

The Dynamics of Conflict in Reasoning

by

Eoin Travers, BSc

Supervisors: Dr Aidan Feeney & Dr Jonathan Rolison

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY (PhD)

SCHOOL OF PSYCHOLOGY

QUEEN'S UNIVERSITY BELFAST

MAY, 2016

© EOIN TRAVERS

ALL RIGHTS RESERVED, 2016

Abstract

This thesis explores conflict during reasoning, using a mouse tracking paradigm that provides a measure of participants' instantaneous attraction towards competing response options. It focuses on two kinds of conflict: conflict between competing sources of information in inductive reasoning, and conflict between fast, automatic Type 1 processes and slower, deliberate Type 2 process in dual process accounts of reasoning. Going beyond traditional analyses of participants' responses, the mouse tracking data reveal under what circumstances conflict occurs, and at what points in time participants were influenced by different factors, as well as something of the qualitative nature of this conflict. Chapter 1 reviews previous work on conflict in reasoning, and methods used in the past to study conflict in cognition. Chapter 2 introduces the details of the mouse tracking paradigm used in subsequent chapters.

Chapters 3 and 4 explore conflict between different sources of information. In Chapter 3, two experiments are presented that pit perceptual cues (visual similarity) against conceptual knowledge (shared category membership) in an inductive reasoning task, using both real and artificial stimuli. This reveals that participants are initially driven by perceptual cues, and only later draw on conceptual knowledge, and also that they are more likely to draw on this knowledge when the properties being reasoned about were related to the distinction between the categories. In Chapter 4, two further experiments are presented that pit associative and structured knowledge against each other in an induction task. Participants were again influenced by both kinds of information, but it was less clear to what extent both forms of knowledge interacted on a single trial.

Chapters 5 and 6 explore conflict in dual process theories of reasoning. Chapter 5 reports a judgement task where participants attempted to categorise people,

and could rely on either descriptions of them or on background base rate statistics. Consistent with a default-interventionist dual process model, participants were strongly influenced by descriptions from early in reasoning, but only drew on statistical information, and sometimes overrode description-cued responses on the basis of it, later on. Chapter 6 presents a variation the mouse tracking paradigm, as participants' cursor movements over up to a minute were analysed as they chose between response options on a task where Type 1 processes cue an intuitively appealing incorrect response, the Cognitive Reflection Test. Results here are again consistent with a default-interventionist perspective, as movements were apparently initially driven by Type 1 processes, and correct responses involving an initial attraction towards the heuristic option.

Chapter 7 discusses the implications of these findings, both for theories of reasoning and for accounts of conflict in cognition. It also explores how the methods used here may prove fruitful in future reasoning research.

Contents

Abstract	iii
Contents	vii
List of Figures	x
List of Tables	xi
Acknowledgments	xii
1 Conflict in Reasoning	1
1.1 Introduction	1
1.2 Knowledge and Reasoning	2
1.3 Dual Processes in Reasoning	9
1.4 Measuring conflict	16
1.5 Thesis Overview	26
2 Methods	28
2.1 Mouse Tracking	28
2.2 Analysing Mouse Data	33
2.3 Statistical Analyses	42
3 Perceptual Similarity vs. Conceptual Knowledge in Induction	48
3.1 Introduction	48
3.2 Experiment 1	52
3.3 Experiment 2	63
3.4 General Discussion	73

4	Associative vs. Structured Knowledge in Induction	76
4.1	Introduction	76
4.2	Experiment 3	80
4.3	Experiment 4	92
4.4	General Discussion	108
5	Descriptions vs. Statistics in Reasoning	110
5.1	Introduction	110
5.2	Method	118
5.3	Results	121
5.4	Discussion	135
6	Intuition vs. Reflection in Reasoning	143
6.1	Introduction	143
6.2	Method	148
6.3	Results	150
6.4	Discussion	161
7	Summary, Discussion and Final Conclusions	165
7.1	Recapitulation	165
7.2	Conflict in Reasoning	168
7.3	Knowledge and Reasoning	171
7.4	Dual Processes and Reasoning	175
7.5	Conflict more broadly	179
7.6	Conclusions	186
	References	220
	Appendices	220
A	Continuous Attraction and Changes of Mind in Mouse Tracking Studies .	222
B	Classifying Cursor Trajectories	225
C	Stimuli from Experiment 1	231
D	Post test scores for stimuli used in Experiment 1	233
E	Stimuli from Experiment 2	234

F	Properties used in Experiment 2	235
G	Association ratings for stimuli used in Experiment 3	237
H	Stimuli used in Experiment 3	238
I	Post test performance for stimuli used in Experiment 3	240
J	Stimuli used in Experiment 4	241
K	Post test performance for stimuli used in Experiment 4	244
L	Bayesian model priors from Experiment 5	245
M	CRT Reasoning Problems from Experiment 6	246

