone, A., Azab, H., Hayden, B., Stafford, T., & Marshall, J.A. (under review). Evidence for the speed-value trade-off: human and monkey decision making is value sensitive.

Contributions: Angelo Pirrone and James A. R. Marshall conceived of the paper; All authors discussed the material; Habiba Azab run the monkey experiment; Angelo Pirrone conducted the human experiment and conducted all analyses; James A. R. Marshall developed the presentation of the model; All authors drafted the paper and all authors approved its content.

Presentations arising from the work presented in this dissertation:

Pirrone, A., Paranamets, P., Marshall, J.A. & Stafford, T. (2016). Choice magnitude and decision time: investigating magnitude sensitivity in value-based decision making. Poster presented at the 38th Annual Conference of the Cognitive Science Society. Philadelphia, USA.

Pirrone, A., Dickinson, A., Gomez, R., Stafford, T. & Milne, E. (2016). Understanding perceptual judgement in autism using the drift diffusion model. Talk presented at the Conference of the Exper-imental Psychology Society. Oxford, UK.

Pirrone, A. (2016). Decision modelling insights in clinical research and perceptual learning: removing confounds from speed-accuracy trade-offs. Talk presented at UCL. London UK.

Pirrone, A., Marshall, J. A. & Stafford., T. (2014). Evolutionary and computational considerations on the Drift Diffusion Model. Talk presented at Decision Making Bristol 2014, Bristol UK.

Pirrone, A., Marshall, J. A. & Stafford., T. (2014). Equal alternatives, value-based decisions and the Drift Diffusion Model. Poster presented at Fourth Symposium on Biology of Decision Making, 2014, Paris France.

To my father.

Acknowledgments

Firstly, I want to express my sincere gratitude to my supervisors, Tom Stafford and James Marshall. You have been great - no exaggeration. Thank you for your guidance, support and inspiration over these three years. I am sure we will keep in contact for a long time and I look forward to develop interesting projects with you in the future.

I want to thank Kevin Gurney and the people in the Neuroeconomics Network (Alvin Pastore, Luana Nunes, Eleni Vasilaki and Roderich Groß) for providing insightful feedback over the years during our meetings. Seb James and Alex Blenkinsop for the interesting chats in the lab, Dan 'the Icelander' for the long discussions on trippy stuff, Liz Milne for proposing me to get involved in a very interesting project and in general all the people in the Department of Psychology with whom I have interacted over the years.

Also, a special thank goes to my labmates in the Behavioural and Evolutionary Theory Lab, in particular Giovanni Reina and Thomas Bose.

Fra, Anas, Fabio, Gabs, Arnav, Brigette, Joe, Toby and Harry: thanks for all the beers, food and chats.

A massive thank goes to the people in Sheng Li's lab at Peking University where I have had the honour of being visiting PhD student for a semester. It feels like home when I think back to China, thank you for the amazing experience. Sheng Li, Xin Xue, Qi Zhang, Mengyuan Gong, Jia Ke, Li Ya...I might have forgotten how to properly spell your names but I won't forget the kindness and hospitality that you have showed to me at PKU. Look forward to seeing you all again and 'xie xie for everything'.

My wife, Clara: together with all the emotional support, you've been my favourite test subject for all my pilot studies...I know that you now hate Random Dot Motion stimuli, but this is the price to pay for marrying me!

My family; a simple, sincere 'thank you' for all the support.

For sure I am forgetting someone: thank you to you too!

1

Introduction

The central topic of this dissertation is decision making: the selection of an option from a set of alternatives. In particular, we are interested in the cognitive processes underlying a decision. Some real life examples of the decisions we make everyday include deciding what to buy at the supermarket, or deciding whether an ambiguous stimulus represents a threat or not. In these situations, the decision maker is assumed to accumulate evidence for each alternative until a decision is made in favour of one option (Gold & Shadlen, 2001). For the sake of simplicity, especially mathematical simplicity, we - as most of the researchers in the field - focus on a particular type of decision: choosing one alternative from a set of two alternatives.

In order to investigate the cognitive processes underlying decision making, researchers can use computational models of decision making. Such models formalise mathematically the contribution of each single process underlying a choice. In this dissertation, we show the benefits of applying computational models: differently from verbal theories, computational models do not rely on inference, but allow a better clarity and completeness, and make quantitative and logically valid predictions about theories by simulating the cognitive process itself, mainly through a computer program (Fum et al., 2007). The benefits of this approach are clear if we consider, for example, the case of categorisation research, in which conclusions based on verbal reasoning (which have influenced the research community for years) have been proven invalid once the theories have been formalised mathematically and tested (for details see Busemeyer & Diederich, 2010). A further important feature of computational models is that of serendipity and emergence (O'Reilly & Munakata, 2000), meaning that computational models allow a new understanding of phenomena that are not accessible otherwise.

The goal of this first chapter is to provide an introduction to the background of our work and introduce some key concepts needed to understand our work. A more specific literature review is then presented within each chapter. The unifying theme of the thesis is the use of computational models of decision making that we have used in order to test or generate hypotheses, however we analyse heterogeneous domains, from perceptual judgement in Autism Spectrum Disorder to adaptive decision making inspired by the behaviour of honeybees. As a consequence of the heterogeneity of the work, an extensive literature review presented in one single introductory chapter would be an odd assortment of concepts of little value.

This dissertation - except for the introductory and concluding chapters - is based on five manuscripts that are either under review, in press or already published.

The chapters of this dissertation, unless otherwise stated, are based on the following original publications:
Chapter 2: Pirrone, A., Dickinson, A., Gomez, R., Stafford, T., & Milne, E. (in press). Understanding perceptual judgement in autism using the drift diffusion model. *Neuropsychology*.
Chapter 3: Pirrone, A., & Stafford, T. (under review). A Drift Diffusion Model decomposition of task-irrelevant perceptual learning.
Chapter 4: Pirrone, A., Marshall, J.A., & Stafford, T. (in preparation). A Drift Diffusion Model account of the semantic congruity effect.
Chapter 5: Pirrone, A., Stafford, T., & Marshall, J. A. (2014). When natural selection should optimize speed-accuracy trade-offs. *Frontiers in neuroscience*, 8.
Chapter 6: Pirrone, A., Azab, H., Hayden, B., Stafford, T., & Marshall, J.A. (under review). Evidence for the speed-value trade-off: human and monkey decision

making is value sensitive.

1.1 The Drift Diffusion Model (DDM)

Perceptual decision making research investigates the process by which perceptual information is combined and guides our behaviour. In a classical experiment in decision making, subjects are presented with a cloud of moving dots, and they have to decide whether there is a coherent motion in the stimulus, usually towards'left' or 'right' (Ball & Sekuler, 1982; Shadlen & Newsome, 2001; Ho et al., 2009). This - only apparently - simple task, defined as a Random Dot Kinematogram (RDK), allows the experimenter to investigate important features of perceptual decision making. For example, the experimenter can vary the percentage of dots moving coherently in one direction, and investigate how the performance of subjects is affected when the difficulty (i.e., the percentage of dots moving coherently) varies. Furthermore, the experimenter can vary the instruction of the experiment, for example by asking a group of subjects to be as fast as possible in making a decision, while asking a second group to be as accurate as possible in making a decision. In this way, the experimenter can investigate the effect of different instructions on the performance of subjects. Computational models of choice (e.g., Ratcliff & McKoon, 2008; Usher & McClelland, 2001; Wagenmakers et al., 2007) have brought to the elucidation of the basic principles underlying perceptual decision making.

One of the key concepts of our work is a computational model of decision making, known as the Drift Diffusion Model (Ratcliff & McKoon, 2008). This computational model of choice has a long history - it was firstly presented in Ratcliff (1978) - and over the years has been refined and applied to a wide variety of tasks and domains (Ratcliff & McKoon, 2008), such as (but not limited to) perceptual judgement (Ratcliff, 2002; Thapar et al., 2003; Voss et al., 2004), memory (Ratcliff, 1978; Ratcliff & McKoon, 1988), signal detection (Ratcliff & Rouder, 1998; Ratcliff et al., 1999, 2004b) and has also been applied to the description of the integration of sensory signals towards a motion-discrimination decision in monkeys (Gold & Shadlen, 2002; Shadlen & Newsome, 2001). In the DDM the decision maker integrates difference in evidence supporting two alternatives until a certain positive or negative threshold is crossed, and a decision is made in favour of that alternative.

In its simplest formulation, defined as 'the *reduced* version', the DDM is the continuous case of a random walk process (Bogacz et al., 2006) and is described by the following equation

$$dx = \mu dt + \vartheta dW, x(0) = 0 \tag{1.1}$$

where dx is the increment in evidence in a small time window dt, μ denotes the mean increase in evidence per unit time and ϑdW denotes an independent Wiener process with mean zero and variance $\vartheta^2 dt$ (Bogacz et al., 2006).

Interestingly, the DDM - in its *reduced* version - implements the Sequential Probability Ratio Test (Wald, 1947; Wald & Wolfowitz, 1948), which is the procedure that gives the shortest decision time given a fixed error rate in a two-alternatives forced-choice task (Bogacz et al., 2006). It is possible to demonstrate (Bogacz et al., 2006) that as discrete samples are taken more frequently and one approaches continuous-time sampling of a variable, the SPRT converges to Equation 1.1. In this way, the DDM is statistically optimal for stationary distributions of evidence in conditions in which the subject has to manage a speed-accuracy trade-off (Bogacz et al., 2006). Given this feature of the model, the DDM not only represents a descriptive model of decision making, but has been proposed also as a a normative model (Basten et al., 2010) towards which, under the influence of natural selection, the decision maker may be supposed to have evolved (but see Pirrone et al., 2014).

It is useful for the purpose of the dissertation to restate that 'optimal', with regards to the DDM, means that the DDM allows the fastest decision for a fixed error rate. Of course, optimal does not mean that the DDM is optimal in general, but only in the specific case in which speed and accuracy have to be optimised, a scenario that describes many types of decisions - such as for example, making a fast and accurate decision while driving and encountering an obstacle on the road. A main theoretical and empirical point of this dissertation, that will be developed in detail in the future chapters, is that in many decision making settings (e.g., choosing between sources of food), subjects should optimise different criteria rather than the speed-accuracy trade-off.

A further reason for the popularity of the DDM is that, as shown by Bogacz et al. (2006), other prominent models of choice, under specific parametrization implement or approximate the DDM, with the exclusion of race models (Vickers, 1970) - models with one accumulator for each alternative that accumulate

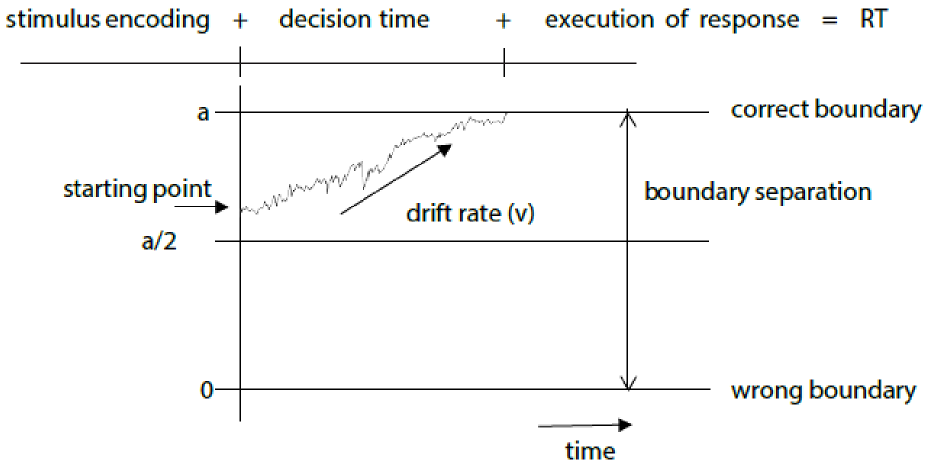


Figure 1.1: Graphical representation of the DDM. [Image taken from Pirrone, Dickinson, Gomez, Stafford and Milne (in press).]

evidence but do not inhibit each other.

Although there are numerous variants of the DDM, throughout this dissertation we will focus in particular on the *extended* version of the DDM as formalised in Ratcliff & McKoon (2008), a more refined and psychologically plausible version of the *reduced* DDM. From now on, unless stated otherwise, when we refer to the DDM, we refer to the *extended* version.

A DDM process is determined by seven parameters (Ratcliff & McKoon, 2008; Vandekerckhove & Tuerlinckx, 2007), as shown in Figure 1.1 and in Table 1.1.

The first, denoted by a, is the boundary separation and it captures the distance between the two thresholds for a decision. When a is small the decision is faster but less accurate since, given noisy fluctuations in the accumulation of evidence, it is more likely to end up at the wrong boundary; when a is large the decision is slower and more accurate. An interpretation for this parameter is therefore the

symbol	parameter	interpretation
a	boundary separation	speed-accuracy trade-off
v	drift rate	stimulus discriminability
z	starting point	bias towards a response
ter	non-decision time	stimulus encoding and execution of motor response
eta	inter-trial variability in v	variability in stimulus discriminability or in attention
s_z	inter-trial variability in z	variability in bias towards a response
s_t	inter-trial variability in ter	variability in non-decision time

Table 1.1: A list and interpretation of the seven parameters of the DDM.

trade-off between speed and accuracy for a decision. Second, is the starting point of evidence accumulation, denoted by z. An interpretation for this parameter is the bias for either response; if z is not equidistant from the boundaries but nearer to the one of the two limits, the subject will be 'biased' to make the choice corresponding to the nearer boundary; when the accumulation of evidence starts at $a/2$ the process is unbiased. In the case of a biased process, fast reaction times (RTs) towards the nearer boundary and slow RTs towards the opposite boundary are predicted, given that the distance from the decision boundary is small in one case and large in the other. Third is the inter-trial variability of z, defined as s_z . Fourth is the drift rate, denoted as v, which represents the mean rate at which information is accumulated over time. This parameter can be interpreted as the quality of the stimulus and the amount of information carried by it for the perceiver. Experimental conditions for which the correct decision is 'easy' will have a higher drift rate compared to more difficult conditions. Also, a further interpretation of this parameter is the sensitivity of a subject towards a stimulus. The accumulation of information varies according to the drift rate and to a fifth parameter, the inter-trial variability in drift rate, denoted by eta. This parameter can be interpreted as the variability in attention or motivation of the decision maker or, in the case of changing stimuli, it can be thought of as the variability in stimulus quality. The last two parameters of the DDM refer to the non-decisions time, since the decision maker has to encode the stimulus and execute the motor response when making a decision. The non-decision component of a RT is denoted by *ter* and its inter-trial variability is defined as s_t.

It is interesting to note that the DDM can account for the full range of correct and incorrect RTs and for the probability of correct and wrong answers, while for example, other diffusion models that have been proposed, account only for accuracy and for RTs of correct choices (Link, 1975; Link & Heath, 1975). Additionally, the DDM offers several advantages in terms of the relation between model parameters, experimental design, and wider theoretical interpretation. The main parameters of the DDM have clear interpretations in terms of psychological processing (e.g., the speed-accuracy trade-off is reflected in the separation of the decision thresholds). Model fitting using the DDM tends to reveal single parameters changing their values to track changes across experimental conditions (see Bogacz et al., 2006; Ratcliff & McKoon, 2008; Krajbich et al., 2010). Inter-related to both of these, the intuitive nature of some aspects of DDM function means that changes to experimental design can often produce clear predictions in terms of DDM parameter change.

1.2 VALUE-BASED DECISION MAKING

Behavioural economics, as it is possible to understand from the definition of the field itself, studies how social and psychological factors influence economic decisions. The main difference with the classical economic approach is that, while economists use a normative approach in describing *how decisions should be made* to optimise a reward function - by assuming that subjects are rational decision makers (Neumann et al., 1944; Von Neumann & Morgenstern, 2007) -, behavioural economics describes *how decisions are actually made* by using experiments that emulate real-life decision problems (Kahneman & Tversky, 2000; Todd & Gigerenzer, 2003). This field of research has lead to the understanding of the heuristics that subjects use in different choice context (Holyoak & Morrison, 2005). However, these explanations only postulate disparate heuristics for different choice settings and cannot describe the time course of a decision. In an attempt to overcome such limitations, recently, sequential sampling models of choice widely studied in the perceptual decision making field, have been used to explain the time course of preference formation in value-based choice in which the subject is no more integrating evidence over time, but value instead, as for example in making a decision between different sources of food.

Tasks involving value-based decisions (Krajbich et al., 2010; Krajbich & Rangel, 2011; Milosavljevic et al., 2010; Krajbich et al., 2012) are different from the usual paradigms used in perceptual decision making; while in a classical RDK experiment stimuli are constantly changing and so the noise is intrinsic to the stimuli, in the experiments used to investigate value-based choices (Krajbich et al., 2010; Krajbich & Rangel, 2011; Milosavljevic et al., 2010; Krajbich et al., 2012) the stimuli (e.g., images of snacks) are non-stochastic, in the sense that the image is non-changing. However, the noisy signal about the value of the stimuli is hypothesized to be generated internally. The process is then a classical diffusion process; the difference in value of the two alternatives is integrated and when the difference crosses a threshold a decision is made. In this way, even though the nature of the comparison is different from a classical perceptual decision making task, it has been shown that the computational problem faced by the decision maker is similar (Krajbich et al., 2010; Krajbich & Rangel, 2011; Milosavljevic et al., 2010; Krajbich et al., 2012). Direct evidence that participants sample noisy values of a choice over time is provided by studies that have focused on how value is represented and integrated (e.g., Towal et al., 2013; Hare et al., 2011; Philiastides et al., 2010; Platt & Glimcher, 1999). For example, Hare et al. (2011) showed that activity in brain areas associated with value-based decision making (the dorsal medial prefrontal cortex and the intraparietal sulcus), correlates with the predictions of a DDM-like neural model.

An important study that has investigated value-based decisions is the one by Krajbich et al. (2010). In this study, subjects had to choose among two stimuli representing food; each stimulus was previously rated by subjects using a Likert scale. The authors created and tested a computational model of value-based binary choice, which is an extension of the DDM in which visual fixation biases the value integration process by a discount parameter on the non-fixated option's value. This model can capture the relationship between fixation and the choice made by subjects in the experiment. Furthermore, the model produced the same fixation biases observed, which supports the hypothesis that fixations may play a causal role in the integration and comparison of the value of the alternatives (cf. Armel et al., 2008). In Krajbich & Rangel (2011), the results of Krajbich et al. (2010) are generalised to trinary value-based choice. Finally, in Krajbich et al. (2012) the same model presented in Krajbich & Rangel (2011) is tested and validated in the case of simple purchase decisions.

The studies presented above, as well as other similar investigations (e.g., Louie, 2013), explicitly support the hypothesis that the DDM can be a unifying computational framework for describing both perceptual decision making and value-based decision making.

1.3 Honeybee Decision Making

A study of selection of nest sites in honeybees (Seeley et al., 2012) has inspired a model of decision making (Pais et al., 2013). This model is one of the prominent aspects of the second part of this dissertation and will be briefly presented here, and in more detail in the future chapters. Before introducing the model, it is useful to describe very briefly the world of honeybees and the problem of nest-site selection. We refer in general to honeybees, however the following studies and observations are based on the most common of species, *Apis millifera*. In an honeybee swarm it is possible to divide the population according to three different roles: the queen bee that has a reproductive role; male bees with the role of reproducing with queen bees from near nests; and a big portion of female worker bees with the fundamental role of foraging, constructing the hive and finding new suitable nest sites (Seeley, 2010). During spring, several workers leave the hive with the queen to form a new colony. In Seeley (2010) there is a detailed and accurate description of the organization and life of a colony of honeybees; of interest for us is the problem of nest selection since it helps to reveal some mechanisms of decision making which, as we will see later, are not accounted for by a whole family of sequential sampling models, including the DDM.

Following Pais et al. (2013), the mechanism of nest selection by honeybees can be formally described by a system of stochastic ordinary differential equations. Although in Pais et al. (2013) this model is described in terms of nest site-selection by honeybees, it can represent any decision making system in which two imperfect accumulators compete to reach a threshold for a decision and are activated and inhibit each other by the value of each of the integrators. Pais et al. (2013) in their work, through approximations and stochastic simulations, have analysed in detail the decision making dynamics emerging from this system and we briefly recall one of the very fundamental results.

The decision dynamics of the model depend on three parameters; (i) the strength of cross-inhibition between evidence accumulating populations (ii) the difference in value between the two alternatives and (iii) the mean value of the two alternatives (Pais et al., 2013). This means that, assuming everything else constant, if the mean value of the alternatives increases, the decision is faster. Surprisingly, this is a feature of decision making that is at odds with models such as the DDM, which by accumulating *difference* in evidence, disregards the mean value of the alternatives. In the DDM, two conditions that have the same difference in value, but different mean value, would be treated as equivalent. For example, assume that a decision maker is presented with two alternatives, which are poor in value and with a fixed difference. If the decision maker is then presented a condition that maintains the same perceived difference in value but has a higher overall value, Pais et al. (2013) model predicts that in this second case the decision maker would have faster RTs. The DDM instead considers the two conditions of this example as being equivalent given that, working by integrating difference in value between the two alternatives, the information regarding the magnitude of the alternatives is not taken into account. It is important to restate that the DDM, works by integrating difference in evidence between the two alternatives and that this feature of the DDM, magnitude insensitivity, has been also addressed elsewhere (e.g., Teodorescu et al., 2015; Tajima et al., 2016).

The dynamics arising from Pais et al. (2013) model allowed us to generate insights into adaptive decision making and show the existence of a specific feature of decision making that cannot be accounted for by celebrated models of choice, magnitude sensitivity.

1.4 Overview of the dissertation

In the first part of this dissertation, we show the benefit of applying a computational decomposition of data, with a particular focus on the DDM, to various domains. Many cognitive processes underlie a decision (e.g., the trade-off between speed and accuracy adopted by the subject, the sensitivity of subjects to a specific stimulus etc) and a computational decomposition of a decision allows to quantify the contribution of each single cognitive process. As we will show, this approach provides crucial insights in some domains.

In the second chapter we apply the DDM to perceptual decision making in Autism Spectrum Disorder (ASD). Research investigating perception and cognition in ASD has focused on RTs and accuracy alone. However, this approach does not allow to quantify the contribution of each specific underlying cognitive process, for example the speed-accuracy trade-off, and can result in misleading conclusions of differences in perceptual sensitivity between groups which may be instead due to differences in other processes (e.g., response criterion or the time to execute the motor response). For the first time, we used the DDM to provide a principled reconciliation of the speed and accuracy of perceptual decision making in individuals with and without ASD performing an orientation discrimination task and we found that: (i) participants with ASD prioritise accuracy over speed, (ii) participants with ASD have a longer non-decision time component and (iii) discriminability does not vary between groups. This result has significant implications for previous studies that on the basis of slower RT claimed an impairment for ASD subjects, or on the basis of greater accuracy claimed an enhancement for ASD subjects in perceptual abilities.

In the third chapter, we perform a decomposition of task-irrelevant perceptual learning, learning to better discriminate below-threshold stimuli to which subjects are merely exposed but are irrelevant to the task. Our results show that task-irrelevant perceptual learning results in an increase in sensitivity for the trained direction accompanied by a decrease in response criterion. We show that this pattern could lead to the mistaken conclusion of no learning effect if only accuracy scores are analysed, as a decrease in response criterion can mask a simultaneous increase in sensitivity. Together with this variation, that is consistent across subjects, other cognitive processes vary non-systematically between the pre-test and the post-test session making an analysis only based on accuracy or reaction times potentially misleading. Our analysis shows the benefits of isolating different cognitive processes that together are responsible for the output of a decision.

In the fourth chapter we analyse a phenomenon known as semantic congruity effect, which refers to the finding that (i) subjects are faster in judging the bigger of two big items or the smaller of two small items, than judging the smaller of two big items or the bigger of two small items and (ii) when the relative size of a standard and a target stimulus coincides, the judgement is facilitated. For this phenomenon, many verbal theories have been proposed. Two families of theories make different predictions regarding the explanation of this effect: one theory states that this effect arises as a consequence of starting point variations while a second theory states that this phenomenon arises as a consequence of variations in drift rate. These two theories can be easily tested within the framework of computational models of choice. With our investigation, we show that the semantic congruity effect results in an increase in rate at which evidence is accumulated in case of congruency between the relative sizes of the standard and target stimuli. Again, we show the benefit of a computational decomposition of perceptual decision making.

In the second part of the dissertation, inspired by the model proposed by Pais et al. (2013), we focus on an aspect of decision making that has important theoretical consequences for models of choice: magnitude sensitivity. Some computational models of choice work by integrating a function between the two alternatives (i.e., ratio or difference for example) and they disregard the overall magnitude of alternatives. In the fifth chapter we focus on a theoretical exposition of why we should expect decision making to be sensitive to the magnitude of the alternatives on the basis of evolutionary arguments, and why this poses a challenges to all models that disregard the magnitude of alternatives, DDM included.

In the sixth chapter, we provide evidence for magnitude sensitivity in decision making for both humans and monkey decision making. Manipulating the magnitude of alternatives can be tricky: if the magnitude of the alternatives is increased while the same *physical* difference between alternatives is kept constant, the two conditions are not psychologically comparable given the well established result in psychophysics of Weber's law (Householder & Young, 1940). On the other hand, deciding a priori which physical quantities would result in the same perceived difference is virtually impossible. Our solution was using the only condition for which it is possible to be sure that even though the magnitude of alternatives is increased, the perceived difference remains the same: equal alternatives. For this particular set of alternatives, both the physical and perceived difference in evidence remains the same when the magnitude of alternatives is increased; in fact, both the physical and perceived difference in evidence is null.

In our investigation we found that increasing the magnitude of equal alternatives results in a decrease in reaction times, both for humans performing a perceptual decision making task and for monkeys performing a value-based decision making task. For such conditions, models that integrate a function between the two alternatives predict magnitude insensitivity and cannot explain this result. In this chapter we discuss the importance of this result for adaptive decision making and for the falsifiability of computational models of choice, in particular for the DDM. Because of its widespread success in matching empirical data, alignment with statistically optimal evidence accumulation, and its centrality within the family of models of decision making, understanding the features and limitations of the DDM algorithm is of great importance.

Finally, in the last chapter, we provide a general conclusion and present future research questions for each of the study presented in this dissertation.

Part I

Decision modeling insights in cognition

2

Understanding perceptual judgement in autism using the DDM

ABSTRACT

In this study, we performed, for the first time, a DDM decomposition of perceptual decision making in ASD. In ASD research involving 2AFC tasks, differ-

ences in reaction times and accuracy alone have been interpreted as perceptual impairments or enhancement in different domains, without taking into consideration factors that underlie a decision such as the speed-accuracy trade-off or the non-decision time necessary to encode the stimulus and execute the motor response. In our investigation, an orientation discrimination task, we found that (i) the drift rate, measure of sensitivity to a stimulus, was not statistically different between ASD participants and controls, (ii) subjects with ASD had a higher non-decision time component than controls, (iii) subjects with ASD had a more conservative criterion for a response than controls. This result has fundamental consequences for research investigating perception in ASD, and we believe it can lead to a re-consideration of the whole field. In particular, differences that have been interpreted as perceptual differences, might instead be differences in response criterion and motor response.

2.1 Introduction

Autism Spectrum Disorder (ASD) was firstly described by Kanner et al. (1943) and 70 years of research have seen a tremendous interest of the scientific community in this neuropsychiatric disorder. ASD is characterised by anomalies in social interaction and language (Kanner et al., 1943), reduced behavioural flexibility (D'Cruz et al., 2013) and repetitive motor mannerisms (Turner, 1999; Watt et al., 2008).

Numerous investigations in ASD have employed 2AFC tasks in order to investigate specific areas of cognition and surprisingly, both impairments and enhancements have been reported, depending on the field. For example, Deruelle et al. (2004) have reported an impairment in facial matching while O'Riordan et al. (2001) have suggested enhanced visual search in ASD subjects. One field in which conflicting results have been reported (including impairments, enhancements or no difference compared to a control group) is motion discrimination (Milne et al., 2005; Bertone et al., 2003; Milne et al., 2002; Dickinson et al., 2016). On the basis of these results supporting impairments or enhancements, theories regarding the neural substrates of ASD have been developed (Pellicano et al., 2005; Pellicano & Burr, 2012; Deruelle et al., 2004; Dickinson et al., 2016; Simmons et al., 2009).

One strong limitation of the studies reported above is that they have focused on accuracy – as it is common in psychophysics - or RTs alone. On the basis of differences between groups in RTs or accuracy, these studies have inferred cognitive differences between groups in information processing, without performing a computational decomposition of the data that requires combining both measures rather than focusing exclusively on accuracy or RTs. Not taking into account the different cognitive processes that underlie a decision may lead, and has led indeed (examples below), to ambiguous interpretation of data. Let us take an example of two groups that have the same sensitivity to a stimulus but have a different speed-accuracy trade-off (Heitz, 2014; Pachella, 1973; Palmer et al., 2005; Stone, 2014; Wickelgren, 1977). Group A is faster, and so by necessity more likely to make more errors, while group B is slower and hence more accurate since more time is used to make a decision. A researcher focusing only on accuracy, might be tempted to conclude that group B is better at performing the task than group A, given the higher accuracy. Conversely, a second researcher focusing on RTs, might be tempted to conclude for an impairment of group B compared to group A, given slower RTs. In this example, a difference in response criterion could be misinterpreted as a difference in sensitivity to a stimulus. Together with the criterion for a response and stimulus discriminability, other mechanisms underlie perceptual judgement: the time to execute the motor response, the bias towards a response, the trial by trial fluctuations in motivation etc (White et al., 2010). These factors as well can confound conclusion as for example a difference in the time to execute the motor response (e.g., pressing a button on a keyboard), resulting in a difference in RTs between two groups, can be interpreted inappropriately as a difference in stimulus discriminability.

Although some methods - other than the application of computational models - have been proposed to control for speed-accuracy trade-offs (e.g., the 'efficiency' score, Townsend & Ashby, 1983; which is defined as RT/ percentage correct), such methods either make strong assumptions, for example that speed and accuracy are linearly related (Seli et al., 2013), or have been shown to be unable to detect differences in speed-accuracy trade-offs under specific circumstances (Bruyer & Brysbaert, 2011) .

A principled way of accounting for speed-accuracy trade-offs and the other cognitive mechanisms underlying choices, is the application of computational models of choice. Here we use the Drift Diffusion Model (Ratcliff, 1978; Ratcliff & McKoon, 2008).