List of Figures

1.1	Screen shot from the triad task used by Bright and Feeney (submitted).	7
1.2	The Stanovich and West (2008) framework for individual differences in reasoning.	14
1.3	The dual process framework proposed by Pennycook, Fugelsang, and Koehler (2015).	16
1.4	The mouse tracking paradigm used by Spivey, Grosjean, and Knoblich (2005)	24
2.1	Spivey et al.'s (2005) mouse tracking paradigm.	29
2.2	Summary measures for cursor trajectories.	34
2.3	Trajectories classed as direct and reversals in Experiment 2.	38
2.4	Parameters for descriptions cursor transitions.	40
2.5	The pitfalls of averaging cursor trajectories.	41
3.1	Screen shot of conflict trial from Experiment 1.	54
3.2	Screen shot from the post-test check, Experiment 1.	56
3.3	Proportion of foil responses given by each participant, per condition, in Experiment 1.	57
3.4	The possible cursor transitions that can occur during a trial.	58
3.5	Time course, separately for each response option, in Experiment 1.	60
3.6	Time course, separately for each condition, in Experiment 1.	60
3.7	Screen shot from Experiment 2.	65
3.8	Proportion of foil responses given by each participant, per condition, in Experiment 2.	67
3.9	Time course of attraction towards the foil, in Experiment 2.	71

3.10 Time course for conflict trials, in Experiment 2	71
4.1 A conflict trial from Bright and Feeney (submitted).	80
4.2 A screen shot from Experiment 3.	82
4.3 Participants' number of foil responses, by condition, Experiment 3. .	86
4.4 The possible transitions that can occur during a trial.	87
4.5 Time course, separately for each condition, Experiment 3.	89
4.6 Time course, separately for each response option, Experiment 3. . .	89
4.7 The log-transformed association ratio, used as a predictor in Experiment 4.	96
4.8 Effect of the association ratio of participants' accuracy, Experiment 4. .	97
4.9 Proportion of correct responses given by each participant, by association ratio, in Experiment 4	98
4.10 Proportion of correct responses that involved reversals, Experiment 4. .	100
4.11 Transition probabilities, as a function of the association ratio, in Experiment 4.	102
4.12 Time course, separately for each response option, in Experiment 4. .	103
4.13 Time course, separately for different association ratios, in Experiment 4.	104
5.1 A screen shot from Experiment 5.	121
5.2 Individual differences in the effect of descriptions on participants' responses, Experiment 5.	123
5.3 Individual differences in the effect of base rates on participants' responses, Experiment 5.	124
5.4 Bayesian analysis of effect of the base rate on rt for description-based responses, Experiment 5.	130
5.5 The possible transitions that can occur during a trial where the base rate cues a response.	131
5.6 The possible transitions that can occur during a trial where the description cues a response.	132
5.7 Time course plots for the effect of descriptions on base rate-cued choices, and vice versa, in Experiment 5.	134

6.1	A screen shot from the CRT, Experiment 6.	149
6.2	A typical cursor trajectory from Experiment 6.	152
6.3	Proportion of mouse cursors each region of the screen over time, Experiment 6.	153
6.4	Change point analysis, Experiment 6.	159
6.5	Proportion of cursor in region of chosen response option, Experiment 6.	159
6.6	Proportion of cursor in region of other response options when correct response was given, Experiment 6.	160
6.7	Proportion of cursor in region of other response options when heuristic response was given, Experiment 6.	160
7.1	A framework for conflict in reasoning	172

List of Tables

1.1	Predictions about conflict from various dual process theories	13
3.1	Descriptive statistics for correct responses, Experiment 1.	56
3.2	Kinds of trajectory in Experiment 1.	57
3.3	Transition probabilities for Experiment 1.	58
3.4	Foil responses in Experiment 2.	66
3.5	Descriptive statistics for correct responses, Experiment 2.	66
3.6	Kinds of trajectory in Experiment 2.	67
3.7	Transition probabilities for Experiment 2.	68
4.1	The procedure followed on reasoning trials in Experiments 3 and 4. .	84
4.2	Kinds of trajectory in Experiment 3.	87
4.3	Transition probabilities for Experiment 3.	87
4.4	Divergence times, Experiment 4.	105
5.1	Examples problems from Experiment 5.	120
5.2	Participants' responses, Experiment 5.	122
5.3	Descriptive statistics for base rate-cued responses, Experiment 5. .	125
5.4	Effects of manipulating description on base rate-cued responses, Experiment 5.	126
5.5	Descriptive statistics for description-cued responses, Experiment 5. .	127
5.6	Effect of descriptions on transition probabilities, Experiment 5. . .	131
5.7	Effect of base rates on transition probabilities, Experiment 5.	132
5.8	Prevous base rate neglect studies.	141
6.1	Bayesian change point analysis, Experiment 6.	155

Acknowledgments

This thesis would not have been possible without the ideas, guidance, and patience of my supervisors, Aidan Feeney and Jonathan Rolison. I thank them for keeping me on message whenever I went off, and keeping me inspired whenever I was frustrated. I also thank the postgraduate community at QUB — they know who they are — for listening to me complain in the understanding that I would listen to them do the same. Thanks are also due to my family and friends in the real world, for their support, and for pretending to be interested in my research. Finally, my heartfelt thanks go to Orla, who got stuck with me as an undergraduate, put up with me as a graduate student, and hopefully gets the reward of listening to me as a doctor. Thank you.