In clinical research, the benefits of applying the DDM have been widely shown. DDM applications to ADHD research have shown that this disorder is associated with deficits in information processing (i.e., in the drift rate) rather than with a non-conservative decision criterion setting (Karalunas et al., 2012; Metin et al., 2013). Another application that highlights the benefits of the DDM applied to clinical research, has shown that subjects with aphasia, who usually have slower RTs in lexical decision tasks, have a more conservative criterion for a response and longer non-decision time compared to control subjects (Ratcliff et al., 2004b) when performing a lexical decision task. The result of slower RTs was previously interpreted as a signature of an impairment in information processing, however Ratcliff et al. (2004b) have shown that also in this case the difference between the two groups is in the criterion for a response and in their non-decision time component, rather than in information processing of lexical stimuli (e.g. word vs. non-word) . A further example, commonly reported to show the benefits of the DDM in clinical research, comes from Ratcliff et al. (2006), who have shown that the general finding of increased RTs in older participants in a variety of perceptual tasks is related to a more conservative criterion for a response and a longer time to execute the motor response, invalidating in this way the previous explanation of data according to which higher RTs were interpreted as a perceptual impairment for older participants.

The aim of this investigation is to apply, for the first time, the DDM to perceptual decision making in individuals with ASD in order to investigate whether cognitive processes not related to information processing of stimuli could provide an explanation for the differences in RTs and accuracy between individuals with and without ASD.

2.2 METHOD

2.2.1 PARTICIPANTS

Note: Participants were recruited by Abigail Dickinson and Rosanna Gomez, under the supervision of Dr Elizabeth Milne. Sixty participants without any history of epilepsy, migraine or seizure took part in this study voluntarily; twenty-eight individuals with ASD (7 females) and thirty-two neurotypical (NT, 11 females). Participants were matched on age [age of the ASD group (M = 33.85, SD = 14.24), age of the NT group (M = 34.40, SD = 14.66), p=.89], and non-verbal IQ measures [Matrix Reasoning T-score of the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999), ASD group (M =59.32, SD=7.02), NT group (M= 56.26, SD = 7.92), p=.14]. Participants in the ASD group had a diagnosis of ASD from clinicians working in the NHS on the basis of Diagnostic Statistical Manual or International Classification of Diseases (Fritz et al., 2000) criteria. Participants in the NT group did not have a diagnosis of ASD and did not have first degree relatives with a diagnosis of ASD. Five participants from the ASD group were excluded from the analyses given that three did not finish the task and two subjects were not collaborative, as shown by an inspection of their performance which was at chance level throughout the experiment. Eleven ASD participants and two NT participants were taking medications at the time of the study. Their inclusion in the analyses did not change the general pattern for both the behavioural and model fitting results. Seventeen ASD participants and nine NT participants had previously taken part in studies on orientation discrimination. Participants with ASD performed the fourth module of the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 1999) in order to elicit specific social and communicative behaviours - one participant did not perform the ADOS test. Four out of twenty-four subjects did not meet the clinical cut-off for ASD inclusion on this scale; however, they had scores above the clinical cut-off for The Social Responsiveness Scale (Constantino & Gruber, 2007) and the Autism Spectrum Quotient (Baron-Cohen et al., 2001), together with a diagnosis of ASD, and they were included in the analyses. Both groups, except four individuals for each group, performed the The Social Responsiveness Scale (Constantino & Gruber, 2007) in order to assess social interaction and communication. Four subjects in the NT group obtained a score above the clinical cut-off, and especially one subject had a very high score. Inclusion of this subject did not affect the general results, both for the behavioural analyses and for the parameters of the DDM, and this subject was therefore included in the analyses. Ethical approval was obtained from the Ethics Committee of The Department of Psychology at The University of Sheffield and informed consent was obtained from subjects involved in the study.

2.2.2 Task, Stimuli and Apparatus

Stimuli were generated using MATLAB and Psychtoolbox (Brainard, 1997) and were presented on a laptop with a linearised screen at a refresh rate of 60 Hz and a resolution of 1366 x 768 pixels.

Subjects performed a 2AFC task consisting in an orientation discrimination task. They were presented a fixation cross for 500 ms, followed by the presentation of 2 Gabor patches (99% contrast Gaussian-windowed sinusoidal gratings; 2.5 cycles per degree) on each side of the fixation cross. The background had a mean luminance of 80 cd/m2. For all subjects, the reference Gabor was the one on the left and it was oriented at 45° for all trials. The rationale behind having the reference stimulus only on the left was to keep the task as simple as possible (compared to the situation in which the reference and target change position over trials). However, this choice was a limitation of our study as it is discussed below. Subjects had to decide by button press whether the target Gabor on the right was tilted clockwise or anticlockwise compared to the reference stimulus. They were instructed to use two fingers of their right hand and they could make a decision in their own time as stimuli remained on screen until subjects made a response. No feedback was provided to subjects. There were 2 levels of Rotation (clockwise, anticlockwise) and 5 levels of Angle (3°, 5°, 7°, 9°, 11°). Subjects performed 50 trials per condition in random order and after 250 consecutive trials they could take a break. Before starting the experiment subjects were familiarised with 2 trials per condition.

2.3 Results

2.3.1 Observed variables

In line with the literature that suggests that RTs below 300 ms can be considered fast guesses (Ratcliff et al., 2006), and RTs above 3000 ms are less likely to be generated from a DDM process but can be considered attentional lapses (Ratcliff et al., 2006), we removed 9% of the data from the analyses. A t-test showed that there was a significant difference between the number of remaining trials for ASD subjects (M = 87.3, SD = 11.01) compared to NT subjects (M = 93.59, SD = 8.88), t(55)=-2.291, p =.01. Interestingly, this difference was due to the high number of slow responses for ASD subjects. If only fast RTs were excluded, the two groups did not differ in the number of remaining trials, t(55) = -1.36, p=.18; however, if only slow RTs were excluded, the two groups differed in the number of remaining trials, t(55) = -2.31, p=.03.

A first inspection of the data showed a clear difference in behavioural results as a function of Rotation, with subjects being faster and more accurate when the target is oriented clockwise. For this reason, although our a priori interest was in Group and Angle difference, we included Rotation in the analyses. All analyses presented are linear regression models containing Angle, Rotation, Group and all interactions. For all analyses we report B, the slope of the regression line, and the associated t-test. As expected, Angle affected correct RTs (Figure 2.1), B = 0.032, t=3.996, p<0.001, and accuracy (Figure 2.2), B = 0.032, t=3.996, p<0.001. For both groups, as the difference in angle between the reference and the target increased, RTs decreased and accuracy increased. There was a Group by Angle interaction on RTs, B=-0.366, t=-3.454, p<0.001, but not on accuracy, B=0.023, t=0.612, p=0.541. The Group interaction effect on RTs suggests that subjects in the ASD group were generally slower. Rotation affected accuracy, B=0.412, t=4.754, p<0.001, with subjects being more accurate for clockwise discrimina-

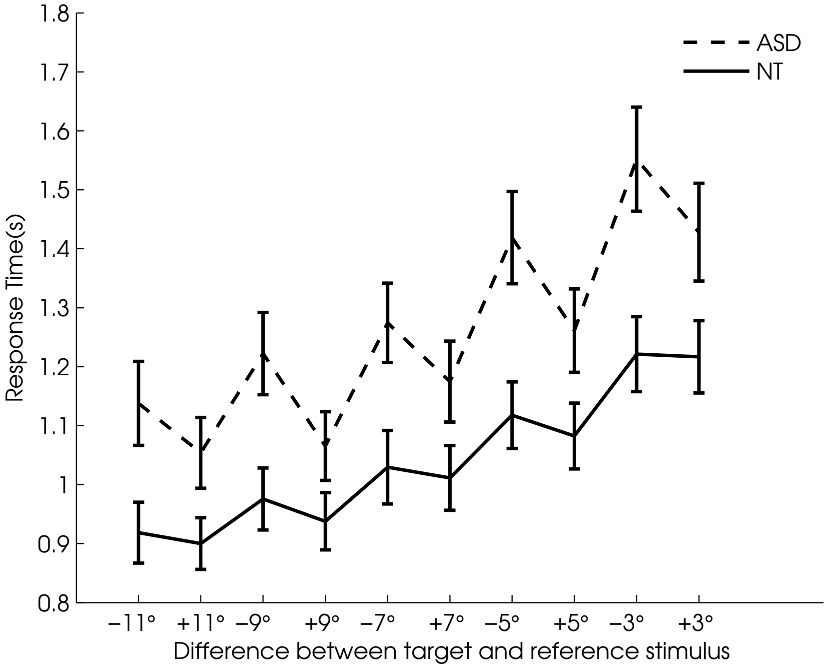


Figure 2.1: Mean response time (for correct responses only) for the two groups of participants (ASD shown by dashed lines) across the different conditions. Anticlockwise rotations are indicated with - ; clockwise rotations are indicated with +. Error bars represent standard error of the mean.

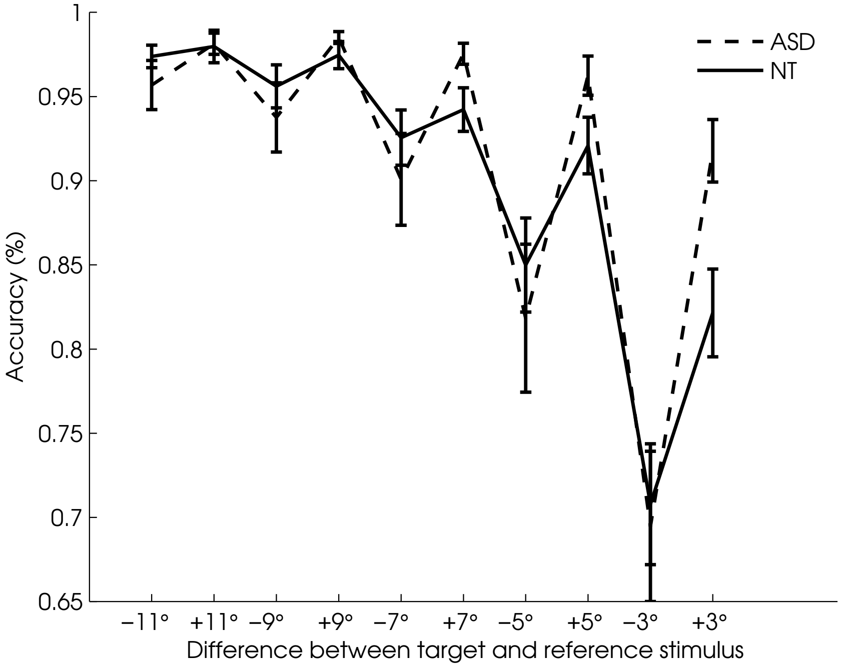


Figure 2.2: Mean accuracy for the two groups of participants (ASD shown by dashed lines) across the different conditions. Anticlockwise rotations are indicated with - ; clockwise rotations are indicated with +. Error bars represent standard error of the mean.

tions than for anticlockwise discriminations. There was an interaction effect between Rotation and Group, B=-0.137, t=-2.582, p=0.010, with ASD participants being generally better for clockwise discriminations and generally worse for anticlockwise discriminations. The effect of Rotation on RTs was not significant, B=-0.293, t=-1.195, p=0.233, and also the interaction of Rotation and Group resulted not significant, B=0.140, t=0.937, p=0.349.

2.3.2 Model Fitting

Estimating the parameters of the DDM requires recovering the full distribution of correct and error responses (Ratcliff & Tuerlinckx, 2002) for each condition. For datasets with few data points and in which the performance of subjects is at ceiling level, meaning that the distribution of wrong responses cannot be estimated, fitting the DDM is problematic (Vandekerckhove & Tuerlinckx, 2008). This is the case of our study in which the mean number of wrong trials is 4 per condition and decreases to 2.5 if +3 and -3 degrees conditions (i.e, the most difficult conditions) are excluded. To overcome the limits of applying the DDM to datasets with few number of trials and/or errors, Wagenmakers et al. (2007) have proposed a method called the EZ-DDM. This method makes two very strong assumptions about the diffusion process; the decision process is assumed to be unbiased, and across trials variabilities are removed. Following concepts introduced in the previous chapters, the decision process is an unbiased, *reduced* DDM. In this way, Wagenmakers et al. (2007), break down the complexity of the fitting procedure and derive analytically values of the drift rate, boundary separation and non decision time, via three simple equations that take as input the mean correct response time, the variance of correct response time and the response accuracy. Another limit of this method is that it does not allow to constrain parameters across conditions and it generates estimates of the three parameters (drift, boundary and non-decision time) separately for each condition. A correction is needed in EZ when accuracy is at ceiling in order to avoid infinite terms in the equations that calculate the parameters; when the probability of a correct response was 1, we replaced this value with $1 - 1/n$, where n is equal to the total number of trials for each condition (i.e., 50 trials, if no trials were removed during the outlier cut-off procedure).

Although we fitted each condition separately, we reasoned that since subjects were presented with conditions of different difficulty in random order, they could not adjust their criterion for a response before each trial was presented and for this reason, for each participant, we computed an average boundary separation based on the average across conditions. At the same time, for each participant, we computed a mean non-decision time component based on the average across conditions.

The drift rate, the parameter that reflects the sensitivity to the stimulus, was averaged across all individuals in each group, and it is plotted against judgement difficulty in Figure 2.3. For the parameters of the DDM, we used the same regression model used for the behavioural analyses.

As expected, Angle had an effect on drift rate , B=0.009, t=2.372, p=0.018, and Rotation had an effect on drift rates, B=0.117, t=2.836, p<0.005. As the differ-

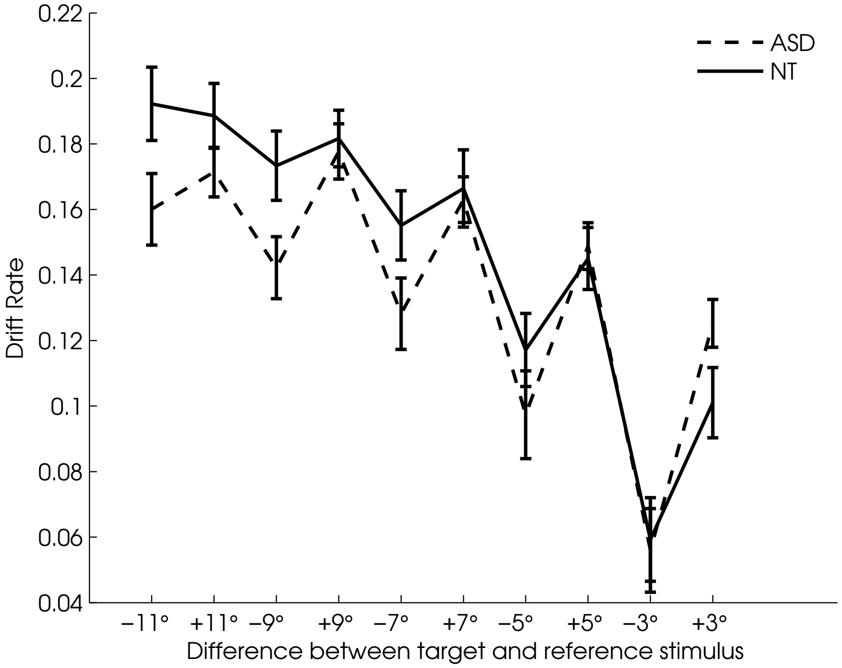


Figure 2.3: Drift Rate for the two groups of participants (ASD shown by dashed lines) across the different conditions. Anticlockwise rotations are indicated with - ; clockwise rotations are indicated with +. Error bars represent standard error of the mean.

ence between the reference and the target stimulus increased, the drift increased and the drift was higher for clockwise conditions compared to anticlockwise conditions.

However, our effect of interest was not significant, as the drift did not differ significantly by group, B=0.002, t=-0.086, p=0.932. None of the interactions between predictors was significant (p>0.145).

Group had a significant effect on both non-decision time (Figure 2.4), B = -0.121, t=-5.377, p<0.001 and boundary (Figure 2.5), B = -0.019, t=-5.351, p<0.001, with the ASD group having in general a more conservative criterion for a response

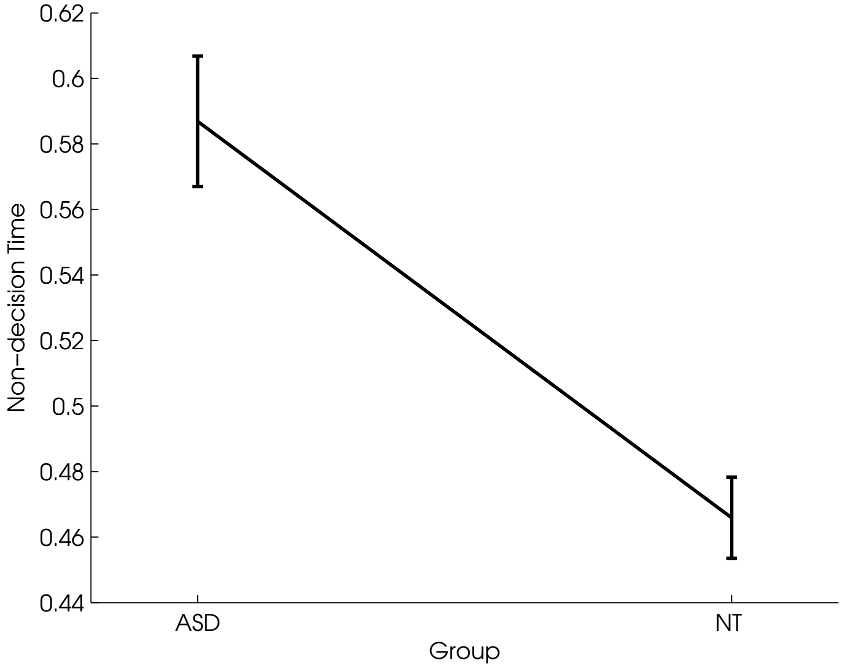


Figure 2.4: Non-decision Time for the two groups of participants (ASD shown by dashed lines) across the different conditions. Anticlockwise rotations are indicated with - ; clockwise rotations are indicated with +. Error bars represent standard error of the mean.

and longer non-decision time (about 100 ms difference as shown by Figure 2.4).

To investigate whether the parameters that we recovered provided a good description of the data in absolute terms, we assessed the goodness of fit of the model. We generated a DDM process with two values for the boundary and the non-decision time based on the average across subjects for the two groups, while the drift was averaged across individuals for each condition, separately for each group. Using this model, in which variabilities across trials were set to zero and the starting point of evidence was set at the same distance from the two boundaries (i.e., the assumptions were the same to those of the model on which we based

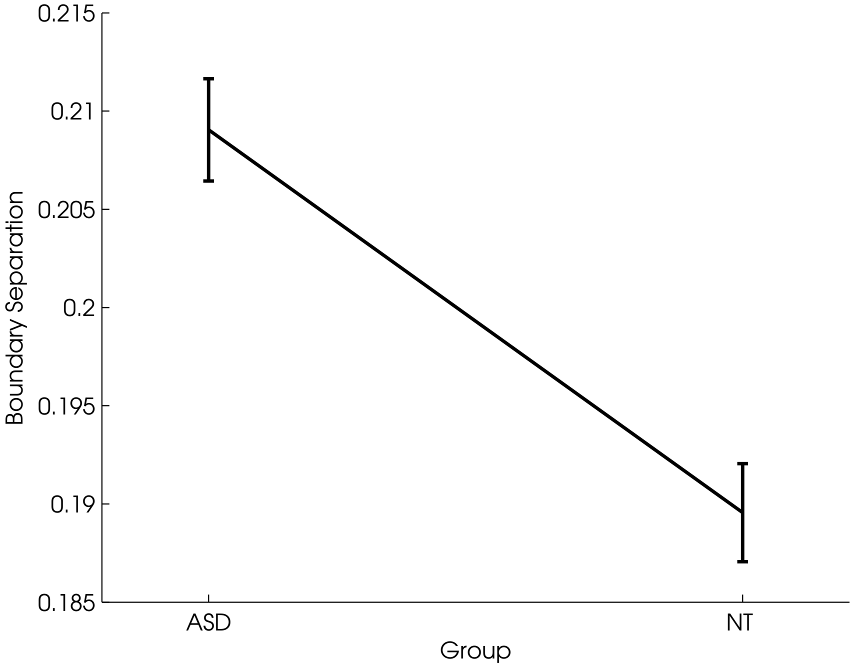


Figure 2.5: Boundary Separation for the two groups of participants (ASD shown by dashed lines) across the different conditions. Anticlockwise rotations are indicated with - ; clockwise rotations are indicated with +. Error bars represent standard error of the mean.

our analysis), we simulated 10000 trials per condition. Figure 2.6 shows comparisons of correct RTs and accuracy between the model and the observed data for the two groups. For both groups, the model can be seen to fit the data well.

In order to investigate whether the process can be assumed as unbiased we compared the difference between mean correct and error responses. In particular, if the process is biased towards one alternative, fast RTs towards that alternative and slow errors towards the opposite threshold are predicted. An inspection of the plots showed no difference in the mean speed of correct and error RTs for each condition. However, this result should be interpreted cautiously given that the mean of wrong RTs is determined by few, potentially unreliable, data points averaged across individuals.

In this study we were limited by the number of trials for each condition; in particular, given that accuracy of subjects was at ceiling level for most conditions, we could not estimate the error distribution and fit the DDM using more complete methods (Vandekerckhove & Tuerlinckx, 2008). Although we showed with a simulation that the model on which our analysis is based can describe the data well, the EZ method is arguably the roughest method to fit the DDM, given that it makes the assumption of the process being unbiased, it removes the across trial variabilities and it does not allow to constrain parameters across conditions. To overcome some of the limitations of EZ, Grasman et al. (2009) proposed a new estimation method, EZ2. However, also this method is not suitable in our case as it does not work when accuracy is at ceiling level and it has been shown to fit the data poorly when drift rates are high and accuracy around ceiling. However, to

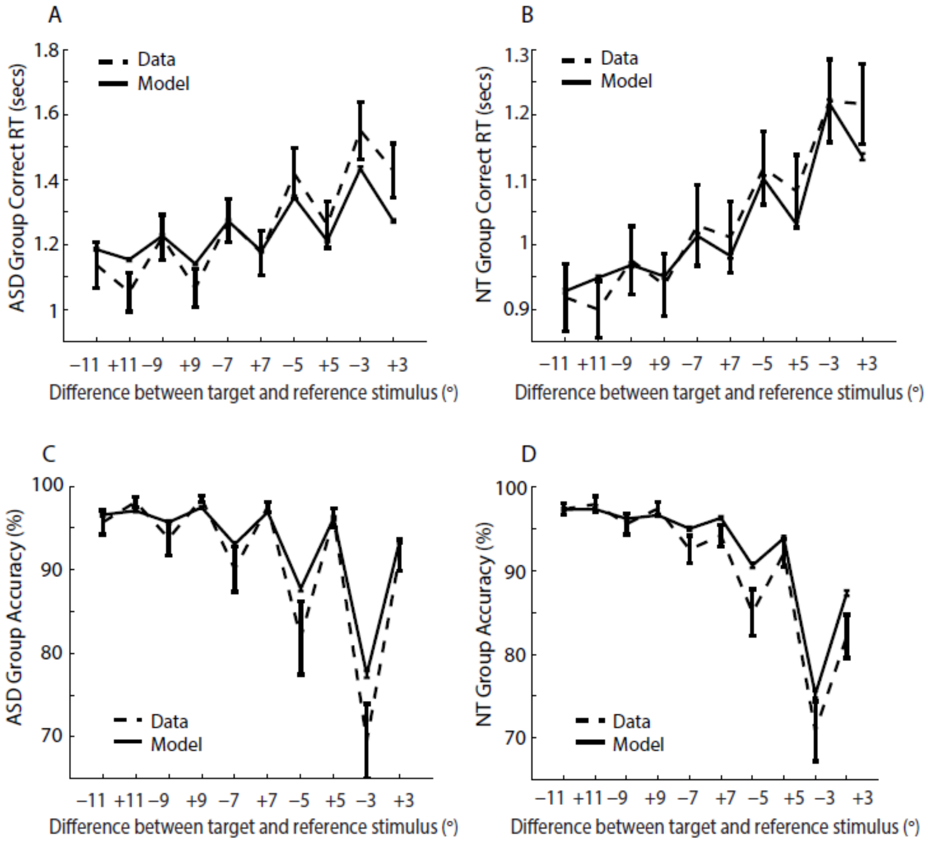


Figure 2.6: Simulated data. A shows comparisons between mean response time (for correct responses only) for the ASD group and for the simulated data. B shows comparisons between mean response time (for correct responses only) for the TD group and for the simulated data. C shows comparisons between percentage of trials correct for the ASD group and for the simulated data. D shows comparisons between percentage of trials correct for the TD group and for the simulated data. Anticlockwise rotations are indicated with - ; clockwise rotations are indicated with +. Error bars represent standard error of the mean.

reassure us that our results are consistent regardless of the fitting method, we decided to fit the full DDM to the pooled dataset for each group. For both groups, all trials from all participants were put together; in this way, from having about 50 trials per condition, we recovered about 1500 trials per condition, making it possible to estimate the distribution of error responses.

To estimate the parameters of the DDM we used the Diffusion Model Analysis Toolbox (Vandekerckhove & Tuerlinckx, 2008) that works by maximising a multinomial likelihood function of the data grouped into 5 quantiles for both correct and error responses (the .1, .3,.5,.7 and .9 quantiles); in the fitting procedure, the lower boundary represented the threshold for a response 'anticlockwise' while the upper boundary represented the threshold for the response 'clockwise'. Similarly to the EZ estimation, we did not constrain parameters across conditions and we removed the across trials variabilities. However, in order to control for starting point effects we allowed the starting point to vary across conditions. Estimated parameters replicated our results. Testing for differences in mean drift, boundary and non-decision time, we found that: the drift rate did not differ significantly between groups $t(18) = .18$, $p = .87$, ASD participants had a more conservative criterion of response, $t(18) = 3.03$, $p = .007$, and longer non-decision time, $t(18) = 4.70$, $p = <.001$. Pooling trials from both groups together can lead to cases in which the pooled data differ greatly from the data averaged across individuals; in our case this would mean that the parameters recovered from the DMAT fitting are not representative of the data on which we based our behavioural analyses. In order to account for this possible confound, we simulated a DDM process with the parameters estimated from the DMAT fitting of the pooled datasets, and we compared it with the behavioural data based on the average across individuals for the two groups; the model in this case fits the data remarkably well, meaning that considering the pooled datasets was a reasonable choice. We also investigated whether the initial trials, during which subjects were learning the task, could affect our results (Browne, 2000): we performed all analyses using only the second half of the trials. For the sake of brevity we do not report here the analysis, however no change in parameter estimates and goodness of fit was observable by fitting the data.

2.4 Discussion

For the first time we have used the DDM to investigate perception in ASD. Subjects performed an orientation discrimination task and our results show that (i) the sensitivity to the stimuli did not differ between the two groups, (ii) subjects with ASD had a more conservative criterion for a response and (iii) the non-decision time component took longer in subjects with ASD.

There are interesting mechanistic interpretations to explain increased non-decision time and increased decision criterion in ASD. The result of increased non-decision time can be linked to the wide literature reporting motor impairments in ASD (e.g., Ming et al., 2007; Bhat et al., 2011). Regarding increased response caution, it is possible that basal ganglia anomalies in ASD (Sears et al., 1999; Qiu et al., 2010) - resulting in anomalies in action selection - prevent subjects with ASD from performing faster and potentially inaccurate responses. Future neuro-imaging research is needed to address these questions directly. Furthermore, future research should investigate whether, within the ASD family of disturbs, clinical measures correlate with DDM parameters. Understanding differences within the ASD family of disturbs could be an important step in the understanding of the contrastive results in the ASD literature.

If we assume that this result can be generalised also to tasks other than orientation discrimination, this result has important consequences for studies that have investigated perception and cognition in ASD and have proposed a difference in sensitivity on the basis of differences in accuracy or reaction times alone. According to our finding, results that have been previously interpreted as differences in information processing should be reconsidered and controlled for differences in response caution and motor response.

Interestingly, the pattern encountered here is different from the one of individuals with ADHD, who show instead a significant difference in drift and not in boundary separation or non-decision time compared to a control group. This difference between ASD and ADHD individuals could lead to better understand the differences and specificities of two clinical disorders which occurrence has been shown to overlap (Polderman et al., 2014).

The difference between clockwise and anticlockwise judgements represents a puzzling result. This response bias was probably given by the fact that the reference stimulus was always oriented clockwise and the target stimulus always appeared on the right of the screen, generating a Simon effect: facilitation towards answering ‘clockwise’. Stimulus rotation did not interact with any other factor in the experiment and the modelling results show that the interaction of Group and Rotation was not significant for any of the parameters, suggesting that our conclusions are not affected by this bias. Future investigations need to avoid a potential bias of this kind, for example by presenting the reference stimulus on top and the target on bottom, with the reference stimulus oriented at 0° rather than 45°.

Regarding the increase in boundary separation, this should be accompanied by an increase in accuracy for the ASD subjects. In our study, the fact that subjects were at ceiling level for most conditions might have masked a significant increase in accuracy. However, although it did not reached significance, for conditions +7°, +5° and +3°, ASD individuals seem to be more accurate than controls, as shown in Figure 2.2. In future investigations, more difficult conditions should be used (i.e., not conditions for which performance is at ceiling) and we believe this would give rise to the observed effect of increased accuracy and increased RTs.

At the moment, we have a new investigation ongoing that overcomes all the above mentioned limitations (number of trials, Simon effect, difficulty of conditions). Results are still preliminary, however promising, and seem to be in line with what we have reported in this study. Regarding number of trials, it is to be mentioned that one common difficulty of performing perceptual decision making tasks with clinical populations is that participants often tolerate a relatively small number of trials. However, here we have shown that, even for a limited number of trials, formal modelling of the underlying processes is beneficial.

In sum, we provide evidence for intact information processing, increased boundary and increased non-decision time component in subjects with ASD. Future research should investigate whether our result is consistent across other domains and whether the result is validated, this would have a strong impact on ASD research. Hopefully, in the foreseeable future, inspired by this work, more labs will apply a computational decomposition of perceptual decision making in ASD research to generate valid conclusions from data.

3

A DDM decomposition of task-irrelevant perceptual learning

Abstract

This study, similarly to the study presented in the previous chapter, addresses whether a phenomenon know as task-irrelevant perceptual learning affects cog-

nitive processes other than the discriminability to a stimulus (e.g., criterion of a response and non-decision time). Even though we change domain, the conclusions of this study resembles those of the previous chapter, in which conclusions from data based only on analysis of accuracy or reaction times are potentially misleading. In this study, change in motion discrimination was assessed after seven days training on a rapid serial visual presentation training task, which included exposure to below-threshold coherent motion that was irrelevant to the task the participant was involved in. Post-training, participants had improved sensitivity for supra-threshold motion discrimination, which was specific to the direction exposed during training. A Drift Diffusion Model decomposition of the effect shows that the improvement is a combination of (i) an increase in rate at which participants accumulate evidence for the direction to which they were exposed and (ii) a decrease in their criterion for a response, meaning that they are willing to make faster but less accurate responses. Together with these differences, which are consistent across participants, other cognitive processes vary non-systematically between the pre-test and the post-test session making an analysis only based on accuracy or reaction times potentially misleading. Our analysis shows the benefits of isolating the different cognitive processes that are involved in perceptual decision making and are affected by perceptual learning.

3.1 INTRODUCTION

It is widely demonstrated that performance increases with practice, both in laboratory and real-life settings. Regarding perception, this phenomenon has been studied from two perspectives; practice research and perceptual learning research. Although these two research lines have obvious commonalities, they have historically focused on two distinct aspects of learning. Practice research has focused on reaction times (RTs) and has documented a decrease in RT related to practice in perceptual tasks (Newell & Rosenbloom, 1981; Logan, 1992), while perceptual learning research has focused on accuracy and has documented an increase in accuracy related to practice in perceptual tasks (Fine & Jacobs, 2002). Recently, a particular type of perceptual learning has captured the attention of researchers, task-irrelevant perceptual learning (TIPL), in which participants learn to better discriminate stimuli to which they are merely exposed but that are irrelevant to the purpose of the experiment (Watanabe et al., 2001; Seitz & Watanabe, 2008). The first example of TIPL comes from Watanabe et al. (2001), in which authors, after testing participants on a motion discrimination task, exposed participants to many days of a rapid serial visual presentation (RSVP) task, on the background of which was presented a random-field motion stimulus with a below threshold coherence in a constant direction across all days of training. Watanabe et al. (2001) found that participants showed an improvement in a post-test motion discrimination task, only for the supra-threshold coherence level in the direction to which they were exposed.

Focusing only on RT or accuracy for a perceptual task has some limitations. Consider a classical perceptual learning task, motion discrimination, in which participants are required to judge whether, within a field of randomly moving dots, a small percentage of dots is moving coherently in one of two directions, say ‘left’ or ‘right’. In this only apparently simple task, different components contribute to those aspects of the decision which can be easily measured, RT and accuracy: (1) the decision criteria adopted by the participant can be more or less conservative, meaning that the participant could focus more on the speed of a decision, and hence be likely to make errors, or alternatively focus more on accuracy, and hence be slower. This is the issue of the ‘speed-accuracy trade-off’, which has received considerable attention previously (Wickelgren, 1977; Heitz, 2014), (2) the RT is a mixture of the decision time and a non-decision time component, which includes time to encode the stimulus and to execute the motor response, for example by pressing one of two buttons on a keyboard, which is independent of any stimulus-characteristics, (3) the participant could be more biased towards answering either left or right, hence having faster RTs and high accuracy for the biased response and slower RTs and lower accuracy for the opposite response, (4) the difficulty of the decision would affect both accuracy *and* RT with participants having faster and more accurate decisions for easy trials and slower and less accurate decisions for more difficult trials, (5) all of the above mentioned mechanisms could vary across trials, either non-systematically, or by these aforementioned decision parameters drifting over the course of an experiment.

If the participant is performing a task on many different days, since the experimenter is interested in the effect of learning over different sessions, it is reasonable to expect that all the above mechanisms could also vary across different days on the basis of factors not related to the experiment (e.g., on one day the participant could be more tired or less collaborative).

Given the ensemble of components at hand in a perceptual task, focusing *only* on RTs or *only* on accuracy is overly simplistic. Unsurprisingly, in some cases, an approach focusing only on RTs or accuracy has even been misleading in generating theories from data. For example, research focusing only on RTs has found differences between faster younger participants and slower older participants. On the basis of this finding, a so called generalised-slowing hypothesis of ageing has been proposed, according to which these data supports the idea of a general deterioration of all processes in the brain of older people (Salthouse, 1996). However, a multitude of studies have applied a computational decomposition of RTs and accuracy in tasks comparing younger and older adults and have shown that, with regards to the components described above, there is no difference in the rate at which the two groups extract information from the stimuli, but older people have a more conservative criterion for a response and in some cases a longer non-decision time component (i.e., they are slower in the motor response of pressing a button to signal a decision). It is this, not generalised-slowing, which leads to slower RTs (e.g., Thapar et al., 2003; Ratcliff et al., 2004b, 2000, 2006, 2010, 2006, 2007). Similarly, Pirrone et al. (2016) have shown that the difference in RTs and accuracy between typically developing and autism spectrum disorder participants can be due to differences in parameters such as the criterion for a response, instead of due to differences in perceptual discrimination.

These two examples show the practical benefit of considering *both* accuracy and RT for perceptual tasks. Arguments from first principles also exist, demonstrating that both accuracy and RT contribute independent information about a decision, and hence combining them allows a more potent analysis (Palmer et al., 2005; Stone, 2014).

To isolate the role of each cognitive process involved in a choice, researchers can adopt computational models of decision making (Ratcliff, 1978; Ratcliff et al., 1999; Usher & McClelland, 2001; Brown & Heathcote, 2008). Computational models of decision making describe the time course of a decision and can isolate the role of each of the cognitive processes involved. Moreover, they can simulate a decision, meaning that the experimenter can use them to make quantitative predictions about a decision making task. In this chapter, we will use the DDM (Ratcliff, 1978; Ratcliff et al., 1999; Ratcliff & McKoon, 2008) as a tool to isolate different components of the processes that contribute to a decision.

Despite widespread success in other domains, the DDM has not been consistently applied in the domain of perceptual learning. In Petrov et al. (2011), the authors performed a DDM decomposition of a fine motion discrimination task. In their study, authors found that perceptual learning was best explained by an increase in drift rate, a decrease in boundary separation and a decrease in both the non-decision time component and its inter-trial variability. In Liu & Watanabe (2012), participants performed a three day perceptual learning coherent motion direction task (i.e., is the RDK all noise or is there some signal?) and authors found an improvement in drift rate *but* with boundary separation decreasing across the days of training. In Dutilh et al. (2009), participants performed a 5 days lexical decision task and authors found that the learning led to an increase in drift rate, a decrease in boundary separation, as well as a significant decrease in non-decision time.

As it is clear from these investigations, RTs and accuracy alone cannot give a full description of the cognitive processes that are most likely to have generated the data, while considering both measures and their distribution can lead to a better understanding of the cognitive processes involved in such tasks. In particular, if perceptual learning is associated with a decrease in the boundary separation, as other studies of task-relevant perceptual learning have found, then assessing perceptual learning via measurement of accuracy will systematically underestimate the true size of perceptual learning (since decreased boundary separation will tend to decrease accuracy, all other things being equal).

Although task-relevant perceptual learning has been decomposed by using the DDM, to our knowledge, no studies have focused on a DDM decomposition of TIPL. In our study, we ran an experiment similar to that presented in Watanabe et al. (2001), and we performed a DDM decomposition of TIPL. Because of the aforementioned multi-component nature of perceptual decision-making, our expectation is that use of the DDM will allow a more accurate assessment of perceptual learning than attention to solely RT or accuracy. Further, the DDM allows us to isolate the component of decision making that reflects a true change to stimulus sensitivity - a change in the drift rate parameter. Because of the potential for non-stimulus related parameters to alter across sessions due to non-experimentally causes factors (such as fatigue or motivation) and because, by their nature, perceptual learning experiments involve testing participants on different days or even weeks, we isolate perceptual learning as an increase in the drift rate for the exposed stimulus (for one participant also the drift of the not exposed direction increased but the increase was greater for the exposed stimulus than for the not exposed stimulus). In this way, we use each participant as their own control, testing them for changes in perceptual decision making for both exposed and not exposed stimuli and thus accommodating non-training related changes in decision parameters.

3.2 Material and methods

3.2.1 Participants

Four right-handed healthy university students (2 males and 2 females, ages: 30, 21, 20, 22 years), with no history of neurological or psychiatric disorders, with normal vision and naïve to the purpose of the study participated voluntarily in the experiment and received a compensation of £50 for their participation. The experiment was approved by the University of Sheffield, Department of Psychology Ethics Sub-Committee, and carried out in accordance with the University and British Psychological Society ethics guidelines. Participants gave their informed consent.

3.2.2 Apparatus

The stimuli were generated on a personal computer using PsychoPy (Peirce, 2009). During the whole experiment, participants had to put their head on a chin rest at a viewing distance of 57 cm from a SONY Multiscan CPD-200ES 17" monitor with a resolution of 1280 x 1024 pixels at a refresh rate of 60 Hz. The experiment was conducted under binocular viewing conditions and participants used a keyboard to make a response.

3.2.3 Motion-Direction Stimuli

We used stimuli similar to those adopted by other studies on task-irrelevant perceptual learning (Watanabe et al., 2001; Seitz & Watanabe, 2008): on a grey background, within a black annulus aperture of 1° - 10°, white dots with a size of 2x2 pixels were moving with a speed of 6°/s and a density of 16.7 $dots/deg^2/s$ on a black background. Signal dots were randomly chosen in each frame, and on each frame, noise dots had a random position. Dots had a limited lifetime of three frames after which they were redrawn in random locations. If any of the signal dots were to move out of the annulus, they were replaced randomly in the stimulus field. The stimuli were generated in real time and two non-cardinal directions were employed in this study, 45° and 135°.

3.2.4 Procedure

The experiment consisted of nine sessions; a pre-test to measure sensitivity for various strengths of motion coherence in the two directions, then seven training sessions consisting of a RSVP task with on the background a random dot motion, and finally a post-test that measured sensitivity for various coherence levels in the two directions, that was equal to the pre-test. Participants came on different days for each session, and could take a maximum of three-days break between sessions.

Due to a computer glitch, data from the first day of training were irrecoverably corrupted for the first two participants who took part in the study. However, participants had performed the whole session and this clearly did not affect the performance in the conditions of interest, the pre-test and the post-test.

3.2.5 Pre/post Motion-Direction Sensitivity Tests

Participants were instructed to pay attention to the stimulus that would be presented for 500 ms and then report as quickly and accurately as possible if the coherent motion was towards up-left (45° on the left with respect to an imaginary vertical reference line) or up-right (45° on the right with respect to a imaginary vertical reference line) by button press. They were instructed to use their right hand index finger to press left on the keyboard for 'up-left', and their middle finger to choose 'up-right'. Participants were instructed that there was always a correct response and were required to fixate the cross at the centre of the screen during the whole task and minimise as much as possible eye movements.

In each trial, a fixation cross in the central circle was presented for 333 ms, followed by the presentation of moving dots for 500 ms, followed by two arrows showing the possible direction of the dots and the text 'Answer:' presented on top of the screen until participants made a response.

Each test stage consisted of 10 blocks x 2 directions (45° and 135°) x 10 motion coherence levels (5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%) x 9 repetitions for a total of 1800 trials and took about 1 hour to complete. Coherence levels were chosen so that for each direction we would have accuracy levels that range from floor to ceiling based on the results of previous pilot studies. During each block the order of presentation of trials was randomised and no accuracy feedback was given to the participants.

After a block of 180 consecutive trials participants were required to take a self-paced break to rest before continuing with this task.

3.2.6 TRAINING SESSIONS

In the training sessions, participants performed a RSVP character identification task and were asked to report in order of presentation two white capital letters (RGB [1,1,1], opacity on a 0 to 1 scale on Python:1, height .9°) in a sequence of 10 capital letters presented in the central circle. Each letter was presented for 33 ms and was followed by a blank interval for 17 ms. Distractors consisted of eight capital black letters (RGB [0,0,0], opacity on a 0 to 1 scale on Python:.3, height .9°). The first and second white letters were presented in one of the first five serial positions and in one of the second five serial positions, respectively. They were determined randomly in each trial. Within the annulus aperture of 1° - 10° participants were presented a motion stimulus in one of the two locations, constant across all training sessions, at a coherence level 5% below their chance level at pre-test, in order to ensure a level reasonably below threshold. For each participant, we computed the motion strength at 50% accuracy by interpolating the psychometric curve predicted by the model free estimation of the psychometric curve described in Zchaluk & Foster (2009), and using MATLAB scripts made available by those authors.

In each trial, a fixation cross in the central circle was presented for 333 ms followed by the presentation of the stimulus for 500 ms followed by a grey screen and the text 'Type in the two white letters' presented on top of the screen until participants responded.

Each test stage consisted of 10 blocks x 108 repetitions for a total of 1080 trials and took about 45 minutes to complete. No accuracy feedback was given to the participants.

After a block of 108 consecutive trials, participants were required to take a self-paced break to rest before continuing with this task.

3.3 RESULTS

3.3.1 BEHAVIOURAL ANALYSES

Figure 3.1 shows the performance of participants in the RSVP over the seven days of training. The performance of participants was mostly stable across the seven days of training. We did not perform any analysis on the RSVP task as our interest is in the TIPL, hence in the change in performance between pre-test and post-test for the exposed and not exposed directions. It is to be mentioned that accuracy during training is much lower than the one reported in Watanabe et al. (2001) of around 60%. Previous pilot studies that we performed, suggest that the performance of participants during training is determined by the contrast of the target, the white capital letters - which in our study is different than that of Watanabe et al. (2001). In our pilot studies, in which the contrast of the target was even lower than the one reported here, this led to a performance even worse of the one of Figure 3.1. However, no relation has been reported between per-

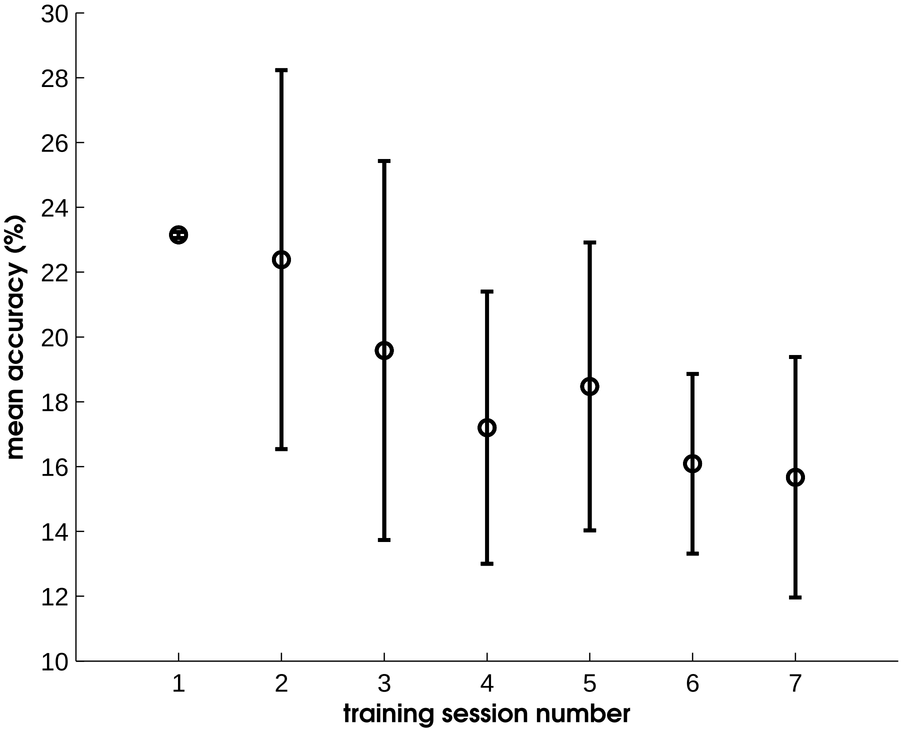


Figure 3.1: Mean percentage correct on the (irrelevant) RSVP task for the four participants across the exposure stage sessions. Error bars are standard errors of the means.

formance in the training and TIPL, meaning that performance during training is irrelevant for the purpose of TIPL.

For all analyses, both behavioural analyses and model fitting, we removed, for each participant, the 2.5% of slowest response, given that a first inspection of data showed the presence of extremely slow outliers. In the following analyses, each subject is analysed separately. Figures 3.2, 3.3, 3.4 and 3.5 show mean RTs and accuracy for each participant for conditions for which they were exposed and for which they were not exposed during the pre-test and the post-test. Participants 1 and 2 were exposed to 45° while participants 3 and 4 were exposed to 135°. T-tests were conducted to investigate overall differences for each participant between the pre-test and the post-test in mean RT and accuracy levels for the exposed and the not exposed directions. Bonferroni corrections were applied on the p-values; this means that probabilities are multiplied by two given that for each dependant variable (e.g., RT and accuracy) we have two subgroups; the exposed and the not exposed direction.

All subjects had a significant decrease in RTs between the pre-test and post-test, for the exposed and the not exposed direction ($p < .001$ in all cases). Participant 1 did not have a change in accuracy for the exposed direction between the pre-test and post-test ($t(9) = -0.34$, $p = 1$) while all other participants had a significant decrease ($p < .001$ in all cases). Regarding accuracy of the not exposed direction, there was not a significant change between pre-test and post-test for any subject ($p > .07$ in all cases).

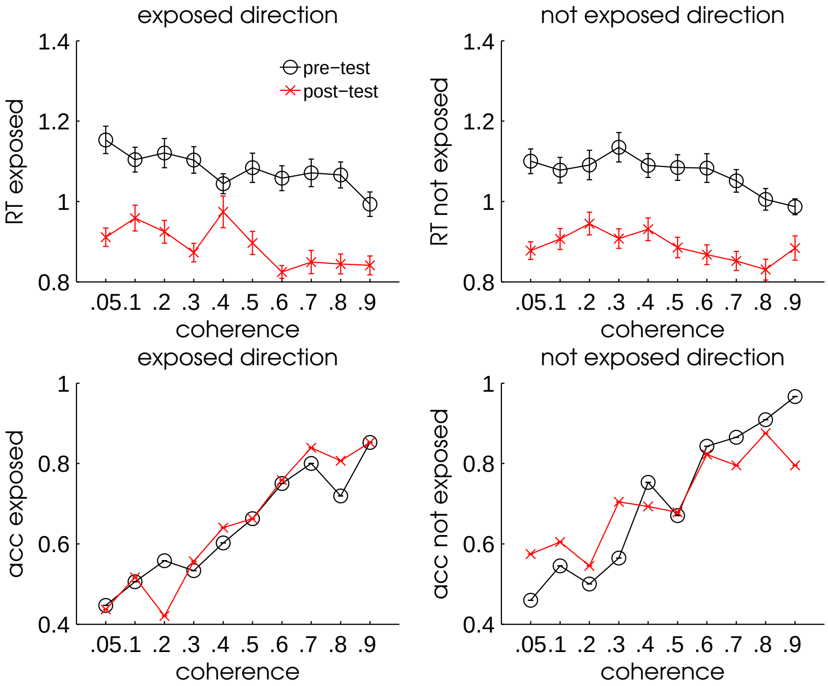


Figure 3.2: (top left) mean RT (± s.e.) for the direction for which participant 1 was exposed, (top right) mean RT (± s.e.) for the direction for which participant 1 was not exposed, (bottom left) mean accuracy for the direction for which participant 1 was exposed and (bottom right) mean accuracy for the direction for which participant 1 was not exposed.

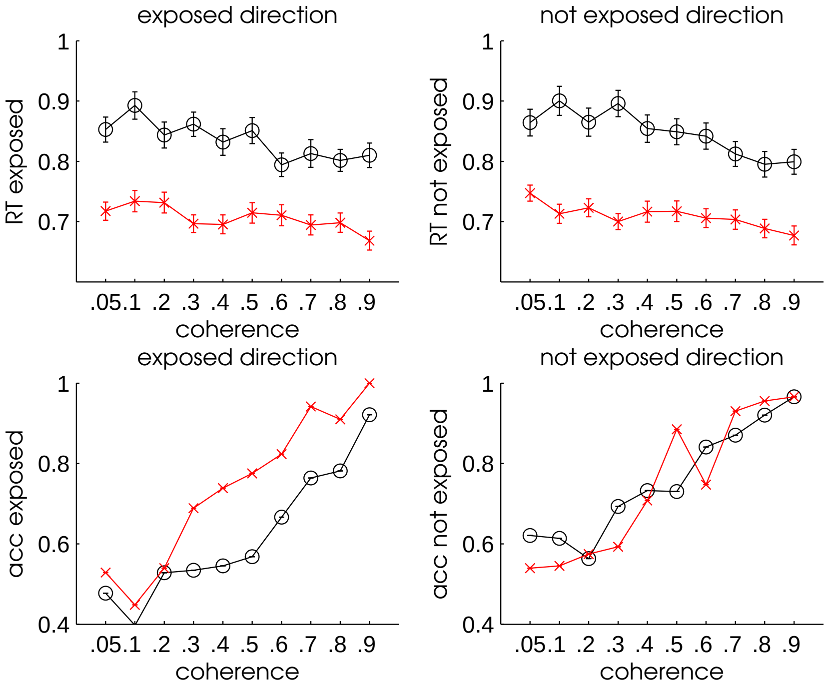


Figure 3.3: (top left) mean RT (± s.e.) for the direction for which participant 2 was exposed, (top right) mean RT (± s.e.) for the direction for which participant 2 was not exposed, (bottom left) mean accuracy for the direction for which participant 2 was exposed and (bottom right) mean accuracy for the direction for which participant 2 was not exposed.

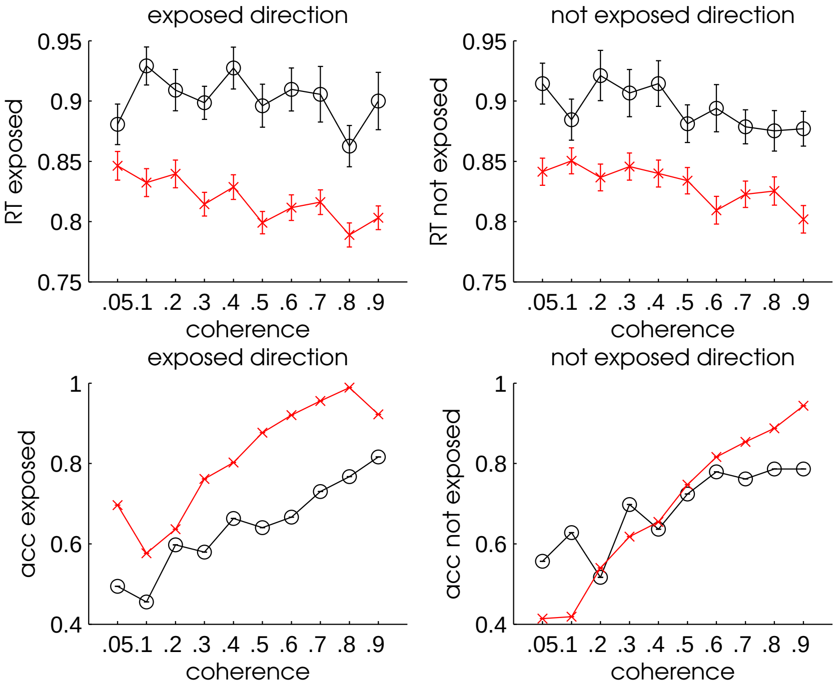


Figure 3.4: (top left) mean RT (± s.e.) for the direction for which participant 3 was exposed, (top right) mean RT (± s.e.) for the direction for which participant 3 was not exposed, (bottom left) mean accuracy for the direction for which participant 3 was exposed and (bottom right) mean accuracy for the direction for which participant 3 was not exposed.

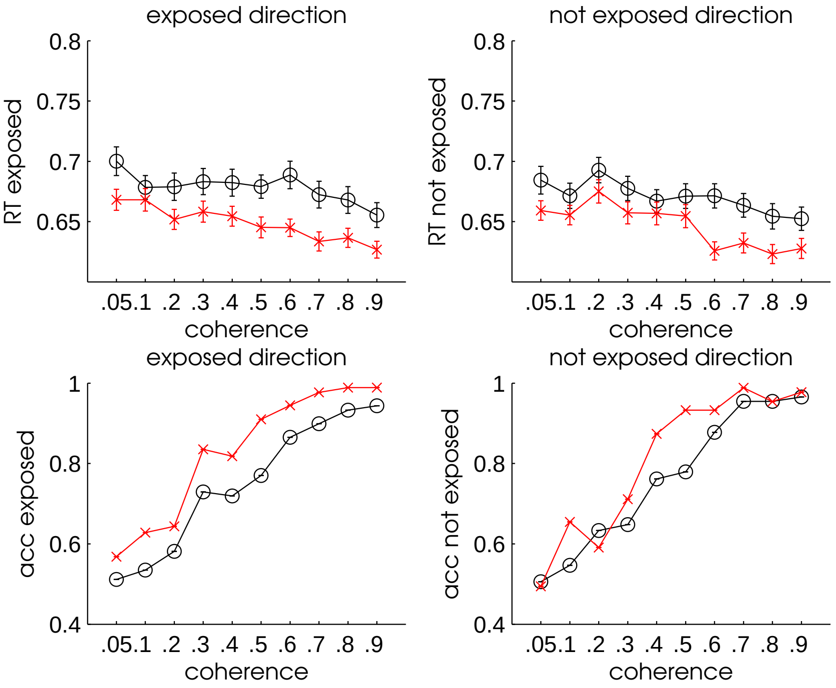


Figure 3.5: (top left) mean RT (± s.e.) for the direction for which participant 4 was exposed, (top right) mean RT (± s.e.) for the direction for which participant 4 was not exposed, (bottom left) mean accuracy for the direction for which participant 4 was exposed and (bottom right) mean accuracy for the direction for which participant 4 was not exposed.

3.3.2 Model fitting

For fitting the diffusion model to RT distributions and proportion of correct and incorrect responses, we used the Diffusion Model Analysis Toolbox (Vandekerckhove & Tuerlinckx, 2007, 2008) for MATLAB. Among the options provided, we chose to estimate parameters by using as the objective function a multinomial likelihood function, which expresses the likelihood of observing a certain proportion of responses in a given number of RT bins and is maximised in order to find the parameter estimates. We decided to represent the reaction time distributions of responses in terms of the classical .1,.3,.5,.7 and .9 quantiles that divide the RT distribution.

For each participant we fitted a model in which the drift rates were free to vary across all conditions while all other parameters were fixed across conditions within the pre-test and the post-test but could vary between pre-test and post-test. Since participants were presented with trials in random order they could not adjust their boundary separation or their starting point of evidence accumulation before the presentation of each trial, hence the assumption of constant boundary and starting point parameters within each session is reasonable, together with stimulus-independent variability in starting point across trials. We assumed a constant the non-decision time component (i.e., stimulus encoding and motor response) between the two directions since it is unlikely that subject would have higher non-decision time (e.g., pressing a button on the keyboard) for one direction compared to the other.

In our model, the lower boundary represents the threshold for answering '135°'

while the upper boundary represents the response '45°'. This means that if the accumulation of evidence reached the upper boundary, the response '45°' was selected, while if the accumulation of evidence reached the lower boundary boundary, the response '135°' was selected. For both the pre-test and the post-test, the value of the starting point was divided by that of the boundary separation in order to give a normalized estimate between 0 and 1, where 0 means complete bias towards the lower boundary ('135°'), 1 means complete bias towards the upper boundary ('45°') and 0.5 represents the unbiased process being equidistant from the two boundaries.

DMAT also allows to calculate estimates of the parameters' standard errors. For each participant, we performed Wald tests for the difference in parameters between pre-test and post-test using the parameter estimates and their standard errors.

The Wald statistic is calculated as follows:

$$Z = \frac{\hat{\vartheta}_{post} - \hat{\vartheta}_{pre}}{\sigma_{post}}$$

where $\hat{\vartheta}_{post}$ is a parameter estimate of the post-test, $\hat{\vartheta}_{pre}$ is a parameter estimate of the pre-test and σ_{post} is the standard error of the parameter of the post-test. When Wald tests are used to test differences between parameters that are constrained across conditions (i.e., for all conditions in our study except drift rates), the corrected reference distribution proposed by Stram & Lee (1994) must be adopted; in this case the Wald statistic Z^2 has been shown to approximate $.5 \cdot \chi^2{}_0 + .5 \cdot \chi^2{}_1$ rather than a $\chi^2{}_1$ distribution.

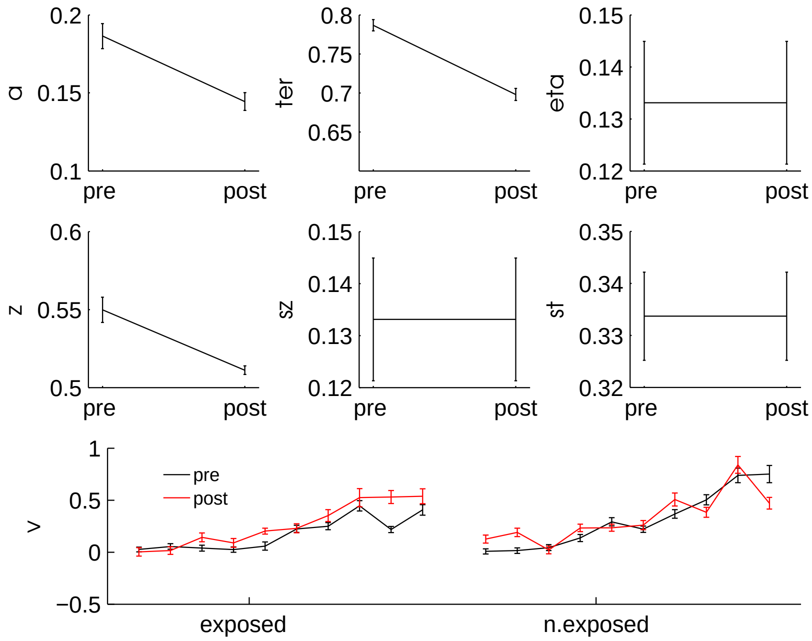


Figure 3.6: DDM parameters for participant 1. In order top left to bottom right: boundary separation, non decision time, across-trials variability in drift rate, starting point, across trials variability in starting point, across trial variability in non-decision time and drift rates. Error bars are standard errors of parameters' estimates.

Also here, results are presented participant by participant. As shown in Figure 3.6, between pre-test and post-test, Participant 1 had a significant decrease in boundary separation (Z = -7.38, p<.001), in non decision time (Z = -11.46, p<.001) and in starting point (Z = -10.36, p<.001). Regarding drift rates, t-tests showed an increase in drift for the exposed direction (t(9) = -2.75, p= .04), while the drift of the not exposed direction did not vary between pre-test and post-test (t(9) = -.42, p= 1). As shown in Figure 3.7, Participant 2 had a significant decrease in boundary separation (Z = -13.14, p<.001), non-decision time (Z = -6.64, p<.001) and starting point (Z = -7.15, p<.001). This subject had an increase in

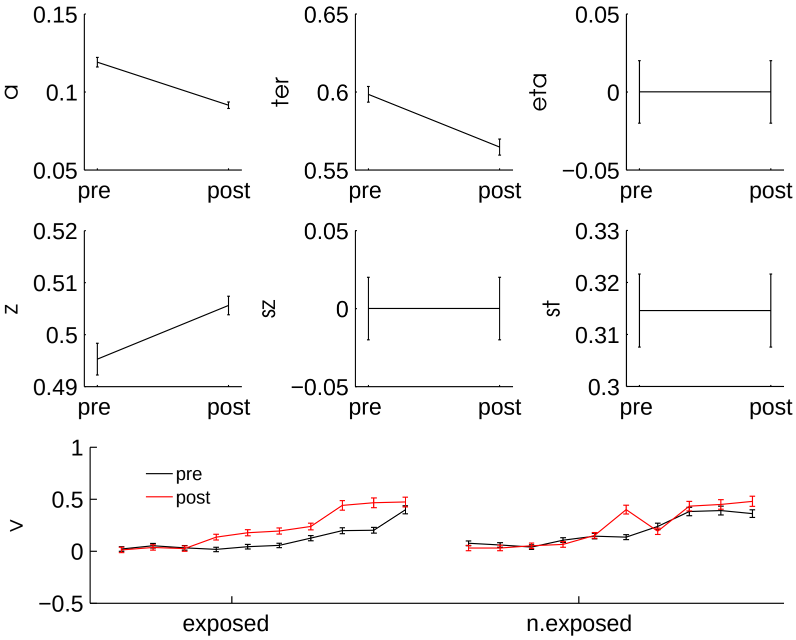


Figure 3.7: DDM parameters for participant 2. In order top left to bottom right: boundary separation, non decision time, across-trials variability in drift rate, starting point, across trials variability in starting point, across trial variability in non-decision time and drift rates. Error bars are standard errors of parameters' estimates.

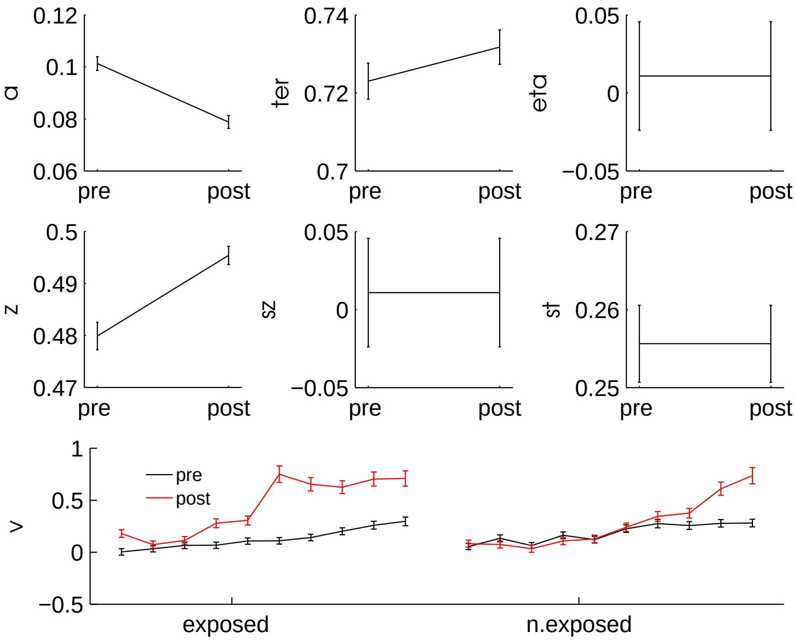


Figure 3.8: DDM parameters for participant 3. In order top left to bottom right: boundary separation, non decision time, across-trials variability in drift rate, starting point, across trials variability in starting point, across trial variability in non-decision time and drift rates. Error bars are standard errors of parameters' estimates.

drift for the exposed direction (t(9) = -3.36, p= .02), while the drift of the not exposed direction did not vary significantly between the pre-test and the post-test (t(9) = -1.15, p= .56). As shown in Figure 3.8, Participant 3 had a significant decrease in boundary separation (Z = -9.18, p<.001), an increase in non-decision time (Z = 1.98, p = .02) and a decrease in starting point (Z = -5.46, p<.001). Between the pre-test and the post-test, the drift of the exposed direction varied significantly (t(9) = -4.83, p= .002) while the drift of the not exposed direction stayed the same (t(9) = -1.61, p= .28). As shown in Figure 3.9, Participant 4 had a significant decrease in boundary (Z = -5.81, p<.001), an increase in non decision time

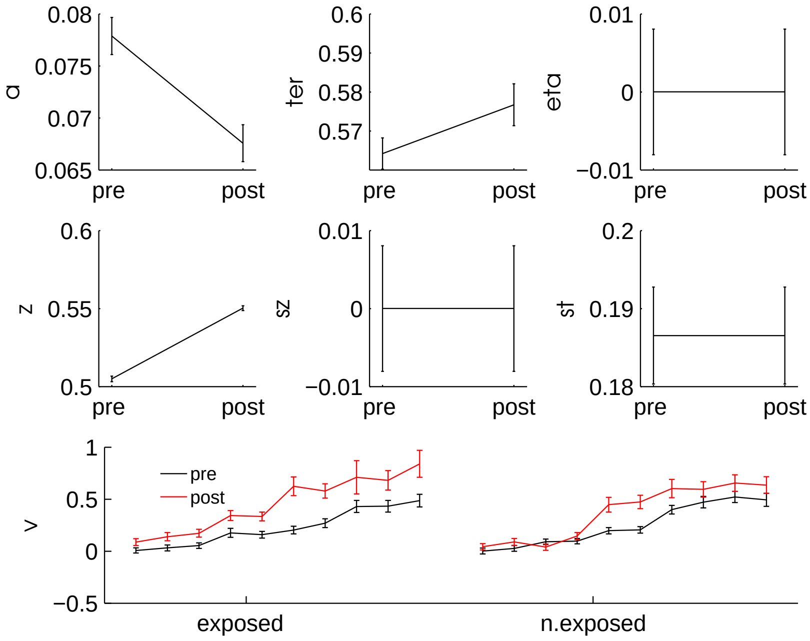


Figure 3.9: DDM parameters for participant 4. In order top left to bottom right: boundary separation, non decision time, across-trials variability in drift rate, starting point, across trials variability in starting point, across trial variability in non-decision time and drift rates. Error bars are standard errors of parameters' estimates.

(Z = 2.33, p=.01) and a decrease in starting point. (Z = -1.37, p=.08). The drift of the exposed direction increased significantly (t(9) = -6.25, p= <.001) as well as the drift of the not exposed direction (t(9) = -3.85, p= .01). Although the drift rates for both directions increased between pre-test and post-test, the relative change between pre-test and post-test for the exposed direction was significantly higher than the relative change between pre-test and post-test for the not exposed direction (t(9) = 3.72, p= .01).

To assess the goodness of fit of the best-fitting model for each participant, fits of the model to the data are represented by quantile probability plots. Figure 3.10 represents quantile probability plots of the data pooled across the four participants against a DDM process with parameters averaged across individuals. On the x-axis of each figure is the probability of a response for the correct and the incorrect boundary. This means that the probability for a correct response is positioned on the right side of the graph (when the probability of response is above chance) while the corresponding probability for an incorrect response is positioned on the left side of the graph at $1 - P(\text{correct response})$. On the y-axis of each plot are reported the conventional .1,.3 .5,.7 and .9 quantiles of the RTs distributions for correct and incorrect responses. For a more detailed explanation and interpretation of quantile probability plots refer to Ratcliff & McKoon (2008). The quantile probability plots show that the model on which our analyses are based fits the data well and without mismatches. Note that some of the error responses for highly discriminable stimuli are affected by few extreme and potentially unreliable measurements, which are less likely to be generated by a

DDM process and to which little weight should be given when assessing the fit.

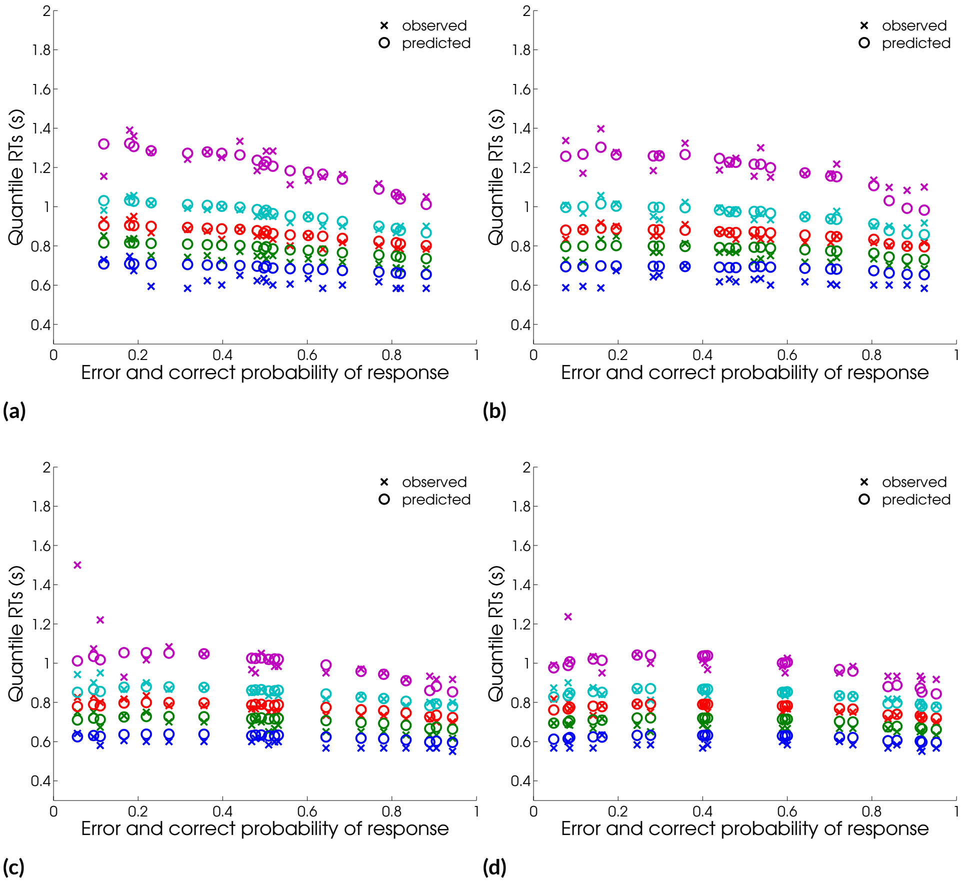


Figure 3.10: Quantile probability plots for the data pooled across individuals and the predictions of the model based on the parameters averaged across individuals. Each plot shows the ten coherence levels for one direction during the pre-test or the post-test. (a) pre-test 45°, (b) pre-test 135°, (c) post-test 45°, (d) post-test 45°. The different colours reflect the different quantiles (.1, .3, .5, .7, .9) of the RT distribution for each condition.

3.4 Discussion

Here, using the DDM (Ratcliff & McKoon, 2008), we have modelled for the first time the processes underlying task-irrelevant perceptual learning in healthy individuals. Participants were exposed to a constant sub-threshold motion direction while performing a RSVP task. Before and after the RSVP task, participants were tested on a pre-test and a post-test motion direction discrimination task. The results indicate that: (i) TIPL affects the drift rate at which participants accumulate evidence for the exposed direction, (ii) TIPL affects the conservativeness of participants' response (iii) non-systematic variations in parameters between the two sessions (e.g., variations in non-decision time, variations in starting point) do not allow a direct comparison of the the decision process only based on accuracy and/or RT.

These findings have important implications for the interpretation of perceptual learning data, both task-relevant and task irrelevant, and, we hope, for the analysis of data collected on different days or for which learning is involved. First, every decision is a mixture of different cognitive processes that can be isolated by this analysis for a more principled interpretation of results. Interpreting learning in terms of latent cognitive variables allows for a more precise investigation of its effect and gives a proper measure of 'true' perceptual learning - change in the drift rate which is related to the quality of input information - while weighting for systematic or random variations in other parameters. In particular, caution should be exercised when comparing data across different sessions. As in previous studies of perceptual learning, participants in our study showed evidence of a change in their speed-accuracy trade-off. Not taking this factor into account can lead to wrong conclusions from data.

In theory, decreased boundary should result in decreasing accuracy for the not exposed direction. In our investigation, participants had to view the stimuli for 500 ms before giving their response; given this constraint, participants had a relatively long time window to make a decision and as a consequence this might have obscured a decrease in accuracy between the pre-test and the post-test that should results from a decreased boundary. For future investigation, we recommend using a shorter presentation of the stimuli (e.g., 200 ms) that is more likely to reveal stronger variations in accuracy and RTs of the not exposed direction between the pre-test and the post-test.

Our results show the risk of directly comparing sessions performed on different days without considering the role of each single parameter. Take for example participant 1: by analysing Figure 3.2 and by testing for differences in accuracy, a researcher may be tempted to conclude that this subject did not have any TIPL since there is not a difference between the accuracy of the first and second session both for the exposed and the not exposed direction. However, the model fitting shows that this participant had higher drift rates for the exposed direction, which is likely to be the signature of TIPL, which is accompanied by a decrease in boundary and variations in the bias towards a response. An increase in drift (faster and more accurate responses) accompanied by a decrease in boundary (faster and less accurate responses) can have as output that accuracy levels stay the same as the increase in drift is combined with a decrease in response conservativeness.

Previous studies have shown that perceptual learning is associated with a decrease in boundary separation (Petrov et al., 2011; Liu & Watanabe, 2012; Dutilh et al., 2009) and we replicated this result also here for TIPL showing consistency across four participants. It has been proposed (Liu & Watanabe, 2012) that this decrease in boundary separation is due to the fact that participants are trying to maximise their reward rate, operatlionalised as the proportion of correct responses divided by the average time between them (Bogacz et al., 2006). In other words, if the quality of information increases (i.e., hence the task becomes 'easier') participants can decrease the time spent for each decision. To the best of our knowledge, this is the first study to report a DDM decomposition of TIPL and the first study to show the systematic parameter variations associated with TIPL.

Regarding other parameters there is not consistency in the literature regarding the effects of learning, and also here we do not observe a clear pattern across participants. For example, regarding the non-decision time component, in previous studies investigating perceptual learning, Petrov et al. (2011) and Dutilh et al. (2009) found a decrease associated with learning, while Liu & Watanabe (2012) found that, although not significant overall, some participants showed an improvement. Here we did not find a consistently decreasing non-decision time component between the two sessions, given that only two out of four participants have a decrease in non-decision time. Our only consistent result is that of decreasing boundary related to learning and an effect on the drift; result that shares some similarities with that by Liu & Watanabe (2012). However, we believe that future work, in which more participants are included, is needed in order to make group inferences regarding non-decision time. Furthermore, it is possible that our training regime was too short to appreciate a decrease in non-decision time for all participants.

Regarding the drift, we show that there is an increase in the drift of the exposed direction, compared to the pre-test, and compared to the drift of the not exposed direction of the post-test when the drift of the not exposed direction increases as well in the post-test. Ideally, we would expect that the drift of the not exposed direction would not differ significantly between pre and post-test. For one participant however the drift of the not exposed direction varies as well; this is unlikely to be an effect of TIPL but rather a random variation in participants' performance that further highlights the importance of a DDM decomposition of learning data.

Although the sample size (N = 4) is low, this is in line with similar studies that have performed a DDM decomposition of learning data (e.g., Dutilh et al., 2009), and it is common practice in perceptual learning research (Lu & Dosher, 2004; Yi et al., 2006; Sigman & Gilbert, 2000; Furmanski & Engel, 2000; Johnson & Leibowitz, 1979; Liu & Weinshall, 2000). Furthermore the consistency in results across participants reassures us about our conclusions.

It is to be mentioned that the training that our participants performed is relatively 'short' if compared with the usual TIPL training of about 20 days, during which TIPL reaches its asymptotic level (Watanabe et al., 2001). To our knowledge, this is also the first study showing that TIPL can occur with only seven days of training, about one third of the training usually performed in TIPL studies. It is sensible to expect that if the days of training increase, the effect on boundary and drift reported here would increase as well and have even stronger consequences on accuracy and RT. Future work, employing more participants and longer training regimes is clearly warranted in order to quantify the rate at which each component is affected by learning, and to quantify the distortion that focusing only on accuracy could lead to.

Overall, a consideration of the different components in decision making shows that the two components which are found to vary systematically all have independent effects on speed and/or accuracy. Whilst increased drift will tend to increase speed and accuracy, decreased boundary separation will tend to decrease both. For these reasons, a decomposition of decision making from these observed variables allows us not only to focus on the different effects of perceptual learning individually, but allows us a more accurate assessment of the extent of increased stimulus sensitivity in perceptual learning. Our study is the first to show this increased sensitivity in task-irrelevant perceptual learning, and does so demonstrating that the other components of decision making are affected in a similar way to as in task-relevant perceptual learning.

4

A DDM account of the semantic congruity effect

Abstract

In this chapter, similarly to what done before, we bring a known psychological phenomenon within the same framework as many other decision phenomena,

and we show how this allows a fresh perspective on previously more loosely specified theories. The semantic congruity effect is that judgements are facilitated (i) when the direction of the comparison of two items coincides with the relative position of the items along the dimension comparison or (ii) when the relative size of a standard and a target stimulus coincides. For example, people are faster in judging 'which is bigger?' for two large items, than judging 'which is smaller?' for two large items. Also, people are faster in judging a target stimulus as smaller when compared to a small standard, than when compared to a large standard. Here, we use the DDM to explain the time course of a semantic congruity effect. Formal modelling of semantic congruity allows the time course of the decision process to be described, using an established model of decision making. Moreover, although there have been attempts to explain the semantic congruity effect within evidence accumulation models, two possible accounts for the congruity effect have been proposed but their specific predictions have not been compared directly, using a model that could quantitatively account for both; a shift in the starting point of evidence accumulation or a change in the rate at which evidence is accumulated. With our computational investigation we provide evidence for the latter, while controlling for other possible explanations such as a variation in non-decision time or boundary separation, that have not been taken into account in the explanation of this phenomenon.

4.1 INTRODUCTION

When subjects are required to judge two stimuli that differ on a single contrastive polar continuum (e.g., 'big' vs. 'small'), subjects are faster to judge which of the two stimuli is higher on that continuum, when the stimuli are high on that particular dimension, and they are faster to judge which of the two stimuli is lower on that continuum, when the stimuli are low on that particular dimension. Furthermore, when subjects are required to judge whether a target stimulus is bigger or smaller than a standard stimulus, subjects are faster when the relative size of the standard and of the target coincides (see Dehaene, 1989). This result, referred to as the *semantic congruity* effect, has been replicated in perceptual and symbolic judgements across different domains, including surface area (Moyer & Bayer, 1976), line length (Petrusic et al., 1998), brightness (Wallis & Audley, 1964), scalar adjectives of quality (Holyoak & Mah, 1982), the distance between two cities (Holyoak & Mah, 1982) and Arabic numerals (Banks et al., 1976; Holyoak, 1978).

Many theories have been proposed to account for the semantic congruity effect. These theories vary greatly in the level of description of the phenomenon, with some theories being able to account for semantic congruity effects only in the case in which comparative instructions are presented to the subject, but not when subjects have to decide whether a target is bigger or smaller than a standard stimulus. For a detailed and exhaustive review of the models proposed for the explanation of the semantic congruity effect, refer to Petrusic (1992) and Leth-

Steensen & Marley (2000); here we present a brief description of some of the theories that have been proposed for the explanation of this phenomenon.

According to the expectancy effect (Banks & Flora, 1977; Marschark & Paivio, 1979), the direction of the comparison (e.g., is the target stimulus bigger than the standard?) prepares the subject for the range of stimuli that will be presented. This results in a facilitation in case of congruency between the comparison and the stimuli. However, even when the comparative is presented together or after the presentation of the stimuli, the semantic congruity effect can still be observed (Holyoak & Mah, 1982), undermining a basic assumption of this model. Alternatively, the semantic coding model (Banks et al., 1976, 1975) explains the congruity effect by referring to linguistic codes; however, this struggles with the finding that even non-human primates show a semantic congruity effect when comparing magnitudes (Cantlon & Brannon, 2005). A further verbal theory, the frequency explanation (Ryalls et al., 1998), explains the semantic congruity effect by the fact that each comparative is associated with one unique dimension during learning (i.e., subjects learn to use 'bigger' for high magnitude stimuli, and 'smaller' for low magnitude stimuli); yet, this explanation struggles with the result that the expectancy effect is found also when subjects are taught new comparisons with novel comparatives (Chen et al., 2014). A further class of models are reference point models (Holyoak, 1978; Holyoak & Mah, 1982; Marks, 1972; Dehaene, 1989; Chen et al., 2014), according to which, subjects, when making a magnitude judgement, compare the numerical value of the stimulus with reference values stored in memory. Under this view, the subject is assumed to establish a reference point near one of the extreme values encountered in a given context and this results in a facilitation when the stimulus to discriminate is nearer to the reference point. From this perspective, the use of reference points has been suggested to affect the strength of evidence accumulation (see Dehaene, 1989; Chen et al., 2014); meaning, for example, that when the magnitude of the standard stimulus coincides with the magnitude of the target, this results in higher rates of evidence accumulation, compared to when there is not congruency between the relative sizes of the two stimuli. Other authors have explained the semantic congruity effect adopting random walk models (Link & Heath, 1975; Link, 1990; Birnbaum & Jou, 1990; Poltrock, 1989); these studies explain the semantic congruity effect as arising from a starting point adjustment dictated by the instructions. However, as argued in Leth-Steensen & Marley (2000), in tasks in which subjects are presented with symmetric differences (i.e., the same number of bigger and smaller comparisons are presented), it is not clear why subjects should adjust their starting point of evidence accumulation towards one of the two alternatives in selection paradigms. Finally, some evidence-accumulation models and instructional pathway interference accounts have been proposed (Petrusic et al., 2008; Petrusic, 1992; Leth-Steensen & Marley, 2000), according to which the semantic congruity is due to a variation in the rate of evidence accumulation in case of congruency/incongruency between the instructions and the relative size of the stimulus pair.

Comparing such theories directly is out of the scope of this work, since some of them are not framed within the evidence accumulation framework. Here, we bring the semantic congruity effect within the same framework as many other decision phenomena; we use the DDM and show how it can account for the semantic congruity effect, by fitting it to behavioural data from a magnitude comparison experiment conducted with human subjects. Since the semantic congruity effect manifests in changes in decision time, the use of the DDM, which explicitly considers the time course of decision-making, is natural. In contrast, some of the heuristic proposals outlined above lack such formal description of how decisions evolve over time, or when they specify how the decision evolves, they do so by adopting ad-hoc models that only make predictions for the specific task but cannot be generalised to other tasks or domains (e.g., the models proposed in Leth-Steensen & Marley, 2000 or in Petrusic et al., 2008), while a unifying framework such as the DDM overcome the limitations of task-specific models. Furthermore, with a DDM decomposition we can investigate which decision parameters account for the semantic congruity effect. Together with the explanations proposed (i.e., drift rate or starting point) other parameters that have never been taken into account, such as non-decision time or boundary separation, could play a role in the semantic congruity effect. For example, the non-decsion time, which has never been taken into account in the previous literature, could as well contribute to a semantic congruity effect given that the congruency/incongruency between the magnitude of the stimuli (or between the instructions and the relative sizes of the target and standard stimulus) could affect the motor response of the subjects.

Usually, in 2AFC tasks parameters such as the starting point of evidence accumulation or the thresholds are assumed to take time to change and are assumed to be set before the stimulus appears (Bogacz et al., 2006); here, however, we assume that the size of the standard, to which subjects pay attention at first during the trial presentation, is apprehended quickly, and it affects the decision process. When there is congruency between the magnitude of the standard and of the target stimulus, this results in a facilitation, compared to when there is not congruency between the relative sizes of the two stimuli. In the literature similar mechanisms that affects the early stages of a decision are described; for example, Provost & Heathcote (2015) provided a similar explanation for a mental rotation task, and in their computational investigation they found that participant adjust their boundary separation on the basis of a property of the stimulus, rotation angle. Also, it is to be mentioned that typically in the kind of tasks in which the DDM is used, subjects evaluate one single stimulus; in this case a change in decision parameters cannot be contingent on the outcome of the decision. However, in our case we have that one feature of the stimulus, the size of the standard stimulus, to which subjects pay attention at first, can affect the subsequent discrimination of the target.

In our experiment subjects had to decide whether a target stimulus was smaller or bigger than a standard array. A stimulus example is reported in Figure 4.1. Our experiment presents some differences with some semantic congruity tasks in which the direction of the comparison is explicitly given, but it is as well a semantic congruity effect (Dehaene, 1989) and similar experimental paradigms have been used before (e.g., Link, 1990; Dehaene, 1989).

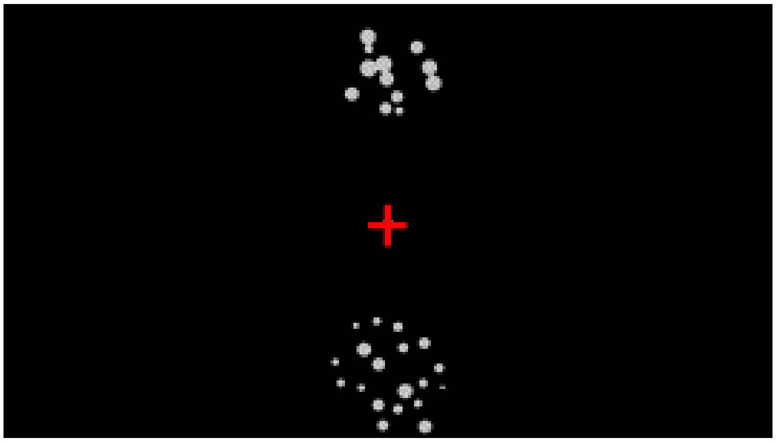


Figure 4.1: Stimulus example. In each trial subjects had to decide whether the array presented on bottom (target) was smaller or bigger in numerosity than the array presented on top (standard). After their response, subjects were presented with a fixation that over the course of 600 ms was varying in size, as a warning signal to maintain fixation at the centre of the screen.

4.2 EXPERIMENT

4.2.1 PARTICIPANTS

Four right-handed subjects, one male, mean age = 20.5 years (SD = 3.2) with normal or corrected-to-normal vision participated voluntarily in the experiment in exchange of credits for course requirements. Each participant was tested in four sixty-minutes sessions on different days. The experiment was approved by the University of Sheffield, Department of Psychology Ethics Sub-Committee, and carried out in accordance with the University and British Psychological Society ethics guidelines and subjects gave their informed consent before performing it.

4.2.2 Materials

The experiments were programmed in Matlab, using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). We used a modification of an established perceptual decision task (Piazza et al., 2010; Revkin et al., 2008b,a; Piazza et al., 2011; Gertner et al., 2012) and a type of 'congruity' task similar to that used by Link (1990) and Dehaene (1989) - similar since also in our case subjects decided whether a target stimulus was bigger or smaller than a standard stimulus, however Link (1990) and Dehaene (1989) used two-digit numbers in their experiments. In our task, participants judged if a cluster of dots presented on the bottom of a laptop screen was 'smaller' or 'bigger' in numerosity than one presented on the top of the screen without counting and responding by button press. On each trial, one array - the standard - contained a fixed numerosity (12 dots for one third of the trials, 24 dots for one third of the trials, 36 dots for the other third), and the other array - the target - contained a varying numerosity that was smaller or bigger than the fixed numerosity by one of seven possible ratios. The ratio defined the difficulty of the judgement, with ratios closer to 1 being harder. The seven ratios, in order of increasing difficulty, were 0.42, 0.50, 0.58, 0.66, 0.77, 0.83, 0.91. The absolute number of dots in each choice pair and a description of conditions is shown in Table 4.1.

There were in total 42 conditions; seven increasing ratios (i.e., increasing difficulty) for each of three levels of standard stimulus magnitude (small, medium and big) for each type of response 'smaller' or 'bigger' (i.e., half of the times the target stimulus was bigger/smaller than the standard). For each trial, subjects had to

Condition	N of Dots	Ratio	Magnitude of Standard	Target (compared to standard) is
1	12 vs 5	0.42	small	smaller
2	12 vs 6	0.5	small	smaller
3	12 vs 7	0.58	small	smaller
4	12 vs 8	0.66	small	smaller
5	12 vs 9	0.75	small	smaller
6	12 vs 10	0.83	small	smaller
7	12 vs 11	0.91	small	smaller
8	12 vs 19	0.42	small	bigger
9	12 vs 18	0.5	small	bigger
10	12 vs 17	0.58	small	bigger
11	12 vs 16	0.66	small	bigger
12	12 vs 15	0.75	small	bigger
13	12 vs 14	0.83	small	bigger
14	12 vs 13	0.91	small	bigger
15	24 vs 10	0.42	medium	smaller
16	24 vs 12	0.5	medium	smaller
17	24 vs 14	0.58	medium	smaller
18	24 vs 16	0.66	medium	smaller
19	24 vs 18	0.75	medium	smaller
20	24 vs 20	0.83	medium	smaller
21	24 vs 22	0.91	medium	smaller
22	24 vs 38	0.42	medium	bigger
23	24 vs 36	0.5	medium	bigger
24	24 vs 34	0.58	medium	bigger
25	24 vs 32	0.66	medium	bigger
26	24 vs 30	0.75	medium	bigger
27	24 vs 28	0.83	medium	bigger
28	24 vs 26	0.91	medium	bigger
29	36 vs 15	0.42	big	smaller
30	36 vs 18	0.5	big	smaller
31	36 vs 21	0.58	big	smaller
32	36 vs 24	0.66	big	smaller
33	36 vs 27	0.75	big	smaller
34	36 vs 30	0.83	big	smaller
35	36 vs 33	0.91	big	smaller
36	36 vs 57	0.42	big	bigger
37	36 vs 54	0.5	big	bigger
38	36 vs 51	0.58	big	bigger
39	36 vs 48	0.66	big	bigger
40	36 vs 45	0.75	big	bigger
41	36 vs 41	0.83	big	bigger
42	36 vs 39	0.91	big	bigger

Table 4.1: Stimuli values for each condition.

decide whether the target stimulus was smaller or bigger than the standard stimulus by pressing 'left' or 'right' on the keyboard. Conditions were chosen so that for each standard stimulus we would have accuracy levels that range from floor to ceiling on the basis of the results of previous pilot studies.

To avoid participants relying upon continuous quantities associated with numerosity (i.e., dot size and envelope area) in this experiment the dot arrays were generated following the method and the MATLAB code provided by Gebuis & Reynvoet (2012). This method was used to produce four sets of images with all possible combinations of correlation (positive vs. negative) between the two features of the stimuli (envelope area, dot size) and dot number.

4.2.3 Procedure

During the whole experiment, subjects had to put their head on a chin rest at a viewing distance of 57 cm from the screen of a 14-inch laptop monitor (Dell Latitude E5430) with a refresh rate of 60 Hz. Subjects were required to fixate a red cross at the centre of the screen. The two dot arrays were presented simultaneously on the screen at ± 4.25 degrees of visual angle from the fixation cross, and participants were asked to judge if the cluster presented on the bottom of the screen was bigger or smaller than the one presented on top by pressing 'left' or 'right' on a keyboard. Each dot was randomly assigned an item size ranging between 0.08 and 0.59 degrees of visual angle. If subjects answered below 300 ms or above 3000 ms the sentence 'Too fast!' or 'Too slow!' was displayed on the screen. After giving a response, subjects were presented with a fixation cross that over the course of 600 ms was varying in size (i.e., small and then bigger for two times), as a warning signal for subjects to pay attention to the centre of the screen, and after subjects were presented with a new trial. Trials were presented in random order across blocks and participants performed 50 trials per condition after a training phase to familiarize them with the task which involved 1 trial per condition. Subjects participated in 4 different sessions on 4 different days (within a week from the first session) for a total of 200 trials per condition and 8400 trials for the whole experiment.

4.3 ANALYSES

4.3.1 BEHAVIOURAL RESULTS

Figure 4.2 shows the psychometric functions averaged across subjects. Figure 4.3 shows mean correct RTs as a function of the experimental condition when data are averaged across participants and RTs lower than 0.3 s and bigger than 3 s are eliminated (about 0.5% of the data). The second column of plots of Figure 4.3 shows mean accuracy averaged across participants. The two plots on the top row show mean RTs and accuracy for conditions for which the standard stimulus was 'small' (i.e., it had 12 dots), the two plots on the middle row show mean RTs and accuracy for conditions for which the standard stimulus was 'medium' (i.e., 24 dots) and the two plots on the bottom row show mean RTs and accuracy for conditions for which the standard stimulus was 'big' (i.e., 36 dots).

Figure 4.3 and Figure 4.2 clearly show the presence of a semantic congruity effect, given that subjects, for conditions having the same ratio (e.g., 12 and 5 dots

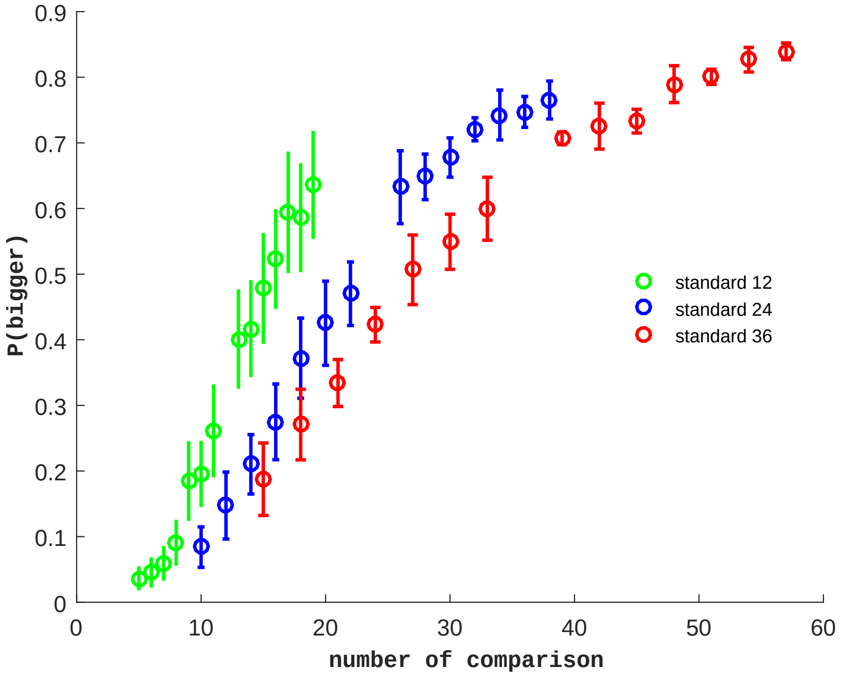


Figure 4.2: Psychometric functions. On the X axis is reported the number of dots of the target stimulus. Error bars are standard errors of the mean.

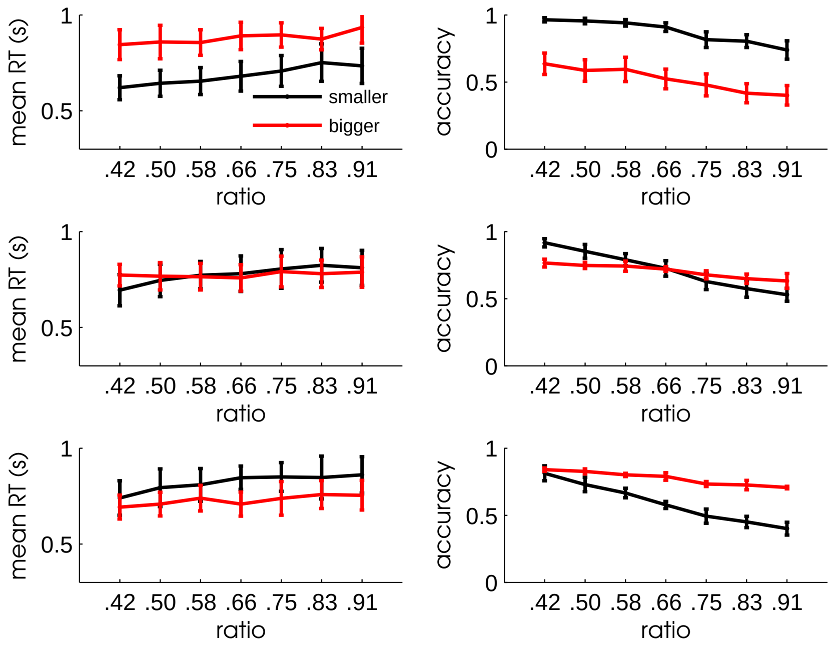


Figure 4.3: Mean corerct RTs, and accuracy levels averaged across subjects for small standard conditions (first row), medium standard conditions (second row) and big standard conditions (third row). Error bars represents standard errors of the mean. The legend shows whether the target stimulus was smaller or bigger than the standard stimulus.

vs 12 and 19 dots), have different RTs and accuracy depending on the congruency between size of the standard and of the target stimulus.

We entered correct RTs and accuracy levels in different mixed-effect regression models for each magnitude level with Ratio as covariate and Correct Response Category (CRC) as factor. In each regression, we included random effects for subject-specific constants and slopes. For each mixed-effect regression, we report the test of fixed effects, and the estimates of fixed effects. Regarding correct RTs, when the standard was small there was an effect of CRC, F(1, 7.58) = 11.94, p = .009, B = -.20 with RT decreasing when the CRC category was 'smaller'. The effect of Ratio, F(1, 13.64) = 19.86, p = .001, B = .20 showed that RT increased as ratio increased. When the magnitude of the standard was medium, there was not an effect of CRC, F(1, 8.77) = .01, p = .92. The effect of Ratio was significant, F(1, 18.03) = 8.99, p = .008, B = .13. When the standard was big, there was not an effect of CRC, F(1,8.61) = 2.88, p = .13, B = .09 although there was an increasing trend for RTs when the CRC was 'smaller' (also shown in Figure 4.3). The effect of Ratio was significant, F(1, 18.21) = 15.47, p=.001, B = .18. Regarding accuracy, when the standard was small there was an effect of CRC, F(1, 7.80) = 12.45, p=.008, B = .36, with accuracy increasing when the CRC category was 'smaller'. The effect of Ratio, F(1, 9.46) = 43.33, $p < .001$, B = -.49 showed that accuracy decreased as ratio increased. When the magnitude of the standard was medium there was not an effect of CRC, F(1, 7.38) = .10, p = .76. The effect of Ratio was significant, F(1, 14.63) = 118.82, $p < .001$, B = -.49. When the standard was big there was an effect of CRC, F(1, 6.30) = 33.11, p=.001, B = -.18, with accuracy decreasing when the CRC category was 'smaller'; also in this case, the effect of Ratio was significant, F(1, 30.61) = 287.36, $p < .001$, B = -.57.

4.3.2 MODEL FITTING

As done for the previous study, for fitting the diffusion model to RT distributions and proportion of correct and incorrect responses, we used the Diffusion Model Analysis Toolbox (Vandekerckhove & Tuerlinckx, 2007, 2008). Here, among the options available, we chose to estimate parameters by using as objective function a chi-square function. We decided to represent the RT distributions of responses in terms of six bins, defined by the boundaries of the .4, .55, .7, 1, 1.5 seconds bins dividing the RT distribution. In the Diffusion Model Analysis Toolbox, the observed response frequencies are compared to the expected response frequencies and a chi-square statistic is minimised to find the best fitting parameters.

For each participant the drift and its variability could be (i) fixed across conditions, or (ii) free to vary across conditions; the boundary separation could be (i) fixed across conditions, or (ii) free to vary across conditions; the starting point and its variability could be (i) fixed across conditions, or (ii) free to vary across conditions; and finally the non-decision time and its between trials variability could be (i) fixed across conditions, or (ii) free to vary across conditions.

All possible combinations of models were fitted to each individual resulting in a total of 16 models per participant. To assess which model best satisfies the trade-off between simplicity and goodness of fit, we used a statistical criterion for model selection, the Bayesian Information Criterion (BIC; Raftery, 1995), calculated as $-2 \cdot log - likelihood(data|model) + k \cdot logN$, where k is the number of free parameters in the model and N the total number of observations. The BIC is a measure of goodness of fit to which a penalty for the introduction of parameters is added. The best model is the model with the lowest BIC value and a difference of ten in BIC scores between two models is considered a strong evidence towards the model with the lowest BIC score. A difference of two BIC scores is the minimum difference to favour a model over another. For all participants, the model in which only the drift rate and its variability were allowed to vary across condi-

		BIC			
parameters free to vary	df	part 1	part 2	part 3	part 4
a	48	36576.14	38024.36	34810.49	35543.56
a eta v	130	36576.23	38260.49	34750.83	34687.46
a z sz	130	36513.37	38444.74	35176.21	34319.85
a eta z sz v	212	37143.49	38928.65	35774.99	34766.12
a ter st	130	37161.02	38628.05	35448.55	36114.33
a ter eta st v	212	37747.68	39298.48	36049.12	36568.87
a ter z sz st	212	37457.55	39141.2	35896.05	36097.81
a ter eta z sz st v	294	38121.66	39577.1	36548.9	36675.68
none	7	36593.38	38040.75	34937.22	35724.95
eta v	89	35754.26	37978.23	34506.52	33565.91
z sz	89	36196.63	38086.65	34610.56	34176.66
eta z sz v	171	36780.5	38604.23	35306.89	34332.28
ter st	89	37226.09	38675.38	35422.08	36123.94
ter eta st v	171	36918.82	38680.12	35062.75	35705.96
ter z sz st	171	36872.35	38844.15	35278.26	34732.73
ter eta z sz st v	253	37074.33	39214.92	35705.16	35427.37

Table 4.2: BIC values for each models and each participant. The first column shows which parameters were allowed to vary in the model (a = boundary, ter = non-decision time, eta = variability in drift, z = starting point, sz = variability in starting point, st = variability in non-decision time, v = drift), the second columns shows the degrees of freedom of each model while the remaining four columns show, for each participant ('part') the BIC value for each model. The best model for each participant is reported in red.

tions, was selected by far as the best model, with differences in BIC scores being always greater than 46 if the best model is compared to the second-best model, showing a very strong preference for this model. BIC scores for each model and participant are reported in Table 4.2.

As it is clear from plotting the drift rate and its across-trials variability recovered from the fitting for each participant - Figure 4.4, Figure 4.5, Figure 4.6 and Figure 4.7 -, the drift rate was (i) a function of the ratio between the standard and the target stimulus (i.e., the difficulty) and (ii) a function of the congruity between the magnitude of the standard and the magnitude of the target stimulus, meaning that in case of congruity between the relative sizes of the standard and of the target, drift rates for 'difficult' conditions are higher in absolute terms compared to when there is not congruity between the relative size of the two stimuli. In each Figure, when drift values are positive, it means that they drifted towards the threshold for the answer 'bigger', while when drift values are negative, it means that the process was directed towards the boundary 'smaller'. Figure 4.7 shows that participant 4, was biased towards answering 'smaller' when the standard stimulus was smaller; even for highly discriminable stimuli.

For each participant, we run separate linear regressions on drift estimates with Ratio as Covariate and Magnitude and CRC as Factors. Our results showed that for all participants the interaction effect of Magnitude and CRC was significant (all $p < .03$). Post-hoc tests with Bonferroni corrections showed that all participants had higher drift rates when the magnitude of the standard was small and the correct response category was 'smaller' compared to when it was 'bigger' (all $p < .001$) and had higher drift rates when the magnitude of the standard was big and the correct response was 'bigger' compared to when it was 'smaller' (all $p < .017$).

Eta did not vary consistently across conditions; this was also confirmed by the fact that no linear or quadratic fitting for *eta* resulted significant (all $p > .087$).

The remaining parameters, for each participant, are shown in Table 4.3.

Fits of the model to the data are represented by quantile probability plots, Figure 4.8, showing that the model in which drift and variabilities are free to vary

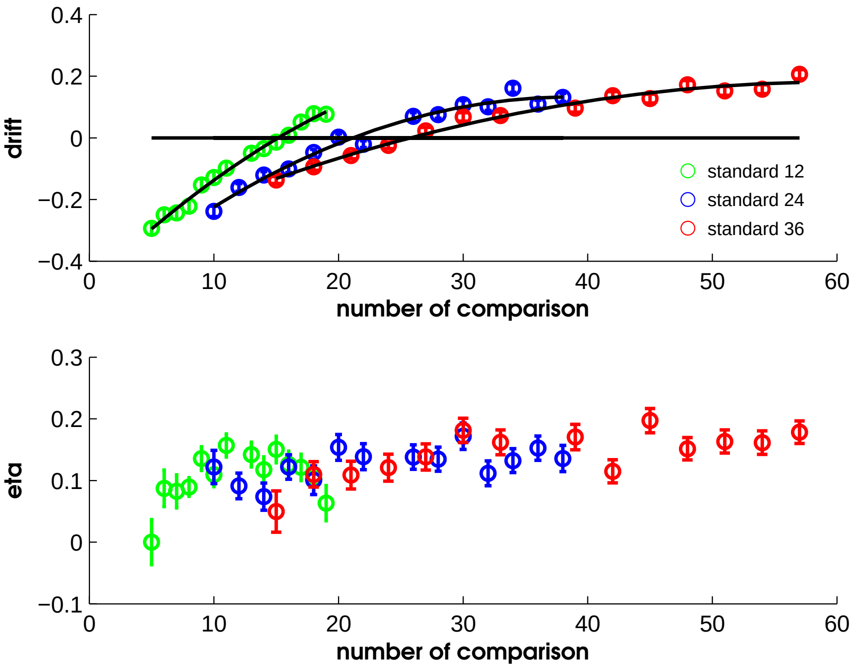


Figure 4.4: Graphical representation of (top) drift rate, (bottom) variability in drift rate across conditions for the first subject. For drift rate we report the best quadratic fitting as a function of the target stimulus. The horizontal line for the drift represents the level at which the drift is 0. Error bars are standard errors of parameters' estimates.

	a	ter	z	sz	st
participant 1	.145	.290	.073	.001	<.001
participant 2	.176	.274	.091	<.001	.084
participant 3	.141	.496	.073	<.001	.446
participant 4	.165	.272	.076	<.001	.039

Table 4.3: Values of a, ter, z, sz and st for each participant.

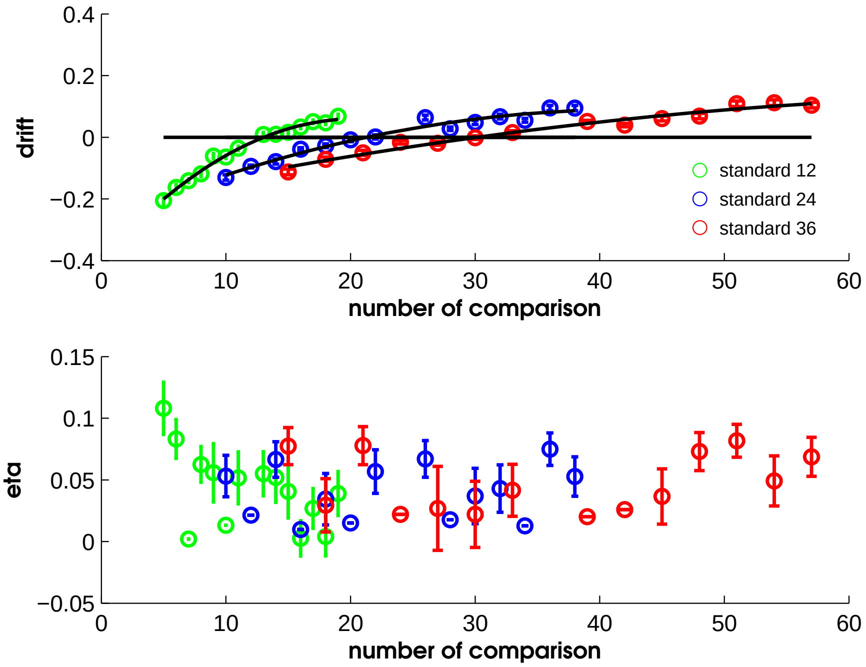


Figure 4.5: Graphical representation of (tom) drift rate, (bottom) variability in drift rate across conditions for the second subject. For drift rate we report the best quadratic fitting as a function of the target stimulus. The horizontal line for the drift represents the level at which the drift is 0. Error bars are standard errors of parameters' estimates.

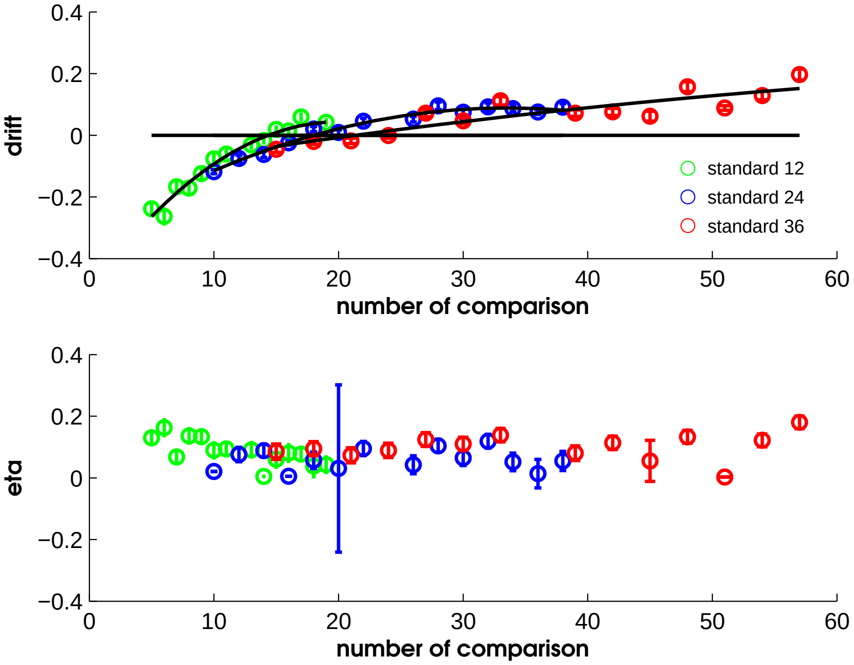


Figure 4.6: Graphical representation of (top) drift rate, (bottom) variability in drift rate across conditions for the third subject. For drift rate we report the best quadratic fitting as a function of the target stimulus. The horizontal line for the drift represents the level at which the drift is 0. Error bars are standard errors of parameters' estimates.

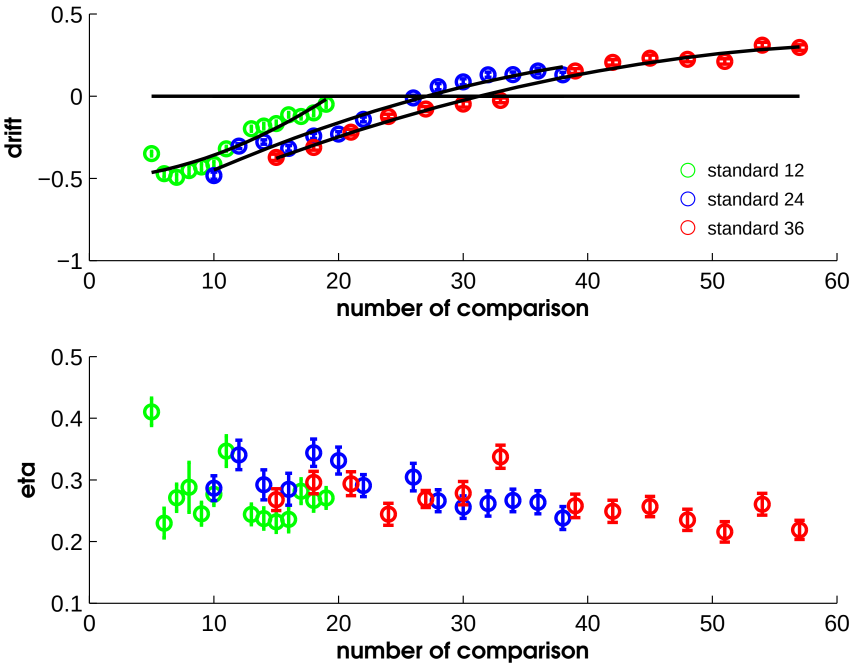
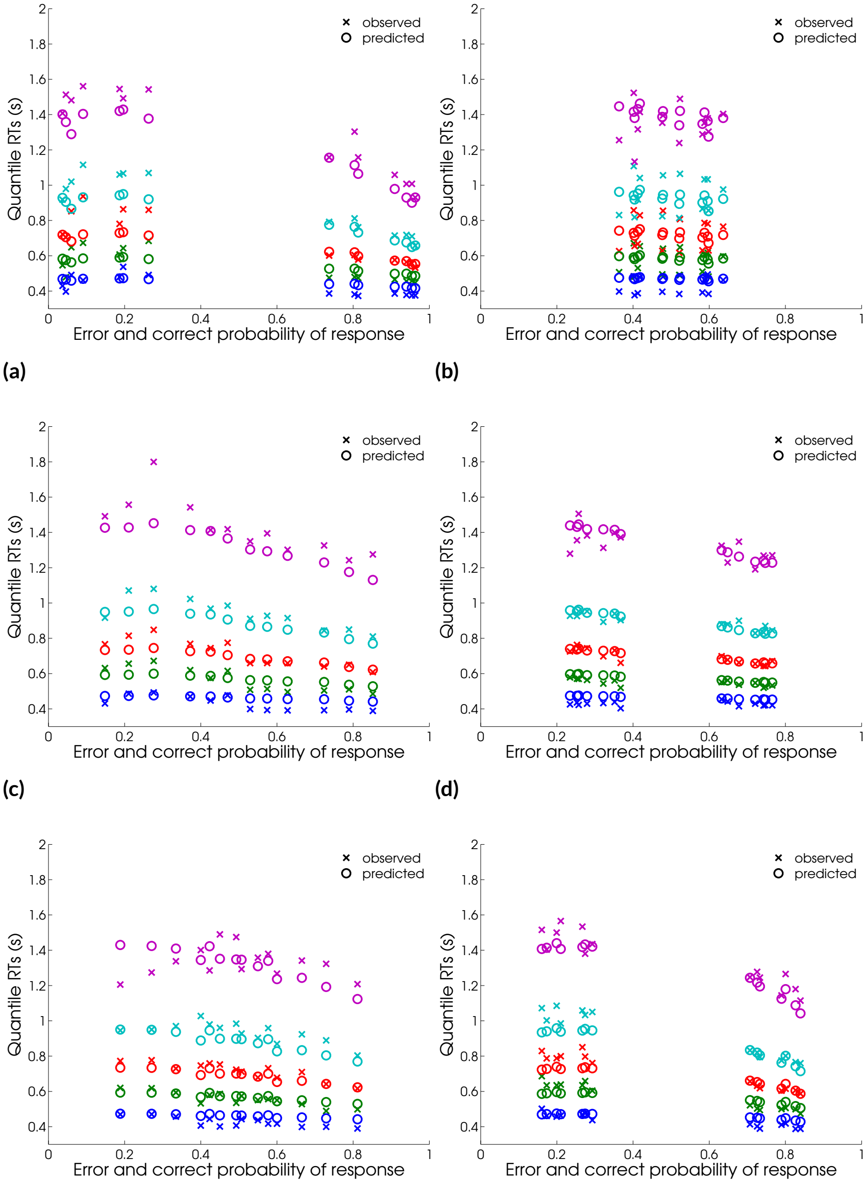


Figure 4.7: Graphical representation of (top) drift rate, (bottom) variability in drift rate across conditions for the fourth subject. For drift rate we report the best quadratic fitting as a function of the target stimulus. The horizontal line for the drift represents the level at which the drift is 0. Error bars are standard errors of parameters' estimates.

across conditions provides a good description of the data. Similarly to the previous chapter, here we compare the predictions of the model based on the parameters averaged across individuals, and the observed data pooled across individuals. Figure 4.8 has 6 plots; the two plots on top show conditions for which the standard was small, the plots on the middle show conditions for which the standard was medium and the plots on the bottom show conditions for which the standard was big. The plots on the left of Figure 4.8 show conditions for which the correct response category was 'smaller', while the plots on the right show conditions for which the correct response category was 'bigger'. Note that, as the behavioural analyses show, for conditions with a high ratio (i.e., high difficulty), the overall performance of subjects dropped below chance in some cases. As a consequence, for these conditions, the probability of a correct choice lays on the left of the graph, and the probability of an incorrect choice is on the right side of the graph, mostly near to chance level. In general, for conditions with highly discriminable stimuli (i.e., conditions with low ratio) little weight should be accorded to the quantiles for error responses since these are mainly influenced by a very limited and potentially unreliable number of measurements given that subjects made very few errors in these extreme conditions. Furthermore, DMAT ignores conditions that have less than eleven errors, hence these conditions are not reported.

The quantile probability plots show that the model obtained from our fitting can capture the averaged data fairly well, especially considering that the data are averaged across four experimental sessions with clear repercussions on the within-



(a) (b) (c) (d) (e) (f)

Figure 4.8: Quantile probability plots showing predictions of the model (recovered from the parameters averaged across individuals) and the data (averaged across individuals).

subject variability, and considering the high number of conditions present in this study.

4.4 Discussion

As we have discussed above, several verbal descriptive theories have been proposed for the explanation of the semantic congruity effect (Banks & Flora, 1977; Marschark & Paivio, 1979; Holyoak & Mah, 1982; Banks et al., 1976, 1975; Ryalls et al., 1998). Here, we have adopted a computational framework, the drift diffusion model (DDM) that is psychologically plausible, mathematically rigorous and that has been shown to fit data in various psychological tasks (Ratcliff, 2002; Thapar et al., 2003; Voss et al., 2004; Ratcliff, 1978; Ratcliff & McKoon, 1988; Ratcliff & Rouder, 1998; Ratcliff et al., 1999, 2004b). Our results show that the DDM (Ratcliff, 1978; Ratcliff et al., 1999; Ratcliff & McKoon, 2008) can account for the data in an experiment in which we have elicited a semantic congruity effect.

We found that the changes in decision time and accuracy associated with the introduction of our manipulation, can be best explained by a change in the drift rate. The drift rate is associated with the discriminability of the experimental condition, as it is commonly assumed in the DDM, but it is also affected by the congruity between the magnitude of the standard stimulus and the magnitude of the target. This effect seems to suggest that subjects were first assessing the numerosity of the standard and then they were assessing the numerosity of the target to decide if it was smaller or bigger compared to the standard; when there was congruency between the relative sizes of the standard and of the target, this resulted in higher drift rates. Specifically, in this study subjects may have learnt to use the two extreme standard magnitudes as reference points for the values 'small' and 'big', since over the four experimental session the numerosity of the standard only consisted of three possible values. This strategy would result in the pattern observed in the data with subjects being faster and more accurate in judging which of the two stimuli is bigger/smaller when there was congruency between the magnitude of the target and of the standard. In our study, subjects had to asses the size of the standard stimulus. Afterwards, subjects had to assess whether the target stimulus was bigger or smaller than the standard and in case of congruency between the sizes two stimuli, the response was faster and more accurate (i.e., drift rates were higher).

The main result of this study is in line with reference point models (see Dehaene, 1989) and with theories in which the congruency between magnitude of the stimulus and the response category affects the strength of the evidence signal (Petrusic et al., 2008; Leth-Steensen & Marley, 2000), while we invalidate theories which interpret the semantic congruity effect as a modification in starting point of evidence accumulation (Link & Heath, 1975; Link, 1990; Birnbaum & Jou, 1990; Poltrock, 1989). However, a key point of the models proposed by Leth-Steensen & Marley (2000) and by Petrusic et al. (2008) is that the semantic congruity effect arises when there is congruency between the comparison instruction and the relative size of the stimuli, while in our case the semantic congruity arises as congruency between the size of the standard and the size of the target stimulus. Further theoretical work - in which such theories are framed within a DDM framework - and experimental work - in which the direction of the comparison is explicitly given - is needed to test Leth-Steensen & Marley (2000) and Petrusic et al. (2008) explanations, given that the experimental paradigm presented here and the conceptual explanation that we provided vary greatly from their conceptualisation of the same phenomenon. Furthermore, for these theories, it has been proposed (Leth-Steensen et al., 2014) that it is the relative size of the stimulus pair that 'primes' the corresponding congruent form of the instruction, resulting in a facilitation in case of congruency. However, it is not clear why an assessment of the relative size of the stimulus pair is even necessary when not explicitly required, as in our case. Also, the result that semantic congruity effects arise even when the standard stimulus and the target stimulus are presented sequentially (Link, 1990; Dehaene, 1989), seems to undermine the role of the size of the stimulus pair in the explanation of semantic congruity effect.

The other principal theories that have been proposed for the explanation of the semantic congruity effect - the expectancy effect (Banks & Flora, 1977; Marschark & Paivio, 1979), the semantic coding model (Banks et al., 1976, 1975) and the frequency explanation (Ryalls et al., 1998) - seem to be already partially falsified by the contrastive results presented in the introduction. Furthermore, the expectancy theory and the semantic coding model do not apply in our study, given that they are dependent on the direction of the comparative instruction that is not used in the current task.

The choice of previous authors (Link & Heath, 1975; Link, 1990; Birnbaum &

Jou, 1990; Poltrock, 1989) of allowing variations in parameters such as the starting point of evidence accumulation, while neglecting the role of other possible mechanisms (i.e., variations in boundary separation), is unclear. Here, we show directly - with the model selection procedure - that neglected mechanisms, such as boundary separation or non-decision time variations, do not play a role in the semantic congruity effect.

Our application of the DDM further highlights the heuristic power of the DDM, and shows that different phenomena that have been previously explained by descriptive and or task-specific theories can be accounted for by sequential sampling models of evidence accumulation and decision making, when the focus is shifted to the computational level of analysis. Our formal account of this phenomenon is parsimonious, as it uses a unifying model of choice rather than proposing an ad-hoc model for the explanation of the phenomenon, and rigorous, as we account for the full distributions of correct and error responses, by taking into consideration all the cognitive processes that underlie a decision.

Part II

Decision modeling insights in adaptive decision making

5

When natural selection should optimize speed-accuracy trade-offs.

Abstract

This chapter is from the paper Pirrone, A., Stafford, T., & Marshall, J. A. (2014). When natural selection should optimize speed-accuracy trade-offs. Frontiers in

neuroscience, 8, 73. Here we argue that although decision making is usually studied from the speed-accuracy trade-off perspective, many decisions are instead described by a different trade-off, that we define as speed-value trade-off and that cannot be accounted for by computational models of choice that integrate difference in evidence supporting two alternatives.

5.1 Introduction

In the previous three chapters we have used the DDM to generate insight into three different domains; here, and in the following chapter we focus instead on which are the limits of the DDM and under which circumstances the DDM ceases to be optimal.

In psychology and neuroscience, and in other disciplines studying decision-making mechanisms, it is often assumed that optimal decision-making means statistical optimality. This is attractive because statistically optimal decision procedures are known, can be simply implemented in biologically-plausible models, and because such models have been shown to give good fits to behavioural as well as neural data. Here we question when statistical optimality is the kind of optimality we should expect natural selection to aim towards, by considering what kinds of loss function should be optimised under different behavioural scenarios. In laboratory settings subjects are often rewarded only on making a correct choice, so optimisation of a zero-one loss function is appropriate, and this is achieved by implementing a statistically-optimal decision procedure that gives the best compromise between speed and accuracy of decision-making. Many naturalistic decisions may also be described by such a loss function; however others, such as selecting food items of potentially different value, appear to be different since the animal is rewarded by the value of the item it chooses regardless of whether it was the best available. We argue that most naturalistic decisions are value-based. Mechanisms that optimise speed-accuracy trade-offs need to be parameterised, using information about the decision problem, in order to deal with value-based decision-making. Mechanisms for value-sensitive decision-making have been described, however, which adaptively change between decision-making strategies without the need for continual re-parameterisation.

5.2 Speed-Accuracy Trade-Offs

It is usually assumed that decision-makers have to decide to be either fast or accurate. When speed is important mistakes are more frequent, while when accuracy is needed decisions are slower. This obvious problem is defined as the speed-accuracy trade-off and is a distinctive feature of many types of decision making (Wickelgren, 1977).

The speed-accuracy trade-off can be explained within the theoretical framework of sequential sampling models of decision making that have been shown to fit behavioural and neural data from human and animal choice tasks (Ratcliff et al., 2004a; Ratcliff & Rouder, 2000; Ratcliff & Smith, 2004; Ratcliff et al., 2003; Busemeyer et al., 2013). In particular, the Drift Diffusion Model (DDM; Ratcliff, 1978) describes choice between two alternatives (*see* Bogacz et al., 2006; Smith & Ratcliff, 2004; Basten et al., 2010) and recently has been shown also to be quantitatively accurate in describing trinary choices (Krajbich & Rangel, 2011) and value-based choices (Krajbich et al., 2010; Krajbich & Rangel, 2011; Milosavljevic et al., 2010; Krajbich et al., 2012), suggesting that the DDM can be thought of as a unifying computational framework for describing decision making (Basten et al., 2010). Moreover, Bogacz et al. (2006) have demonstrated that several connectionist decision-making models can approximate the DDM under specific conditions. The DDM is a special case of the statistically-optimal Sequential Probability Ratio Test (SPRT; Wald, 1947; Wald & Wolfowitz, 1948). In the DDM noisy sensory evidence supporting the alternatives is integrated over time until the net evidence in favour of one alternative exceeds a certain positive or negative threshold value, precipitating a decision for the corresponding alternative. These thresholds can be varied to compromise optimally between the average speed and accuracy of decisions.

5.3 Speed-Value Trade-Offs

In situations where decisions are rewarded according to whether they are correct or not, optimising the speed-accuracy trade-off is sensible. When decisions are rewarded according to the value of the option chosen, however, a different criterion needs to be optimised. This can be illustrated with the simplest case of choosing between two equal value options; here there is no decision accuracy, since choosing either option is ‘correct’. Similarly, there is no difference in average evidence for which of the two options is more valuable, meaning that the SPRT/DDM will only reach a decision by integrating sufficient noise to cross a decision threshold. Thus in this scenario there is no speed-accuracy trade-off to manage; the optimal decision is to choose anything as quickly as possible. The fundamental insight is that for certain decisions, speed-value trade-offs are more appropriate to optimise, rather than speed-accuracy trade-offs.

The SPRT/DDM can be optimised to take account of the value of the alternatives but, as we discuss here, doing so requires knowledge of the decision problem faced. The thresholds for an optimal decision depend on the goals of the decision maker and are task specific. By way of example, one route to accounting for the values associated with different decision outcomes is to minimise an extended version of the Bayes Risk (BR). BR is a linear combination of expected decision delay and expected terminal decision loss, first proposed by Wald and Wolfowitz (Wald & Wolfowitz, 1948), and assumes that decision makers seek to minimise a cost function that is the weighted sum of decision times (DTs) and error rate (ERs). This was subsequently extended by Edwards to also account for non-zero rewards for incorrect decisions (Edwards, 1965; Bogacz et al., 2006). Formally Edwards' extension of BR, which implements Wald and Wolfowitz's version as a special case, can be defined as

$$BR_E = c_1 DT + c_2 \begin{pmatrix} ER \\ 1 - ER \end{pmatrix} \tag{5.1}$$

where c_1 is the cost of observing the stimulus per unit time, while c_2 is a row-vector specifying the payoffs from incorrect and correct choices (Bogacz et al., 2006). If $c_2 = (k\ 0)$, where $k > 0$ is a constant, then Wald and Wolfowitz's original BR is recovered. Several studies demonstrate that, under specific circumstances, subjects choose decision thresholds close to those that minimise BR_E (Busemeyer & Rapoport, 1988; Mozer et al., 2001). Bayes risk is not the only criterion proposed to date that decision-makers might optimise. Bogacz *et al.* survey alternatives, such as reward-rate, however these alternatives are all calculated based on decision-accuracy, which requires explicit parameterisation based on the values of correct and incorrect choices (Bogacz et al., 2006). We therefore concentrate our analysis on Bayes risk. Bayes risk can be used to optimise value-sensitive decision-making; for example in a decision between two equal alternatives, each having value v if chosen, we would set the vector $c_2 = (v\ v)$ (*e.g.* dashed green line in Figure 5.1), thus simplifying equation 1 above to

$$BR_E = c_1 DT + v. \tag{5.2}$$

Equation 2 shows us that, intuitively, an optimal decision-maker in our equal-alternatives scenario should minimise decision-time DT, since doing so incurs no penalty as the error rate ER no longer features. However, using Bayes risk in this way requires the values of the alternatives to be known on a case by case basis, as shown in Figure 5.1. Subjects might learn the values of incorrect and correct choices over time, for example when trials are blocked in psychophysical experiments (*see* Bogacz et al., 2006). However in the following we argue that in most naturalistic decision scenarios decision-makers will not have this opportunity, and will therefore use other mechanisms that directly optimise speed-value trade-offs, rather than optimising decisions indirectly via optimisation of

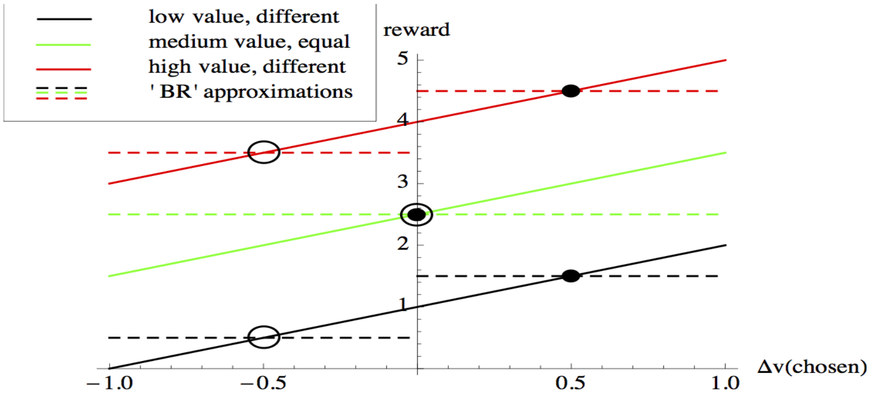


Figure 5.1: The accuracy-based component of Bayes Risk (BR_E as defined by equation 1) can be used to approximate a value-based reward scheme. In value-based decisions individuals are rewarded according to the value $|v| + \Delta v$ of the option they choose (solid lines), where $|v|$ is the average value of the alternatives under consideration, and Δv is the deviation from this average of the value of the option chosen by the subject. With knowledge of the values of the alternatives, BR_E can be used to optimise value sensitive decision-making as described in the main text; for example the dashed lines show payoffs used in BR_E for: options having values of 0.5 and 1.5 units (black), options having equal values of 2.5 and 2.5 units (green) and options having values of 3.5 and 4.5 units (red). Intersections between payoffs selected for BR_E (dashed lines) with value-based reward (solid lines of matching colours) correspond to choice scenarios between different-valued options for which BR_E implements reward-by-value of the selected option; these intersections represent choice scenarios involving 'poor' (hollow circles) and 'good' (filled circles) options having particular values. However, the cost parameters for BR_E need to be recalculated according to the values of the options under consideration; for example, although the difference in the values of the alternatives does not change from the low-value (black) to the high-value (red) scenarios, since their absolute values change the BR_E payoffs need to be recalculated in each case. As described in the text, value-sensitive decision-mechanisms have been described that are able adaptively to deal with a variety of such decision scenarios, without re-parameterisation.

the speed-accuracy trade-off with an appropriate payoff matrix c_2.

5.4 NATURALISTIC DECISIONS ARE USUALLY VALUE-BASED

We argue that most naturalistic decisions faced by animals, including humans, are value-based, in that the animal is rewarded according to the value of the option it chooses. Such a view on decision-making is not new to behavioural ecologists, where a long tradition exists of studying behaviours such as mate choice and foraging (Davies et al., 2012) or nest-site selection (Stroeymeyt et al., 2014). Recently many studies have focused on how value and reward are represented and integrated during the decision process (Platt & Glimcher, 1999; Sugrue et al., 2004; Padoa-Schioppa & Assad, 2006; Rangel et al., 2008; Kable & Glimcher, 2009; Krajbich et al., 2010; Philiastides et al., 2010; Hare et al., 2011; Krajbich & Rangel, 2011; Louie & Glimcher, 2012; Tsetsos et al., 2012; Cassey et al., 2013; Towal et al., 2013); however, in psychology and neuroscience, experiments are usually designed such that there is always a correct choice, and only correct choices are rewarded (*see* Gold & Shadlen, 2003; Bogacz et al., 2006). While studying behaviour in psychophysical tasks is beneficial in that it gives a well-controlled decision environment, our point is that only rewarding subjects when they make correct choices may not correspond to the kind of decisions animals, and their neural circuitry, have typically evolved to deal with. Even in the value-based decision experiments cited above, which are analysed using the DDM, it is typical to only present subjects with a choice between options known to have *different* values. Moreover, even though some studies have looked at how reward information is integrated (Rorie et al., 2010; Gao et al., 2011), much of this work has not yet focused on the trade-off between value and speed. While usually in the decision-making literature the optimal behaviour is to optimise speed-accuracy trade-offs, and subjects can apparently do this (Busemeyer & Rapoport, 1988; Bogacz et al., 2006), we argue that these scenarios are not representative of many naturalistic settings, and that there is great value in considering how subjects make value-sensitive decisions and how these should be optimised. In the following section we discuss theory that may be useful for this.

At least one important class of naturalistic decisions does require optimisation of speed-accuracy trade-offs; these are life-or-death decisions. If we analyse for example the case of an animal attempting to forage while avoiding predators (Trimmer et al., 2008), a slow-but-accurate decision would mean being killed by the predator, a maximal loss. On the other hand if the decision is fast-but-inaccurate the animal would escape even when the stimulus is not a predator, and this would mean losing food. The best strategy for the animal is thus that which optimises the speed-accuracy trade-off, taking into account the payoffs arising from the different decision outcomes; hence Trimmer *et al.*'s hypothetical animal is modelled with a single-threshold DDM, with evidence sufficient to cross that single decision threshold leading to the animal taking anti-predator action such as running away.

Recent modelling work inspired by studying another value-sensitive decision-making system, collective nest-site selection by honeybees (Seeley et al., 2012), has described a very simple mechanism able to adaptively account for the value of different decision outcomes, with minimal parameter tuning (Pais et al., 2013). This simple model implements a variety of sophisticated decision-making strategies; for example, when equal but low-value alternatives are presented, a decision deadlock is maintained that can be broken should a third, higher-value alternative, be made available. However if equal-but-high-value alternatives are presented, or sufficient time passes, deadlock is spontaneously and randomly broken (Pais et al., 2013). This is the rationale behind the speed-value trade-off: as the overall value of the alternatives increases, reaction times decreases. Conversely, as the overall value of the alternatives decreases, reaction times increases.

This is particularly interesting, since the classic DDM is insensitive to the absolute value of the alternatives under consideration, and only integrates the difference in their values. When differences between alternative values are sufficient, the value-sensitive mechanism of Pais *et al.* becomes closer to a classic DDM, allowing speed-accuracy trade-offs to be managed, although not optimised, through modification of decision thresholds. All of the different behavioural regimes of the model arise without direct parameterisation regarding alternatives' values, simply through the dependence of the model's dynamics on the mean values of inputs to its integrator populations; this allows the model to adaptively respond to different decision scenarios on a trial-by-trial basis, which cannot be achieved in pure DDM models without the decision-maker having access to explicit information on the decision-task at hand. Modifications to DDM-type models have been proposed to deal with trial-by-trial variability such as online estimation of task parameters (Deneve, 2012) or the use of time-dependent change in parameters such as decision-thresholds, urgency signals or asymmetry of inhibition (Ditterich, 2006; Hanks et al., 2011; Drugowitsch et al., 2012; Thura et al., 2012); fundamentally however these modifications are still interpreted under the assumption that decision speed vs accuracy is the trade-off to be maximised, unlike the model of Pais et al. (2013) in which the dynamics are naturally interpreted in terms of value vs time trade-offs. Pais *et al.*'s mechanism also exhibits other characteristics of natural value-discrimination systems, such as Weber's law of just-noticeable difference; interestingly Weber's law arises from the deterministic dynamics of the mechanism rather than from noise processes (Pais et al., 2013) (*cf.* Deco & Rolls, 2006; Deco et al., 2007). However, as it will be discussed in more detail in the following chapter, many other sequential sampling models (e.g., Usher & McClelland, 2001; Teodorescu et al., 2015; Drugowitsch et al., 2012; Kacelnik et al., 2011) can account for magnitude sensitivity together with the models presented in Pais et al. (2013).

5.6 Conclusion

The study of speed-accuracy trade-offs has been tremendously fruitful for psychology, neuroscience and animal behaviour, and will doubtless prove fruitful for many years to come. Yet as we have argued here most naturalistic decisions, which animals' brains should have evolved to optimise, are value-based rather than accuracy-based. This leads us to argue that the drift-diffusion model, which optimises speed-accuracy trade-offs, is not an ideal computational framework to describe value-based decision-making; although it has had some success in describing particular experiments on value-based decision-making, discussed in the section 'Speed-Accuracy Trade-Offs', as we have shown here the DDM requires special case-by-case parameterisation to implement true value-based decision-making. We suggest that this limits the generality of the DDM as a unifying framework for all ecologically-relevant decision-making problems. However recent theory has presented mechanisms that can manage value-sensitive decision problems without the additional informational requirements of the DDM. At the same time, experimental and theoretical psychologists and neuroscientists have started to tackle problems of value-based decision-making. We have presented our arguments for value in terms of animal decision-making, but unicellular organisms and individual cells also make decisions (*e.g.* Perkins & Swain, 2009; Latty & Beekman, 2011), and value is likely to be similarly important for these. We believe that the evolutionary perspective we have presented here should motivate further research into value-sensitivity and decision-making.

6

Evidence for the speed-value trade-off

ABSTRACT

In this chapter we test a prediction of the theoretical arguments presented in the previous chapter, magnitude sensitivity.

Complex natural systems from brains to bee swarms have evolved to make adaptive multifactorial decisions. Recent theoretical and empirical work suggests

that many evolved systems may take advantage of common motifs across multiple domains. We are particularly interested in value sensitivity as a mechanism to resolve deadlocks adaptively. This mechanism favours long-term reward maximization over accuracy in a simple manner, because it avoids costly delays associated with ambivalence between similar options; speed-value trade-offs have been proposed to be evolutionarily advantageous for many kinds of decision. A key prediction of the value-sensitivity hypothesis is that choices between equally-valued options will proceed faster when the options have a high value than when they have a low value. However, value-sensitivity is not part of idealised choice models such as the DDM. Here we examine two different choice behaviours in two different species, perceptual decisions in humans and economic choices in rhesus monkeys, to test this hypothesis. We observe the same value sensitivity in both human perceptual decisions and monkey value-based decisions. These results endorse the idea that neural decision systems make use of the same basic principle of value-sensitivity in order to resolve costly deadlocks and thus improve long-term reward intake.

6.1 Introduction

Adaptive decision-making is a hallmark of intelligent complex systems at all levels of biological complexity. Such systems can monitor inputs and then calculate effective responses to them with impressive efficiency and flexibility. A major goal is the elucidation of the basic computational principles underlying mechanisms for decision making, from perceptual decision making, to value-based decision making to social decisions (Krajbich et al., 2015).

Decision mechanisms are often studied from the perspective of the speed-accuracy trade-off. That is, the decision-maker is assumed to optimize choices based on two competing cost functions, the cost of inaccurate choices and the cost of delays imposed by longer deliberations. This trade-off function has been a central aspect of models of decision-making in psychology, neuroscience, and behavioural ecology (*e.g.* Bogacz et al., 2006; Chittka et al., 2009). However for many decisions, such as food choice, decision-makers should optimize value, not accuracy, and decision-making processes should take this fact into account (Pirrone et al., 2014; Teodorescu et al., 2015). Both the cost of a decision - in time taken and risk of error - and benefit of a decision - in reward - may frequently depend on the value of options. When referring to 'overall value' we mean the magnitude or intensity of the stimuli or reward under consideration; in this case value can have a relation with hedonistic concepts as 'reward' or be related to the physical dimension of stimuli. For example, by comparing two lights, we would say that the brighter one has a higher value. At the same time, of two sources of the same food we would say that the bigger has an higher value. It seems reasonable to assume a positive correlation in many ecological scenarios between stimulus magnitude (or salience) and fitness value; for example, a brighter fruit may be riper and thus more nutritionally beneficial (Schaefer et al., 2008), or a high intensity cue may indicate a more dangerous situation (Teodorescu et al., 2015). Prominent computational models of choice work by integrating difference (Ratcliff & McKoon, 2008) or ratio (Brown & Heathcote, 2008) in evidence between alternatives, thus disregarding information related to the absolute value of the alternatives under consideration (Teodorescu et al., 2015; Pirrone et al., 2014). Such systems may also exhibit decision deadlock between equal alternatives, which can be solved by adding urgency signals, asymmetry of inhibition or collapsing thresholds (Ditterich, 2006; Thura et al., 2012), however, these additions are motivated by avoiding long reaction times in low evidence trials, without explicit reference to implementing ecologically-relevant option magnitude sensitivity.

Consider, for example, a forager who encounters two food items. Laboratory formalism treats this choice as independent of other events (Bogacz et al., 2010), but if in the subject's natural environment food item availabilities and qualities are drawn from typical environmental distributions, then an optimal agent will be more willing to reject both items if they are matched and relatively low in value and instead search for a larger food item. However, if both items are matched and high in value, there is no sense in waiting, nor is there any benefit in deliberation between them. This decision-maker will thus be faster to respond to high-value stimuli than to low-value ones, even if their ratio or difference is identical.

A nonlinear model of decision-making, inspired by observations of house-hunting honeybees (Seeley et al., 2012), has been proposed that implements precisely this value-sensitive deadlock-breaking behaviour (Pais et al., 2013). The dynamics of the model are such that decisions between equal options below a value threshold result in deadlock, but deadlock is spontaneously broken for options above this value threshold; the value threshold is determined by a single biologically-relevant parameter, strength of cross-inhibition between evidence accumulating populations (Pais et al., 2013). An adaptive strategy is to progressively increase this parameter so that equal low-value alternatives that result in decision deadlock will eventually result in deadlock breaking (Pais et al., 2013); under this schedule high-value equal alternatives will result in deadlock breaking before low-value equal alternatives, and hence exhibit shorter reaction times in the former case. As the decision-maker moves from maintaining to breaking decision deadlock, change in the stochastic dynamics around the deadlock point corresponds to a sign change in the Ornstein-Uhlenbeck (O-U) process

$$\dot{x} = Bx + \sigma\eta \tag{6.1}$$

from stable ($B < 0$) to unstable ($B > 0$) (Pais et al., 2013). In equation 6.1 x represents state of the decision process, with 0 corresponding to decision deadlock and a decision being reached when x crosses a positive or negative threshold, η is a Wiener process, or Brownian motion, and σ is its standard deviation

Additionally, when differences between options are large enough the decision-mechanism approximates the classical drift-diffusion model of decision-making (Pais et al., 2013)

$$\dot{x} = A + \sigma\eta, \tag{6.2}$$

where x represents integrated evidence with 0 corresponding to equal evidence, and A is the strength of drift, which is a function of the difference between mean evidence strengths (Ratcliff, 1978). If there is no such difference then $A = 0$ and the decision variable will only cross a decision threshold through integrating sufficient noise; importantly, if decision thresholds have been set high (indicating a prioritisation of decision accuracy) and do not change, then this will take a correspondingly long time. While Pais *et al.* present a model of collective behaviour, corresponding non-linear neural models with qualitatively similar properties can be found (Bose et al., 2016).

A first demonstration of value sensitivity in human decision making comes from Teodorescu et al. (2015), and some preliminary results about magnitude sensitivity are also present in Teodorescu & Usher (2013). In Teodorescu et al. (2015) subjects were required to choose the brighter of two grey patches presented on the screen. Compared to a baseline condition, the authors increased the overall value of the alternatives while holding the ratio or the difference between the mean luminances of the two grey patches constant. Their results demonstrate that subjects show a sensitivity to the overall value of the alternatives both in the condition where the difference and the condition where the ratio are maintained constant but the overall value is increased. However, to the best of our knowledge, no study to date has investigated value sensitivity as a mechanism to break decision deadlocks for equal alternatives. We hypothesized that value sensitivity, exhibited by a model of decision-making in honeybee swarms (Pais et al., 2013), will also be observed in neural decision systems. We therefore measured the effects of value on matched-value decisions in two different contexts, perceptual decisions in humans and reward-based decisions in rhesus monkeys. In both cases, decisions of interest (i.e., equal alternatives) were embedded in a larger set of decisions between options of unequal value. In both cases, we observed a significant decrease in reaction time with increasing value for matched-value options. These findings are readily predicted by a value-sensitive model, but are not predicted by many classical models, except under implementations or assumptions that we discuss in our final remarks.

6.2 METHODS

6.2.1 HUMAN PERCEPTUAL DECISION TASK

For the Human Perceptual Decision Task, all procedures were approved by the University of Sheffield, Department of Psychology Ethics Sub-Committee (DESC), and carried out in accordance with the University and British Psychological Society (BPS) ethics guidelines. Subjects gave their informed consent before participation. We examined the behaviour of 9 human subjects (1 male, mean age = 18.8 years, SD = 1.64). All subjects had normal or corrected-to-normal vision and participated voluntarily in the experiment in exchange for course credit. Each subject was tested in a single sixty minute session.

Stimuli were programmed in Matlab, using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007), and were presented on a Mitsubishi Diamond Pro 2070sb 22" CRT monitor. Materials and procedure were similar to those used by Teodorescu et al. (2015), with the only exception being the addition of the equal-alternatives conditions.

As done by Teodorescu et al. (2015), we defined as 'multiplicative' the condition that held the same ratio between the two alternatives as in a baseline condition while increasing the overall value, and we defined as 'additive' the condition

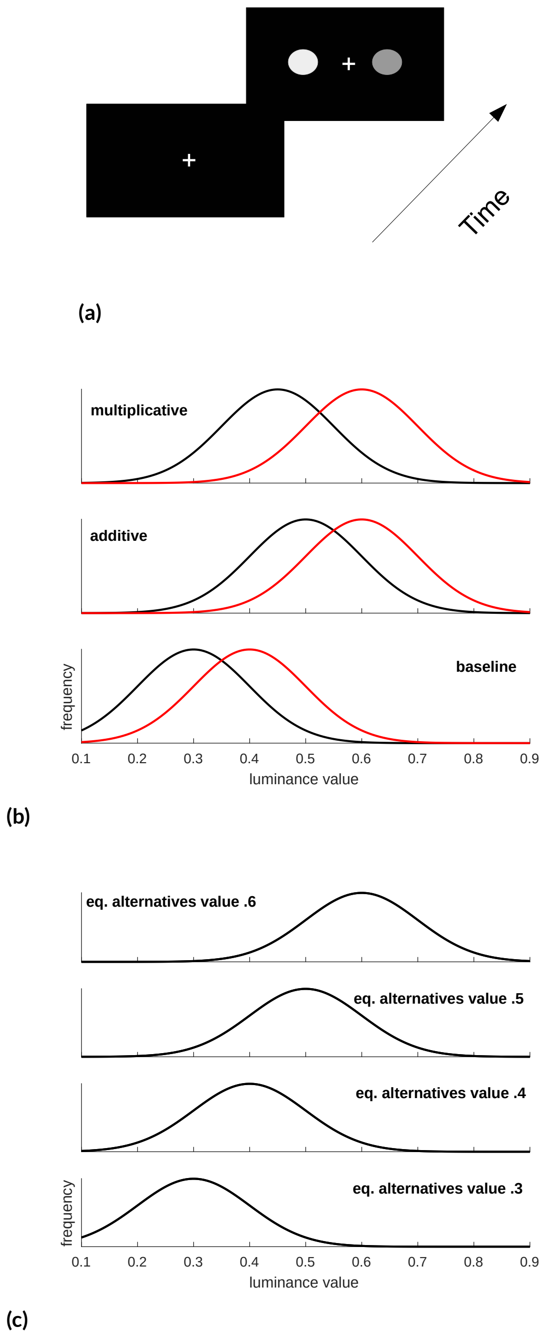


Figure 6.1: Top (a) Time course of an experimental trial; the stimulus remained on screen until subjects responded and 1 second after that subjects were presented a new trial. Middle (b) Luminance value distributions for the target (red) and non-target (black) alternatives. In the baseline condition alternatives had a mean of 0.4 and 0.3 respectively, hence a difference of 0.1 and a ratio of 4/3. In the additive condition, alternatives had a mean of 0.6 and 0.5 respectively, hence maintaining the same difference of 0.1 as in the baseline condition but giving a ratio of 6/5. In the multiplicative condition alternatives had a mean of 0.6 and 0.45 respectively, hence a difference of 0.15 but same ratio of 4/3 as in the baseline condition. Bottom (c) Equal alternative conditions of increasing value, respectively 0.3, 0.4, 0.5, 0.6.

in which the difference between the two alternatives was kept constant as in a baseline condition while the overall value of the alternatives was increased.

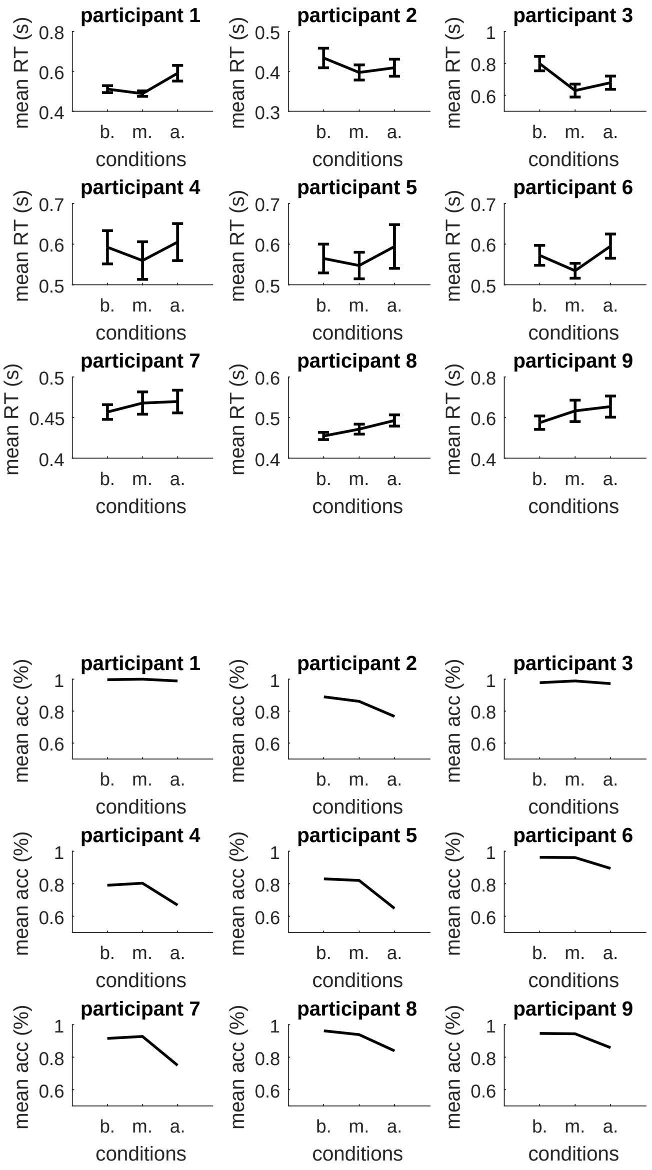
Stimuli consisted of two homogeneous, round, grey patches on a black background. The width of each patch was 1.2 cm; the distance between the centres of the two grey patches was 6.2 cm. A fixation cross was positioned between the two patches. The baseline array consisted of grey levels normally distributed around means of 0.4 and 0.3 (scale: 0 to 1.0), the multiplicative condition around means of 0.6 and 0.45, the additive condition of 0.6 and 0.5 and the four equal alternatives conditions were distributed respectively around means of 0.3, 0.4, 0.5 and 0.6; all conditions had a standard deviation of 0.1. On each frame, a Gaussian random variable with mean 0 and standard deviation of 0.01 was added to the mean grey level of each patch. If the final computed grey level was below 0.1, it was rounded to 0.1. The screen had a refresh rate of 60 Hz and subjects were positioned at 57 cm with their head on a chin rest. Order presentation of the two grey patches was counter-balanced for each subject. In the remainder we will refer to the four equal-alternatives conditions of increasing value with regards to their intensity (i.e., condition 0.3, condition 0.4, condition 0.5 and condition 0.6). Typical stimuli and value luminance distributions for the two alternatives are represented in Figure 6.1.

6.2.2 PROCEDURE

The two grey patches were presented simultaneously on the screen and subjects were asked to decide which of the two was brighter by pressing ‘left’ or ‘right’ on a keyboard using their left and right index fingers. One second after giving a response they were presented with a new trial. Subjects were not informed about the presence of equal-alternatives conditions or about the presence of a multiplicative and additive condition. Subjects performed 1400 trials of which 320 (22.9 %) were baseline trials and 180 (12.9 %) for each of the remaining conditions. After each block of 60 trials, subjects were asked to take a break and were presented on the screen with their accuracy and reaction times for the block. Accuracy was only computed for non-equal alternatives trials. Subjects were instructed to be as fast and accurate as possible and to maintain their fixation on the cross at the centre of the screen throughout each block. Before the experiment they were presented with 14 training trials (2 trials for each condition) to familiarise them with the task. No feedback was provided after each trial. No additional conditions or measures were collected.

6.3 Results of Human Study

No fast data were excluded from the following analyses, given that fast responses are particularly relevant for this study. However, we excluded slow responses over 3 seconds excluding in this way about 1 % of the data. To assess the effect of our manipulation on RTs and accuracy levels for the baseline, the additive and the multiplicative conditions, we show for each participant mean RTs, Figure 6.7a, and mean accuracy, Figure 6.7b with bars representing 95% confidence intervals of the mean. In interpreting a graph that shows 95% confidence intervals, when a confidence interval does not overlap with a specific value, it is possible to conclude



(a)

(b)

Figure 6.2: Top (a) Mean RTs for the baseline (b.), multiplicative (m.) and additive (a.) conditions for each subject. Bars represents 95% confidence intervals. Bottom (b) Mean accuracy levels for the baseline, multiplicative and additive conditions for each subject.

		estimate	t stat.	p value
ppt 1	intercept	1.05	18.94	<.001
	slope	-0.02	-1.17	0.242
ppt 2	intercept	0.57	19.39	<.001
	slope	-0.03	-2.49	0.013
ppt 3	intercept	1.25	23.08	<.001
	slope	-0.09	-4.35	<.001
ppt 4	intercept	0.74	20.68	<.001
	slope	-0.04	-3.09	0.002
ppt 5	intercept	0.66	22.55	<.001
	slope	-0.02	-2.29	0.022
ppt 6	intercept	0.71	30.48	<.001
	slope	-0.02	-2.42	0.016
ppt 7	intercept	0.51	49.54	<.001
	slope	-0.01	-1.86	0.063
ppt 8	intercept	0.55	48.10	<.001
	slope	-0.01	-2.54	0.011
ppt 9	intercept	0.91	21.95	<.001
	slope	-0.05	-3.02	0.003

Table 6.1: Estimate of slope and intercept, t statistic, and p value for the linear effect of value on RTs for the equal alternatives conditions (human participants).

that there is a statistical difference between the estimates of the values of interest at a false negative rate equal or lower than .05. For example, for the first participant, for mean RTs the graphs show that for the additive condition the subject was significantly slower than for the baseline or the multiplicative condition. For the multiplicative conditions the first participant did not differ from the baseline in mean RT. Regarding accuracy levels, the trend is consistent across subjects with subjects being generally less accurate for the additive condition compared to the baseline or the multiplicative, while the multiplicative condition remains the same as the baseline. Generally the accuracy of subjects is high especially for the baseline and the multiplicative condition with participant 1 and 3 being at ceiling level for all conditions. Regarding RTs however, there is no consistent pattern in how decision time varies across the three unequal conditions. In our experiment the display screen was not linearised with respect to brightness. Interestingly, the pattern reported coincides with what would be predicted given Weber fractions (Ross & Brodie, 1987) for the multiplicative condition compared to the baseline; by maintaining the same physical ratio between the two alternatives, compared to the baseline condition, the multiplicative condition results in similar accuracy and RTs. At the same time, the fact that the same physical difference between the two alternatives is maintained in the additive condition, leads to a smaller perceived difference of the additive condition compared to the baseline, resulting in slower and less accurate responses.

Recall that our interest is on the equal alternatives. To assess whether the effect of value on equal alternatives was consistent across subjects (Figure 6.3) we ran for

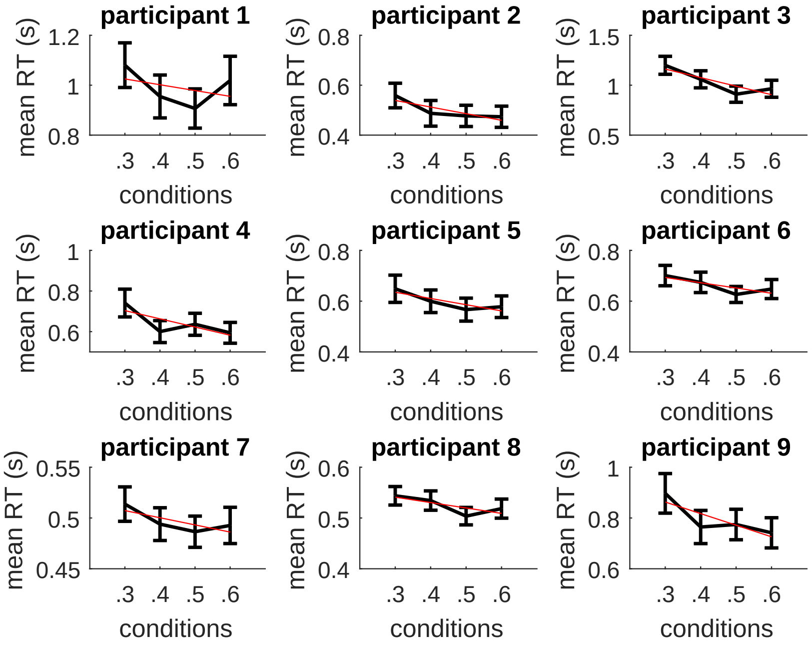


Figure 6.3: Top (a) Mean RTs for the equal alternatives conditions of increasing magnitude for each subject. Bars represents 95% confidence intervals. The red line represents the linear regression line on mean RTs.

each of the nine subjects a linear regression on mean RTs with value as predictor. For eight out of nine participants the regression slope was significantly non-zero or had a non-zero trend. Estimates of the slope, intercept, and significance levels are reported in Table 6.1.

6.4 Experiment on Monkeys

This experiment was performed by Habiba Azab at Rochester University, under the supervision of Benjamin Hayden. However, I have participated in the development of the experimental protocol. The section 'Basic Procedures', and the section 'Monkey Behavioural Task' have been written by Habiba Azab and Benjamin Hayden and are taken from the paper currently under review. I have performed the analyses on the monkey data.

6.4.1 Basic Procedures

The basic procedures used in this study were based on existing protocols used for other experiments (Blanchard et al., 2013). All procedures were approved by the University of Rochester Institutional Animal Care and Use Committee and were designed and conducted in compliance with the Public Health Service's Guide for the Care and Use of Animals. Four male rhesus monkeys (*Macaca mulatta*) served as subjects. Each animal was outfitted with a small prosthesis using a standard technique (Hayden et al., 2008). Animals received analgesics and antibiotics after all surgeries. Animals were slowly habituated to laboratory conditions and trained to perform oculomotor tasks for liquid reward. Standard reinforcement training was used with only positive rewards; punishment was never used, nor was aversive conditioning.

In each session, the animal was transported from the colony at the University of Rochester to the testing room, about 100 feet away in the same building. The testing room was built specifically for primate studies and houses a computer

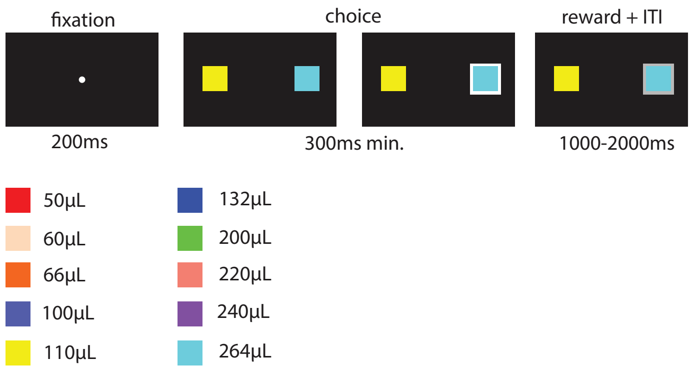


Figure 6.4: Time course of an experimental trial and reward values for the two alternatives for the monkey experiment. RGB values for stimuli were as follows: - red: [255 0 0] - off-white: [255 218 185] - orange: [255 102 0] - indigo: [78 90 200] - yellow: [255 255 0] - blue: [0 0 255] - lime green: [0 255 0] - pink: [250 128 114] - purple: [160 32 240] - cyan: [0 255 255] - white (fixation dot): [255 255 255].

screen and floor plate for firm mounting of the ergonomically designed primate chair (Crist). Animals made all task-relevant decisions using gaze shifts to selected targets. Horizontal and vertical eye positions were sampled at 1000 Hz by an infrared eye-monitoring camera system (SR Research). Stimuli were controlled by a computer running Matlab (MathWorks) with Psychtoolbox (Brainard, 1997) and Eyelink Toolbox (Cornelissen et al., 2002).

A standard solenoid valve controlled the duration of water delivery (Parker). We estimated the precision of fluid volume delivered by the solenoid across the range of open time commands used in this study. All reward volumes were measured and confirmed. Fluid access was controlled outside of experimental sessions.

6.4.2 Monkey Behavioural Task

A two-alternative forced choice task was used to study the effect of overall magnitude of the decision variable on reaction time in macaques. The task is a computerized implementation of a simple economic choice task, of the type we and others have long used. This task uses the same basic structure as several other tasks in the lab, including those used to study risk (Blanchard et al., 2014), intertemporal choice and foraging (Blanchard & Hayden, 2015), and curiosity (Blanchard et al., 2015). The key novel elements of this task were the use of simultaneous option presentation with speeded responses. We used a computerized presentation, with a standard LCD monitor placed 144.8 cm (57 inches) inches in front of the monkey in a darkened room. Screen resolution was 1024x768. All trials were identical aside from the specific values and colours used. On each trial, monkeys first fixated on a small white central spot (50 px diameter, 200 ms duration) to indicate their willingness to initiate the trial. Successful fixation led to the immediate presentation of two choice options; monkeys were allowed to select the choice option (by shifting gaze to it) immediately; no minimum initial fixation was required, nor were monkeys required to look at both options before making a choice. The computer selected two options independently and at random, with a uniform distribution. It then presented them 300 pixels to the left and right of the central spot. Both stimuli were squares (200 pixels wide) in one of 10 colours. The colors we used were red, off-white, orange, indigo, yellow, blue, lime green, pink, purple and cyan. These colours indicated the size of the reward offered by this option, according to the following scheme: red: 50μL, off-white: 60μL, orange: 66μL, indigo: 100μL, yellow: 110μL, blue: 132μL, lime green: 200μL, pink: 220μL,, purple: 240μL , and cyan: 264μL. We chose these particular reward values carefully to allow us to have several ratios with different magnitudes. Thus, while subjects saw trials in all possible combinations of the above 10 stimuli, we were particularly interested in subsets of trials that form the focus of our analyses. Subjects had extensive experience with the reward-colour mappings of most colours in this hierarchy of rewards from previous experiments (specifically: red, orange, yellow, blue, lime green, purple and cyan; Blanchard & Hayden, 2015; Strait et al., 2016). To ensure that this familiarity did not introduce any special bias, we extensively familiarized our subjects with the rewards offered by new colours in several training sessions prior to testing. Following presentation, the subject then selected an option by shifting their gaze toward it. Subjects were required to maintain fixation on their choice for 300ms. Failure to maintain fixation led to deselection of the option and returned the monkey to the choice state. Thus, monkeys were allowed to inspect the options without committing to them if they wanted. Once the subject successfully completed fixation, the reward was given and an inter-trial interval of 1 s, 1.5 s, or 2 s began. The particular ITI on a given trial was selected at random from a uniform distribution. Options remained on the screen during reward delivery and throughout the inter-trial interval. Typical stimuli and reward values for the two alternatives are represented in Figure 6.4. No additional measures or conditions were collected.

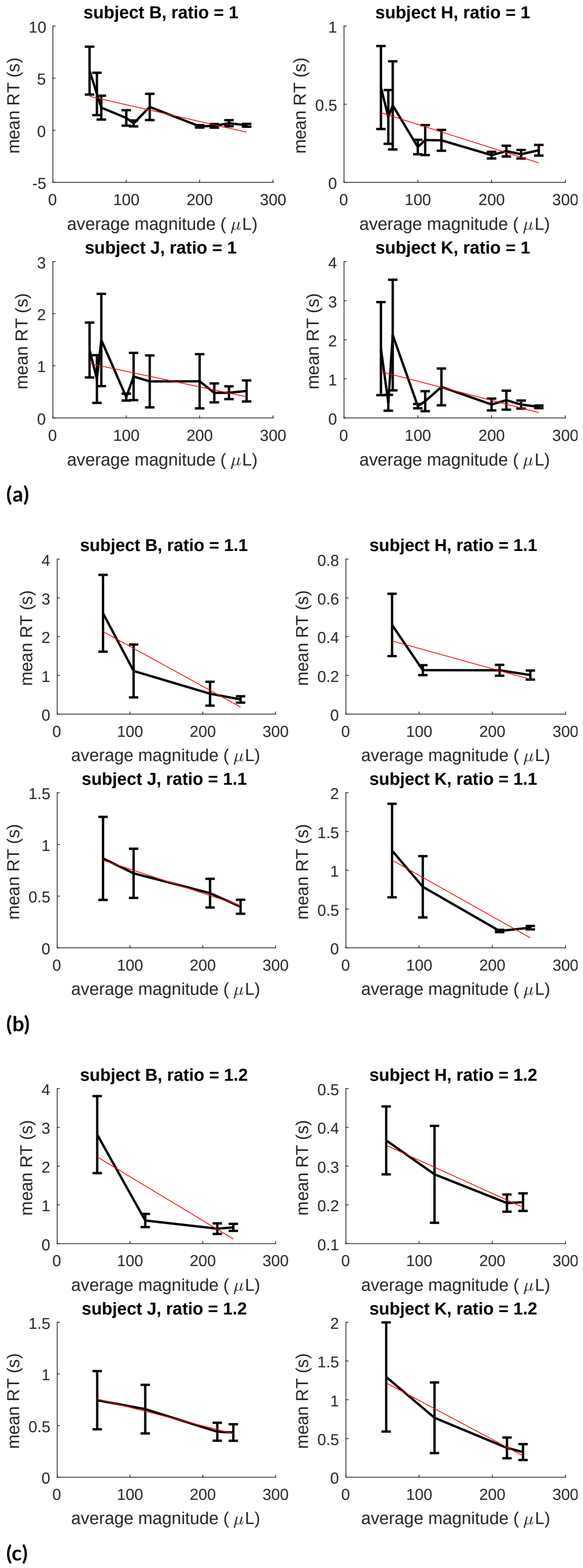


Figure 6.5: Mean RTs for conditions with constant ratio of increasing magnitude for each subject. Bars represents 95% confidence intervals. The red line represents the linear regression line on mean RTs. Top (a) ratio =1 (equal alternatives). Middle (b) ratio =1.1. Bottom (c) ratio =1.2.

6.5 Results of Monkey Study

All subjects initially performed over 9000 trials of this task (subject B: 9132 trials, subject H: 11652 trials, subject J: 11150 trials, subject K: 10230 trials). The exact number of trials performed by each subject was constrained by the subject's willingness to work on any given day, and the need to start them on different tasks. Subjects were highly accurate in their choices (overall accuracy: 85.41%; subject B: 87.64%; subject H: 87.69%; subject J: 88.93%; subject K: 77.01%). These values are all significantly greater than chance (two-sided binomial test, all $p < 0.0001$). No fast data were excluded from the following analyses but we removed the slowest 0.5 % of trials per subject, which represents unreasonably slow RTs. Given the variability across subjects in mean RT, we could not use a single common value for an upper cutoff as done for the human data. To assess the effect of our manipulation on correct RTs we show for each participant mean RTs with bars representing 95% confidence intervals for all those ratio conditions for which more than three magnitude levels were present, separately for each participant; Figure 6.5, Figure 6.6 and Figure 6.7.

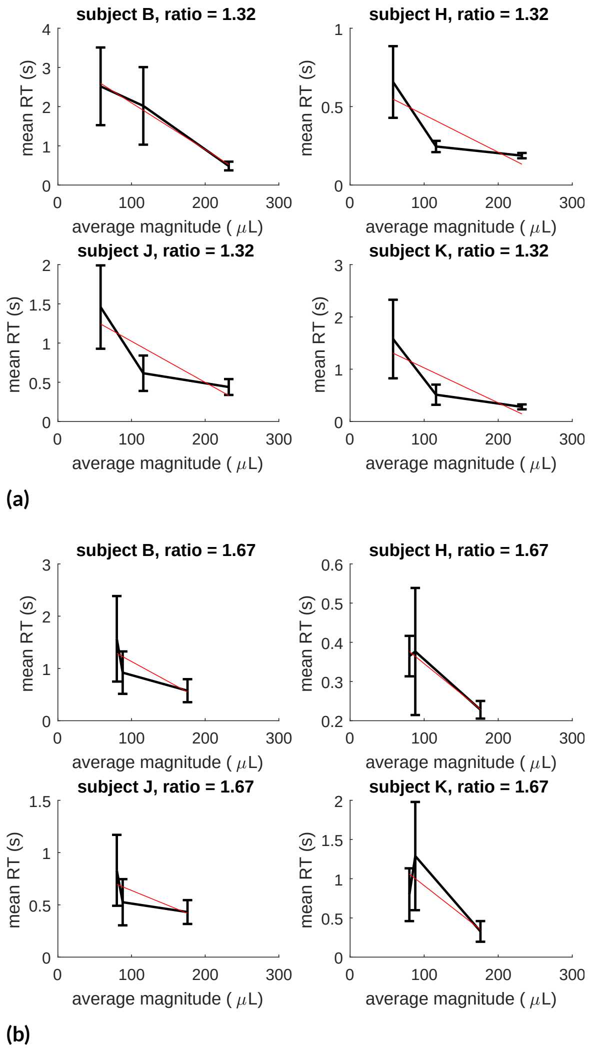


Figure 6.6: Mean RTs for conditions with constant ratio of increasing magnitude for each subject. Bars represents 95% confidence intervals. The red line represents the linear regression line on mean RTs. Top (a) ratio =1.32. Bottom (b) ratio =1.67.

		estimate	t stat.	p value
subject B	intercept	4.06	4.85	0.001
	slope	-0.01	-3.13	0.014
subject H	intercept	0.52	8.13	<.001
	slope	0.00	-3.82	0.005
subject J	intercept	1.19	6.05	<.001
	slope	0.00	-2.46	0.039
subject K	intercept	1.43	3.69	0.006
	slope	0.00	-2.06	0.073

Table 6.2: Estimate of slope and intercept, t statistic, and p value for the linear effect of value on RTs for the equal alternatives conditions (monkey participants).

Our condition of interest is the condition for which the ratio is 1, meaning that the two alternatives that the subject were presented with were equal in value. For these conditions, as done for the human experiment, we ran for each of the four subjects a linear regression on RTs with value as predictor. For three participants the regression slope was significantly non-zero while for one participant it had a non-zero trend. Estimates of the slope, intercept, and significance levels are reported in Table 6.2.

6.6 Discussion

Influenced by a model of value-sensitive decision-making (Pais et al., 2013) and by evolutionary and ecological arguments (Pirrone et al., 2014; Teodorescu et al., 2015) we have investigated the effect of the overall value of the alternatives on decision making, in humans and in monkeys. In line with these arguments, our ini-

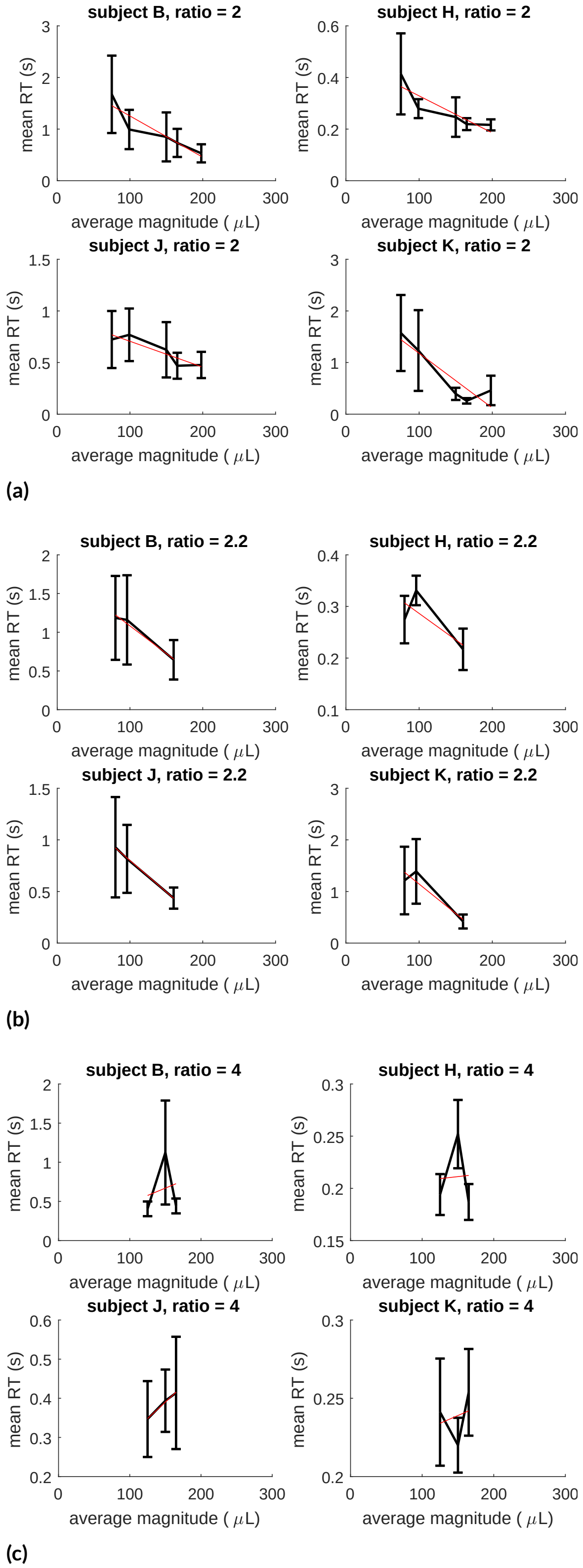


Figure 6.7: Mean RTs for conditions with constant ratio of increasing magnitude for each subject. Bars represents 95% confidence intervals. The red line represents the linear regression line on mean RTs. Top (a) ratio =2. Middle (b) ratio =2.2. Bottom (c) ratio =4.

tial prediction was that an effect of the overall value of the alternatives should be present also for 'equal' alternatives: fast decision times when the overall value of the alternatives is high and slow decisions when the overall value is low. Both the perceptual decision-making experiment on humans and the economic decision-making experiment on monkeys provide evidence that the overall value of the alternatives affects response times. These effects are not predicted by classical models of choice which integrate only differences between or ratios of alternatives. Value sensitivity might seem to be counter-intuitive if considered from a speed-accuracy trade off perspective. From a speed-accuracy point of view, choices involving more valuable options may be more costly to make mistakes on, so we might expect decision making to shift towards a low error regime and, hence, be slower. Instead, we observe the opposite since when the overall value is increased, subjects are faster and could open themselves to making more errors. This result, for the value-based task is in line with a 'satisficing' perspective where a 'good enough' choice is preferred rather than the 'best', and as a consequence accuracy in decisions over small differences is sacrificed in favour of quick responses (Pirrone et al., 2014; Kacelnik et al., 2011).

Unfortunately, due to a programming error, in our experiment the display screen was not linearised (gamma corrected) with respect to brightness. This means that our results hold for physical rather than perceived multiplicative and additive shifts with respect to the baseline. As a consequence, our results on non-equal alternative conditions are not directly comparable to those of Teodorescu et al. (2015). For example, they found a difference in performance between the baseline and multiplicative conditions which we did not, probably due to our stimuli being shifted by a smaller physical amount. However, the equal alternative conditions, which are the focus of our work, do not suffer from issues related to normalisation. The consistency across subjects for these conditions, as shown in Figure 6.3, represents a simple but effective test of value-sensitivity in human perceptual decision-making for deadlock breaking.

Relevant to our monkey experiment, regarding the unequal conditions (i.e., the ratio between the two alternatives is not 1), no analyses were performed. These conditions were presented to allow subjects to focus on the task; clearly an experiment consisting only of equal alternatives would be unreasonable as for all trials each choice would be random by necessity. Moreover, these unequal conditions do not allow to test for value sensitivity given that when the ratio between two alternatives is kept constant but the overall value is increased, also the discriminability between the two alternatives increases -assuming constant noise- resulting in decreasing RTs. This means that although for unequal alternatives RTs generally decrease as magnitude increases, it is not possible to dissociate the effect of magnitude from the effect of increased disriminability between the two alternatives. However, also for the monkey data the presence of equal alternatives conditions (e.g., ratio=1) allows us to test and confirm value sensitivity in monkey reward-based decision making.

A strength of presenting both sets of data using different species and domains is that this finding seem to suggest that value guides decision making, regardless of the specific domain. We believe that this supports the idea of a single common mechanism underlying decision making that given evolutionary pressures is value sensitive for perceptual stimuli and for rewards (Pirrone et al., 2014; Teodorescu et al., 2015).

Our point, argued in the previous chapter, is that most naturalistic decisions are value-based rather than accuracy-based, in the sense that decision-makers are rewarded by the value of the alternative chosen, regardless of whether it was the best available. Although decision-making is traditionally studied within the speed-accuracy tradeoff perspective, this alternative viewpoint suggests that a speed-value tradeoff (Pirrone et al., 2014) could be the most relevant decision tradeoff to manage in various naturalistic settings (Bateson & Kacelnik, 1998). We believe that the value-sensitivity shown in simple tasks such as those presented in this paper is a signature of this evolutionarily-plausible strategy.

These findings stand in contrast to celebrated models of choice. For example, the Drift Diffusion Model (Ratcliff & McKoon, 2008) assumes that the subject integrates difference in evidence supporting two alternatives until a decision boundary is crossed and a decision is made in favour of that alternative. This reliance on evidence difference rather than evidence value entails predictions of equal RTs for choices between two options of equal difference regardless if they are two high value options or two low value options.

Theoretically, value sensitivity of the kind we have demonstrated here can be explained by a number of models in addition to the one we took as our starting point (Pais et al., 2013). Teodorescu et al. (2015) show that under the neurally plausible assumption that processing noise increases with stimulus value, then a difference-based diffusion model becomes value sensitive and can make similar predictions. Other computational models of choice such as the Leaky Competing Accumulator (Usher & McClelland, 2001, LCA) can also give rise to similar patterns. The LCA at the early stages of accumulation shows a sensitivity to the overall value of the alternatives and at the later stages approximates a DDM (Bogacz et al., 2006), hence it is a value sensitive model. At the same time, models in which the cost of accumulating evidence increases over time (Drugowitsch et al., 2012) or sequential choice 'race' models (as compared to models in which the decision maker explicitly compares options) in which agents choose an option that exceeds a fixed threshold of acceptability (Kacelnik et al., 2011) are in line with the value sensitive reaction time results presented here. Further theoretical effort should be made to determine which empirical data on value-sensitivity can be explained by which models, and attempt to discriminate between them on this basis. As noted in earlier work, the nonlinear dynamics of models that explicitly implement value-sensitive decision-making give rise to a further prediction, of decision hysteresis (Pais et al., 2013), which may motivate further experimental investigation.

Our results were inspired by a model of choice that involves explicit mutual inhibition in economic and perceptual decisions. Neural activity in several reward regions in the brain shows evidence of mutual inhibition during economic decisions. These regions include the ventromedial prefrontal cortex (Strait et al., 2014), ventral striatum (Strait et al., 2015), orbitofrontal cortex (Padoa-Schioppa, 2011), dorsal premotor area (Pastor-Bernier et al., 2012), and parietal cortex (Louie et al., 2011). Human neuroimaging results also support this (Hunt et al., 2015; Jocham et al., 2012; Hunt et al., 2012). While a direct link between this literature and the present study remains speculative, the similarity is nonetheless striking. Future work will be required to determine whether these neural processes instantiate the mechanism that our investigation was motivated by.

In conclusion, we hypothesise that far from being an artefact of imperfect implementation, longer RTs with low-value alternatives and shorter RTs with high-value alternatives are diagnostic of an adaptive decision strategy for the uncertain environments faced by decision making systems, at different level of biological complexity, and in various domains.

7

Conclusion

7.1 Overview of the studies and main findings

Our research has focused on computational models of choice. Computational models of choice are a useful, powerful and necessary tool for research in cognition. The computational approach has offered us a way to investigate hypotheses in in various domains. Furthermore, it has allowed us to investigate new features of decision making. In the first part of this dissertation we have focused on the Drift Diffusion Model (Ratcliff & McKoon, 2008) and we have applied it to three domains: orientation discrimination in Autism Spectrum Disorder, task-irrelevant perceptual learning and magnitude comparison research. Applying the DDM to these fields has allowed us to provide important insights into such phenomena.

Regarding perception in ASD, we have shown that with regards to our task, subjects with ASD did not differ in sensitivity compared to control subjects, while their main difference was in their criterion for a response and in their non-decision time component. This finding has fundamental implications for the investigation of perception in ASD in which, surprisingly enough, conclusions based on RTs or accuracy data have never taken into account speed-accuracy trade-offs or non-decision time effects.

Regarding TIPL, we have shown as well how this phenomenon is not monolithically affecting only one of the decision mechanisms underlying a decision but has instead a diffuse effect on other mechanisms, such as the criterion for a response. Not taking into account such effects can lead to ambiguous estimates of learning and interpretation of data.

Regarding magnitude research, we have shown that competing theories for the explanation of the semantic congruity effect can be reconciled within the DDM framework and their predictions tested.

In the second part of this dissertation, we have instead focused on another model of choice, the model presented in Pais et al. (2013). This model has a particular feature that has never been taken into consideration in decision making research so far, magnitude sensitivity.

Firstly, in chapter 5, we have argued from a theoretical perspective why we should expect decision making to have evolved towards magnitude sensitivity and why for many types of decision, a particular type of trade-off, that we have defined as the speed-value trade-off, should be maximised instead of the 'classical' speed-accuracy trade-off.

Finally, in chapter 6, we have provided evidence for the existence of the speed-value trade-off across species and domains.

The results of the second part of this work have important consequences for decision making, both theoretically and computationally. First, these results are important for the falsifiability of computational models of choice (Heathcote et al., 2014). One problem with computational models of choice is that for *most* tasks they tend to make common predictions; for example, say that a researcher performs a RDK experiment in which the difficulty of the task and the instructions regarding whether to be fast or accurate are manipulated. Most likely, regardless of the computational model of choice used, the researcher would find that the difficulty of the task affects the parameter associated with the discriminability (e.g., the drift rate in the DDM or in the LCA) while the speed/accuracy instructions would instead affect the parameter related to the boundary separation. This means that models generally tend to make common predictions and similar accounts for the decision making data they are applied to. Very few are the predictions in decision making that are only met by some models of choice and cannot be accounted by others: magnitude sensitivity is one of these predictions. The investigation of magnitude sensitivity allowed us to compare the qualitative predictions of decision making models. In particular, *relative* models of choice predict magnitude insensitivity for equal alternatives of increasing magnitude, while *absolute* models of choice predict a decrease in the decision time when equal alternatives have an increase in magnitude. By showing magnitude sensitivity in decision making we suggest that *relative* models of choice are not able to describe a phenomenon that could have an important role in making a decision.

7.2 FUTURE RESEARCH

Together with some interesting results, I believe that the most exciting aspect of our work is represented by the hypotheses that it has generated for future research, in particular for two fields: ASD research and research on the falsifiability of computational models of choice.

Regarding ASD research, previous research has reported results that have been interpreted as enhancement or impairments in specific areas of perception (Bertone et al., 2003; Milne et al., 2002; Deruelle et al., 2004; Bertone et al., 2003; O'Riordan et al., 2001; Dickinson et al., 2016). This confusion affects even specific fields such as motion discrimination, for which both impairments and enhancements have been reported (Bertone et al., 2003; Milne et al., 2002). An interesting question for future research is investigating whether such contrastive results can be reconciled by a computational approach and in particular by accounting for the speed-accuracy trade-off and by taking into account differences in non-decision time and boundary separation. Also, future research should investigate what are the causes for the increased boundary separation in subjects with ASD. We believe that subjects with ASD give a higher weight to reward prediction errors (i.e., errors have a stronger impact) compared to control subjects, and as a consequence they use a more conservative criterion for a response that open themselves to less errors.

With regards to magnitude sensitivity research, future investigations are warranted to explore which mechanism gives rise to magnitude sensitivity; for example it could be dependent on the lateral inhibition between the accumulators, or it could be dependent on the noise that rather than being constant is instead magnitude dependent. Future research should quantify the predictions of different models that in principle could account for magnitude sensitivity and compare them. On the other side, for models that cannot account for magnitude sensitivity, an interesting question is whether such models should be modified in order to account for magnitude sensitivity. Also, future research is needed to quantify the loss in explanation in various domains when magnitude sensitivity is not taken into account. It could be for example, that perceptual decision making is only minimally affected by magnitude sensitivity while reward decision making in ecological settings is instead maximally affected by magnitude sensitivity.

Regarding TIPL and perceptual learning research in general, we believe that studies that have documented an absence of learning based on analysis of accuracy data should be reconsidered, as we have shown that variation in the cognitive processes underlying a decision can mask a variation in accuracy. For example, as discussed in Liu & Watanabe (2012), claims that external feedback does not affect learning (e.g., Ball & Sekuler, 1982; Watanabe et al., 2001) are based on the analysis of accuracy data alone. An inclusion of RTs and a computational decomposition of the effect could nevertheless show that a null effect is instead due to the combined effect of a decrease in boundary separation accompanied by an increase in discriminability. Furthermore, other parameters that might be associated with learning such as the non-decision time component, have an effect that only manifest in RTs and cannot be revealed by an inspection of accuracy data alone.

Finally, regarding magnitude comparison research, further experiments should adopt a procedure in which the direction of the comparison is explicitly given (e.g., choose the smaller/bigger) and in which a computational investigation of the results is performed in order to corroborate our findings. More importantly, we strongly discourage using ad-hoc theories, but instead believe that psychological phenomena, if possible, should be studied within the framework of unifying computational models, as for example the DDM.

References

Armel, K. C., Beaumel, A., & Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making*, 3(5), 396–403.

Ball, K. & Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science*, 218(4573), 697–698.

Banks, W. P., Clark, H. H., & Lucy, P. (1975). The locus of the semantic congruity effect in comparative judgments. *Journal of Experimental Psychology: Human Perception and Performance*, 1(1), 35.

Banks, W. P. & Flora, J. (1977). Semantic and perceptual processes in symbolic comparisons. *Journal of Experimental Psychology: Human Perception and Performance*, 3(2), 278.

Banks, W. P., Fujii, M., & Kayra-Stuart, F. (1976). Semantic congruity effects in comparative judgments of magnitudes of digits. *Journal of Experimental Psychology: Human Perception and Performance*, 2(3), 435.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (aq): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of autism and developmental disorders*, 31(1), 5–17.

Basten, U., Biele, G., Heekeren, H. R., & Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences*, 107(50), 21767–21772.

Bateson, M. & Kacelnik, A. (1998). Risk-sensitive foraging: decision making in variable environments. *Cognitive ecology*, (pp. 297–341).

Bertone, A., Mottron, L., Jelenic, P., & Faubert, J. (2003). Motion perception in autism: a "complex" issue. *Journal of cognitive neuroscience*, 15(2), 218–225.

Bhat, A. N., Landa, R. J., & Galloway, J. C. C. (2011). Current perspectives on motor functioning in infants, children, and adults with autism spectrum disorders. *Physical Therapy*, 91(7), 1116–1129.

Birnbaum, M. H. & Jou, J.-W. (1990). A theory of comparative response times and "difference" judgments. *Cognitive Psychology*, 22(2), 184–210.

Blanchard, T. C. & Hayden, B. Y. (2015). Monkeys are more patient in a foraging task than in a standard intertemporal choice task. *PloS one*, 10(2), e0117057.

Blanchard, T. C., Hayden, B. Y., & Bromberg-Martin, E. S. (2015). Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity. *Neuron*, 85(3), 602–614.

Blanchard, T. C., Pearson, J. M., & Hayden, B. Y. (2013). Postreward delays and systematic biases in measures of animal temporal discounting. *Proceedings of the National Academy of Sciences*, 110(38), 15491–15496.

Blanchard, T. C., Wilke, A., & Hayden, B. Y. (2014). Hot-hand bias in rhesus monkeys. *Journal of Experimental Psychology: Animal Learning and Cognition*, 40(3), 280.

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.

Bogacz, R., Hu, P. T., Holmes, P. J., & Cohen, J. D. (2010). Do humans produce the speed–accuracy trade-off that maximizes reward rate? *The Quarterly Journal of Experimental Psychology*, 63(5), 863–891.

Bose, T., Reina, A., & Marshall, J. A. R. (2016). Neural models of value-sensitive decision-making. *manuscript in preparation*.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, 10, 433–436.

Brown, S. D. & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive psychology*, 57(3), 153–178.

Browne, M. W. (2000). Cross-validation methods. *Journal of mathematical psychology*, 44(1), 108–132.

Bruyer, R. & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: is the inverse efficiency score (ies) a better dependent variable than the mean reaction time (rt) and the percentage of errors (pe)? *Psychologica Belgica*, 51(1), 5–13.

Busemeyer, J., Diederich, A., Heathcote, A., Rangel, A., Rieskamp, J., & Usher, M. (2013). Sequential sampling models representing a unifying framework of human decision making. *Symopsium, COGSCI Annual Meeting*.

Busemeyer, J. R. & Diederich, A. (2010). *Cognitive modeling*. Sage.

Busemeyer, J. R. & Rapoport, A. (1988). Psychological models of deferred decision making. *Journal of Mathematical Psychology*, 32(2), 91–134.

Cantlon, J. F. & Brannon, E. M. (2005). Semantic congruity affects numerical judgments similarly in monkeys and humans. *Proceedings of the National Academy of Sciences of the United States of America*, 102(45), 16507–16511.

Cassey, T. C., Evens, D. R., Bogacz, R., Marshall, J. A. R., & Ludwig, C. J. H. (2013). Adaptive sampling of information in perceptual decision-making. *PloS one*, 8(11), e78993.

Chen, D., Lu, H., & Holyoak, K. J. (2014). The discovery and comparison of symbolic magnitudes. *Cognitive psychology*, 71, 27–54.

Chittka, L., Skorupski, P., & Raine, N. E. (2009). Speed–accuracy tradeoffs in animal decision making. *Trends in Ecology & Evolution*, 24(7), 400–407.

Constantino, J. N. & Gruber, C. P. (2007). *Social responsiveness scale (SRS)*. Western Psychological Services Los Angeles, CA.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The eyelink toolbox: eye tracking with matlab and the psychophysics toolbox. *Behavior Research Methods, Instruments, & Computers*, 34(4), 613–617.

Davies, N. B., Krebs, J. R., & West, S. A. (2012). *An Introduction to Behavioural Ecology*. John Wiley & Sons, 4th edition.

D'Cruz, A.-M., Ragozzino, M. E., Mosconi, M. W., Shrestha, S., Cook, E. H., & Sweeney, J. A. (2013). Reduced behavioral flexibility in autism spectrum disorders. *Neuropsychology*, 27(2), 152.

Deco, G. & Rolls, E. T. (2006). Decision-making and weber's law: a neurophysiological model. *European Journal of Neuroscience*, 24(3), 901–916.

Deco, G., Scarano, L., & Soto-Faraco, S. (2007). Weber's law in decision making: integrating behavioral data in humans with a neurophysiological model. *The Journal of Neuroscience*, 27(42), 11192–11200.

Dehaene, S. (1989). The psychophysics of numerical comparison: A reexamination of apparently incompatible data. *Perception & psychophysics*, 45(6), 557–566.

Deneve, S. (2012). Making decisions with unknown sensory reliability. *Frontiers in Neuroscience*, 6.

Deruelle, C., Rondan, C., Gepner, B., & Tardif, C. (2004). Spatial frequency and face processing in children with autism and asperger syndrome. *Journal of autism and developmental disorders*, 34(2), 199–210.

Dickinson, A., Bruyns-Haylett, M., Smith, R., Jones, M., & Milne, E. (2016). Superior orientation discrimination and increased peak gamma frequency in autism spectrum conditions. *Journal of abnormal psychology*, 125(3), 412.

Ditterich, J. (2006). Stochastic models of decisions about motion direction: behavior and physiology. *Neural Networks*, 19(8), 981–1012.

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *The Journal of Neuroscience*, 32(11), 3612–3628.

Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2009). A diffusion model decomposition of the practice effect. *Psychonomic Bulletin & Review*, 16(6), 1026–1036.

Edwards, W. (1965). Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology*, 2(2), 312–329.

Fine, I. & Jacobs, R. A. (2002). Comparing perceptual learning across tasks: A review. *Journal of vision*, 2(2), 5–5.

Fritz, A., Percy, C., Jack, A., Shanmugaratnam, K., Sobin, L., Parkin, D. M., Whelan, S., et al. (2000). *International classification of diseases for oncology.* Number Ed. 3. World Health Organization.

Fum, D., Del Missier, F., & Stocco, A. (2007). The cognitive modeling of human behavior: Why a model is (sometimes) better than 10,000 words. *Cognitive Systems Research*, 8(3), 135–142.

Furmanski, C. S. & Engel, S. A. (2000). Perceptual learning in object recognition: Object specificity and size invariance. *Vision research*, 40(5), 473–484.

Gao, J., Tortell, R., & McClelland, J. L. (2011). Dynamic integration of reward and stimulus information in perceptual decision-making. *PloS one*, 6(3), e16749.

Gebuis, T. & Reynvoet, B. (2012). The role of visual information in numerosity estimation. *PloS one*, 7(5), e37426.

Gertner, L., Arend, I., & Henik, A. (2012). Effects of non-symbolic numerical information suggest the existence of magnitude–space synesthesia. *Cognitive processing*, 13(1), 179–183.

Gold, J. I. & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in cognitive sciences*, 5(1), 10–16.

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.

Gold, J. I. & Shadlen, M. N. (2003). The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. *The Journal of Neuroscience*, 23(2), 632–651.

Grasman, R. P., Wagenmakers, E.-J., & van der Maas, H. L. (2009). On the mean and variance of response times under the diffusion model with an application to parameter estimation. *Journal of Mathematical Psychology*, 53(2), 55–68.

Hanks, T. D., Mazurek, M. E., Kiani, R., Hopp, E., & Shadlen, M. N. (2011). Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *The Journal of Neuroscience*, 31(17), 6339–6352.

Hare, T. A., Schultz, W., Camerer, C. F., O'Doherty, J. P., & Rangel, A. (2011). Transformation of stimulus value signals into motor commands during simple choice. *Proceedings of the National Academy of Sciences*, 108(44), 18120–18125.

Hayden, B. Y., Nair, A. C., McCoy, A. N., & Platt, M. L. (2008). Posterior cingulate cortex mediates outcome-contingent allocation of behavior. *Neuron*, 60(1), 19–25.

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

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in neuroscience*, 8, 150.

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

Holyoak, K. J. (1978). Comparative judgments with numerical reference points. *Cognitive Psychology*, 10(2), 203–243.

Holyoak, K. J. & Mah, W. A. (1982). Cognitive reference points in judgments of symbolic magnitude. *Cognitive Psychology*, 14(3), 328–352.

Holyoak, K. J. & Morrison, R. G. (2005). *The Cambridge handbook of thinking and reasoning*. Cambridge University Press.

Householder, A. S. & Young, G. (1940). Weber laws, the weber law, and psychophysical analysis. *Psychometrika*, 5(3), 183–193.

Hunt, L. T., Behrens, T. E., Hosokawa, T., Wallis, J. D., & Kennerley, S. W. (2015). Capturing the temporal evolution of choice across prefrontal cortex. *Elife*, 4, e11945.

Hunt, L. T., Kolling, N., Soltani, A., Woolrich, M. W., Rushworth, M. F., & Behrens, T. E. (2012). Mechanisms underlying cortical activity during value-guided choice. *Nature neuroscience*, 15(3), 470–476.

Jocham, G., Hunt, L. T., Near, J., & Behrens, T. E. (2012). A mechanism for value-guided choice based on the excitation-inhibition balance in prefrontal cortex. *Nature neuroscience*, 15(7), 960–961.

Johnson, C. A. & Leibowitz, H. W. (1979). Practice effects for visual resolution in the periphery. *Perception & Psychophysics*, 25(5), 439–442.

Kable, J. W. & Glimcher, P. W. (2009). The neurobiology of decision: consensus and controversy. *Neuron*, 63(6), 733–745.

Kacelnik, A., Vasconcelos, M., Monteiro, T., & Aw, J. (2011). Darwin's "tug-of-war" vs. starlings'"horse-racing": how adaptations for sequential encounters drive simultaneous choice. *Behavioral Ecology and Sociobiology*, 65(3), 547–558.

Kahneman, D. & Tversky, A. (2000). *Choices, values, and frames*. Cambridge University Press.

Kanner, L. et al. (1943). Autistic disturbances of affective contact.

Karalunas, S. L., Huang-Pollock, C. L., & Nigg, J. T. (2012). Decomposing attention-deficit/hyperactivity disorder (adhd)-related effects in response speed and variability. *Neuropsychology*, 26(6), 684.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, 36(14), 1–1.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, 13(10), 1292–1298.

Krajbich, I., Hare, T., Bartling, B., Morishima, Y., & Fehr, E. (2015). A common mechanism underlying food choice and social decisions. *PLoS Comput Biol*, 11(10), e1004371.

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in psychology*, 3, 193.

Krajbich, I. & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, 108(33), 13852–13857.

Latty, T. & Beekman, M. (2011). Speed–accuracy trade-offs during foraging decisions in the acellular slime mould physarum polycephalum. *Proceedings of the Royal Society B: Biological Sciences*, 278(1705), 539–545.

Leth-Steensen, C. & Marley, A. (2000). A model of response time effects in symbolic comparison. *Psychological Review*, 107(1), 62.

Leth-Steensen, C., Petrusic, W. M., & Shaki, S. (2014). Enhancing semantic congruity effects with category-contingent comparative judgments. *Frontiers in psychology*, 5.

Link, S. (1990). Modeling imageless thought: The relative judgment theory of numerical comparisons. *Journal of Mathematical Psychology*, 34(1), 2–41.

Link, S. & Heath, R. (1975). A sequential theory of psychological discrimination. *Psychometrika*, 40(1), 77–105.

Link, S. W. (1975). The relative judgment theory of two choice response time. *Journal of Mathematical Psychology*, 12(1), 114–135.

Liu, C. C. & Watanabe, T. (2012). Accounting for speed–accuracy tradeoff in perceptual learning. *Vision research*, 61, 107–114.

Liu, Z. & Weinshall, D. (2000). Mechanisms of generalization in perceptual learning. *Vision research*, 40(1), 97–109.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: a test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 883.

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (1999). Autism diagnostic observation schedule (ados) manual. *Los Angeles, CA: Western Psychological Services*.

Louie, K. (2013). Integrating salience and value in decision making. *Proceedings of the National Academy of Sciences*, 110(40), 15853–15854.

Louie, K. & Glimcher, P. W. (2012). Efficient coding and the neural representation of value. *Annals of the New York Academy of Sciences*, 1251(1), 13–32.

Louie, K., Grattan, L. E., & Glimcher, P. W. (2011). Reward value-based gain control: divisive normalization in parietal cortex. *The Journal of Neuroscience*, 31(29), 10627–10639.

Lu, Z.-L. & Dosher, B. A. (2004). Perceptual learning retunes the perceptual template in foveal orientation identification. *Journal of Vision*, 4(1), 5–5.

Marks, D. F. (1972). Relative judgment: A phenomenon and a theory. *Perception & Psychophysics*, 11(2), 156–160.

Marschark, M. & Paivio, A. (1979). Semantic congruity and lexical marking in symbolic comparisons: An expectancy hypothesis. *Memory & Cognition*, 7(3), 175–184.

Metin, B., Roeyers, H., Wiersema, J. R., van der Meere, J. J., Thompson, M., & Sonuga-Barke, E. (2013). Adhd performance reflects inefficient but not impulsive information processing: A diffusion model analysis. *Neuropsychology*, 27(2), 193.

Milne, E., Swettenham, J., & Campbell, R. (2005). Motion perception and autistic spectrum disorder: a review. *Current Psychology of Cognition*, 23(1/2), 3.

Milne, E., Swettenham, J., Hansen, P., Campbell, R., Jeffries, H., & Plaisted, K. (2002). High motion coherence thresholds in children with autism. *Journal of Child Psychology and Psychiatry*, 43(2), 255–263.

Milosavljevic, M., Malmaud, J., Huth, A., Koch, C., & Rangel, A. (2010). The drift diffusion model can account for value-based choice response times under high and low time pressure. *Judgement & Decision Making*, 5, 437–449.

Ming, X., Brimacombe, M., & Wagner, G. C. (2007). Prevalence of motor impairment in autism spectrum disorders. *Brain and Development*, 29(9), 565–570.

Moyer, R. S. & Bayer, R. H. (1976). Mental comparison and the symbolic distance effect. *Cognitive Psychology*, 8(2), 228–246.

Mozer, M. C., Colagrosso, M. D., & Huber, D. E. (2001). A rational analysis of cognitive control in a speeded discrimination task. In *Advances in Neural Information Processing Systems* (pp. 51–57).

Neumann, J. v., Morgenstern, O., et al. (1944). *Theory of games and economic behavior*, volume 60. Princeton university press Princeton.

Newell, A. & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive skills and their acquisition*, 1, 1–55.

O'Reilly, R. C. & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. MIT press.

O'Riordan, M. A., Plaisted, K. C., Driver, J., & Baron-Cohen, S. (2001). Superior visual search in autism. *Journal of Experimental Psychology: Human Perception and Performance*, 27(3), 719.

Pachella, R. G. (1973). *The interpretation of reaction time in information processing research*. Technical report, DTIC Document.

Padoa-Schioppa, C. (2011). Neurobiology of economic choice: a good-based model. *Annual review of neuroscience*, 34, 333.

Padoa-Schioppa, C. & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, 441(7090), 223–226.

Pais, D., Hogan, P. M., Schlegel, T., Franks, N. R., Leonard, N. E., & Marshall, J. A. R. (2013). A mechanism for value-sensitive decision-making. *PloS one*, 8(9), e73216.

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of vision*, 5(5), 1.

Pastor-Bernier, A., Tremblay, E., & Cisek, P. (2012). Dorsal premotor cortex is involved in switching motor plans. *Volitional inhibition: the gateway for an efficient control of voluntary movements*, (pp.6).

Peirce, J. W. (2009). Generating stimuli for neuroscience using psychopy. *Frontiers in Neuroinformatics*, 2, 10.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, 10(4), 437–442.

Pellicano, E. & Burr, D. (2012). When the world becomes 'too real': a bayesian explanation of autistic perception. *Trends in cognitive sciences*, 16(10), 504–510.

Pellicano, E., Gibson, L., Maybery, M., Durkin, K., & Badcock, D. R. (2005). Abnormal global processing along the dorsal visual pathway in autism: a possible mechanism for weak visuospatial coherence? *Neuropsychologia*, 43(7), 1044–1053.

Perkins, T. J. & Swain, P. S. (2009). Strategies for cellular decision-making. *Molecular Systems Biology*, 5(1).

Petrov, A. A., Van Horn, N. M., & Ratcliff, R. (2011). Dissociable perceptual-learning mechanisms revealed by diffusion-model analysis. *Psychonomic bulletin & review*, 18(3), 490–497.

Petrusic, W. M. (1992). Semantic congruity effects and theories of the comparison process. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4), 962.

Petrusic, W. M., Baranski, J. V., & Kennedy, R. (1998). Similarity comparisons with remembered and perceived magnitudes: Memory psychophysics and fundamental measurement. *Memory & cognition*, 26(5), 1041–1055.

Petrusic, W. M., Shaki, S., & Leth-Steensen, C. (2008). Remembered instructions with symbolic and perceptual comparisons. *Perception & psychophysics*, 70(2), 179–189.

Philiastides, M. G., Biele, G., & Heekeren, H. R. (2010). A mechanistic account of value computation in the human brain. *Proceedings of the National Academy of Sciences*, 107(20), 9430–9435.

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., Dehaene, S., & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, 116(1), 33–41.

Piazza, M., Fumarola, A., Chinello, A., & Melcher, D. (2011). Subitizing reflects visuo-spatial object individuation capacity. *Cognition*, 121(1), 147–153.

Pirrone, A., Dickinson, A., Gomez, R., Stafford, T., & Milne, E. (2016). Understanding perceptual judgement in autism using the drift diffusion model. *Neuropsychology*.

Pirrone, A., Stafford, T., & Marshall, J. A. (2014). When natural selection should optimize speed-accuracy trade-offs. *Frontiers in neuroscience*, 8.

Platt, M. L. & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, 400(6741), 233–238.

Polderman, T., Hoekstra, R., Posthuma, D., & Larsson, H. (2014). The co-occurrence of autistic and adhd dimensions in adults: an etiological study in 17 770 twins. *Translational psychiatry*, 4(9), e435.

Poltrock, S. E. (1989). A random walk model of digit comparison. *Journal of Mathematical Psychology*, 33(2), 131–162.

Provost, A. & Heathcote, A. (2015). Titrating decision processes in the mental rotation task. *Psychological review*, 122(4), 735.

Qiu, A., Adler, M., Crocetti, D., Miller, M. I., & Mostofsky, S. H. (2010). Basal ganglia shapes predict social, communication, and motor dysfunctions in boys with autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 49(6), 539–551.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological methodology*, 25, 111–164.

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, 9(7), 545–556.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59.

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.

Ratcliff, R., Gomez, P., & McKoon, G. (2004a). A diffusion model account of the lexical decision task. *Psychological Review*, 111(1), 159.

Ratcliff, R. & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological review*, 95(3), 385.

Ratcliff, R. & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, 20(4), 873–922.

Ratcliff, R. & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9(5), 347–356.

Ratcliff, R. & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology*, 26(1), 127.

Ratcliff, R. & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111(2), 333.

Ratcliff, R., Spieler, D., & Mckoon, G. (2000). Explicitly modeling the effects of aging on response time. *Psychonomic Bulletin & Review*, 7(1), 1–25.

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004b). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and aging*, 19(2), 278.

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.

Ratcliff, R., Thapar, A., & McKoon, G. (2006). Aging and individual differences in rapid two-choice decisions. *Psychonomic Bulletin & Review*, 13(4), 626–635.

Ratcliff, R., Thapar, A., & McKoon, G. (2007). Application of the diffusion model to two-choice tasks for adults 75-90 years old. *Psychology and aging*, 22(1), 56.

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and iq in two-choice tasks. *Cognitive psychology*, 60(3), 127–157.

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), 438–481.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological review*, 106(2), 261.

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008a). Does subitizing reflect numerical estimation? *Psychological Science*, 19(6), 607–614.

Revkin, S. K., Piazza, M., Izard, V., Zamarian, L., Karner, E., & Delazer, M. (2008b). Verbal numerosity estimation deficit in the context of spared semantic representation of numbers: A neuropsychological study of a patient with frontal lesions. *Neuropsychologia*, 46(10), 2463–2475.

Rorie, A. E., Gao, J., McClelland, J. L., & Newsome, W. T. (2010). Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (lip) of the macaque monkey. *PLoS One*, 5(2), e9308.

Ross, H. E. & Brodie, E. E. (1987). Weber fractions for weight and mass as a function of stimulus intensity. *The Quarterly Journal of Experimental Psychology*, 39(1), 77–88.

Ryalls, B. O., Winslow, E., & Smith, L. B. (1998). A semantic congruity effect in children's acquisition of high and low. *Journal of Memory and Language*, 39(4), 543–557.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological review*, 103(3), 403.

Schaefer, H. M., McGraw, K., & Catoni, C. (2008). Birds use fruit colour as honest signal of dietary antioxidant rewards. *Functional Ecology*, 22(2), 303–310.

Sears, L. L., Vest, C., Mohamed, S., Bailey, J., Ranson, B. J., & Piven, J. (1999). An mri study of the basal ganglia in autism. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 23(4), 613–624.

Seeley, T. D. (2010). *Honeybee democracy*. Princeton University Press.

Seeley, T. D., Visscher, P. K., Schlegel, T., Hogan, P. M., Franks, N. R., & Marshall, J. A. R. (2012). Stop signals provide cross inhibition in collective decision-making by honeybee swarms. *Science*, 335(6064), 108–111.

Seitz, A. R. & Watanabe, T. (2008). Is task-irrelevant learning really task-irrelevant? *PloS one*, 3(11), e3792.

Seli, P., Jonker, T. R., Cheyne, J. A., & Smilek, D. (2013). Enhancing sart validity by statistically controlling speed-accuracy trade-offs. *Frontiers in psychology*, 4, 265.

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), 1916–1936.

Sigman, M. & Gilbert, C. (2000). Learning to find a shape. *Nature neuroscience*, 3(3), 264–269.

Simmons, D. R., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P., & Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision research*, 49(22), 2705–2739.

Smith, P. L. & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, 27(3), 161–168.

Stone, J. V. (2014). Using reaction times and binary responses to estimate psychophysical performance: an information theoretic analysis. *Frontiers in neuroscience*, 8, 35.

Strait, C. E., Blanchard, T. C., & Hayden, B. Y. (2014). Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. *Neuron*, 82(6), 1357–1366.

Strait, C. E., Sleezer, B. J., Blanchard, T. C., Azab, H., Castagno, M. D., & Hayden, B. Y. (2016). Neuronal selectivity for spatial positions of offers and choices in five reward regions. *Journal of neurophysiology*, 115(3), 1098–1111.

Strait, C. E., Sleezer, B. J., & Hayden, B. Y. (2015). Signatures of value comparison in ventral striatum neurons. *PLoS Biol*, 13(6), e1002173.

Stram, D. O. & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, (pp. 1171–1177).

Stroeymeyt, N., Jordan, C., Mayer, G., Hovsepian, S., Giurfa, M., & Franks, N. (2014). Seasonality in communication and collective decision-making in ants. *Proceedings of the Royal Society B: Biological Sciences*, 281(1780), 20133108.

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *science*, 304(5678), 1782–1787.

Tajima, S., Drugowitsch, J., & Pouget, A. (2016). Optimal policy for value-based decision-making. *Nature Communications*, 7.

Teodorescu, A. R., Moran, R., & Usher, M. (2015). Absolutely relative or relatively absolute: violations of value invariance in human decision making. *Psychonomic Bulletin & Review*, (pp. 1–17).

Teodorescu, A. R. & Usher, M. (2013). Disentangling decision models: From independence to competition. *Psychological review*, 120(1), 1.

Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging*, 18(3), 415.

Thura, D., Beauregard-Racine, J., Fradet, C.-W., & Cisek, P. (2012). Decision making by urgency gating: theory and experimental support. *Journal of Neurophysiology*, 108(11), 2912–2930.

Todd, P. M. & Gigerenzer, G. (2003). Bounding rationality to the world. *Journal of Economic Psychology*, 24(2), 143–165.

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.

Townsend, J. T. & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. CUP Archive.

Trimmer, P. C., Houston, A. I., Marshall, J. A. R., Bogacz, R., Paul, E. S., Mendl, M. T., & McNamara, J. M. (2008). Mammalian choices: combining fast-but-inaccurate and slow-but-accurate decision-making systems. *Proceedings of the Royal Society B: Biological Sciences*, 275(1649), 2353–2361.

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences*, 109(24), 9659–9664.

Turner, M. (1999). Annotation: Repetitive behaviour in autism: A review of psychological research. *Journal of child psychology and psychiatry*, 40(6), 839–849.

Usher, M. & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review*, 108(3), 550.

Vandekerckhove, J. & Tuerlinckx, F. (2007). Fitting the ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review*, 14(6), 1011–1026.

Vandekerckhove, J. & Tuerlinckx, F. (2008). Diffusion model analysis with matlab: A dmat primer. *Behavior Research Methods*, 40(1), 61–72.

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, 13(1), 37–58.

Von Neumann, J. & Morgenstern, O. (2007). *Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)*. Princeton university press.

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, 32(7), 1206–1220.

Wagenmakers, E.-J., Van Der Maas, H. L., & Grasman, R. P. (2007). An ez-diffusion model for response time and accuracy. *Psychonomic bulletin & review*, 14(1), 3–22.

Wald, A. (1947). *Sequential Analysis*. Wiley and Sons.

Wald, A. & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 19(3), 326–339.

Wallis, C. & Audley, R. (1964). Response instructions and the speed of relative judgements. *British Journal of Psychology*, 55(2), 121–132.

Watanabe, T., Náñez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature*, 413(6858), 844–848.

Watt, N., Wetherby, A. M., Barber, A., & Morgan, L. (2008). Repetitive and stereotyped behaviors in children with autism spectrum disorders in the second year of life. *Journal of autism and developmental disorders*, 38(8), 1518–1533.

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence*. Psychological Corporation.

White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology*, 54(1), 39–52.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41(1), 67–85.

Yi, D.-J., Olson, I. R., & Chun, M. M. (2006). Shape-specific perceptual learning in a figure-ground segregation task. *Vision research*, 46(6), 914–924.

Zchaluk, K. & Foster, D. H. (2009). Model-free estimation of the psychometric function. *Attention, Perception, & Psychophysics*, 71(6), 1414–1425.

This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Egenolff-Berner Garamond, a revival of Claude Garamont's humanist typeface. A template that can be used to format a PhD thesis with this look and feel has been released under the permissive MIT (xii) license, and can be found online at github.com/suchow/Dissertate or from its author, Jordan Suchow, at suchow@post.harvard.edu.