CHAPTER 1

Conflict in Reasoning

1.1 INTRODUCTION

Reasoning is not a simple task. Below the surface of everyday judgement, decision making, and high-level thinking, there are a range of underlying processes, many of which we are only beginning to understand. Unsurprisingly, amid the constant interaction of these processes, there are many points at which conflict arises in reasoning, and in cognition more broadly. Much previous work (M. M. Botvinick, Cohen, & Carter, 2004; Spivey, 2007; van Veen & Carter, 2006) has addressed the role of conflict within and between individual low-level cognitive processes. In this thesis, I focus on conflict at a higher-level: conflict in reasoning.

I focus on two major points at which conflict is thought to arise during reasoning. First, many forms of reasoning, and inductive reasoning in particular, must draw on various sources of information, including new information, and prior knowledge. Conflict can arise when these different sources of information suggest different inferences or actions. Second, many problems are thought to be attempted using either fast, automatic, effortless *Type 1* processes, or using slower, deliberate, effortful *Type 2* processes, and an extensive *dual processes* tradition focuses on the characteristics of these processes and how they interact. Here, conflict arises when multiple responses are cued simultaneously. This thesis presents four experimental chapters, two studying conflict between mental representations in induction, and two studying conflict between Type 1 and Type 2 processes in reasoning.

In most experimental studies of reasoning, researchers record and analyse only participants' responses, or their responses and their response times. This approach has produced considerable insights, but it also has limitations. In particular, it is difficult to infer what processes actually drive reasoning by analysing only responses which are recorded after these processes have ended. Therefore, to study conflict in reasoning, I use a mouse tracking paradigm. This technique, popular in the studies of conflict in simpler cognition, involves recording the position of the computer mouse cursor as participants decide between alternative response options. By analysing these cursor trajectories, it is possible to infer to what extent participants are drawn towards each option, over time. In particular, this method allows me to infer two things in this thesis: what responses were participants drawn towards giving *before* giving their ultimate response, and at what points in time were they drawn towards each response.

The remainder of this first chapter is organised as follows. First, I discuss previous work on the different kinds of information that can be drawn on in inductive reasoning, and the possibility of conflict between them. Next, I introduce dual process theories of cognition, and review recent work on the nature of conflict within such theories. Lastly, I present an overview of methods used to study conflict in cognition, and introduce the mouse tracking paradigm used throughout this thesis.

1.2 KNOWLEDGE AND REASONING

In inductive reasoning, we make inferences that go beyond what we already know to reach conclusions that, while not guaranteed to be correct, can often be useful (Bruner, 1973). To do this, we must often make use of information from various sources, including both knowledge retrieved from memory — which can come in many forms — and information inherent to the problem at hand. Thus, inductive reasoning may involve conflict not between two qualitatively different kinds of cognitive processing, as in the dual process theories discussed below, but between various kinds of information that could be used as the basis for our inferences.

In this thesis, I focus in particular on category-based induction (CBI), where properties we know to be possessed by members of some categories are generalised, or *projected*, to other, related categories (Rips, 1975). For instance, on learning

that carrots have a property, we may infer that rabbits (who eat carrots), or possibly parsnips (biologically similar to carrots, and used in similar contexts), could also have this property. Studies of CBI typically use categories such as these from the natural world, which are well-known to participants, and can be related in a variety of ways.

1.2.1 TYPES OF KNOWLEDGE IN CATEGORY-BASED INDUCTION

Over the years, many theories of category-based induction have been proposed (see Feeney & Heit, 2007, and Hayes, Heit, & Swendsen, 2010, for reviews). Naturally, these theories differ in terms of what kinds of knowledge they claim people draw on during CBI. Following Bright and Feeney (2014b), I divide these theories into two classes: those based on *structured* knowledge, and those based on *associative* knowledge.

In theories based on structured knowledge (Griffiths & Tenenbaum, 2009; Kemp & Tenenbaum, 2003, 2009; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Shafto, Kemp, Bonawitz, Coley, & Tenenbaum, 2008; Tenenbaum, Griffiths, & Kemp, 2006), inferences are guided by information about the categories to which things belong, and the relationships between various categories. The simplest such structure is category membership itself: learning a property of some category members may lead us to project that property to other members of that category, or, in other words, to generalise it. More sophisticated models are based on the relationships between categories. For instance, Osherson et al.'s (1990) similarity-coverage model represents species of plant and animal as branches on a taxonomic tree, and predicts generalisation of a property to a new species as a function of both the similarity between species that do have the property and the new species, and the extent to which those species cover the lowest super-ordinate category that encompasses all of them as well as the new species.

However, this taxonomic structure is not appropriate for all inferences, and more recent accounts (Griffiths & Tenenbaum, 2009; Kemp & Tenenbaum, 2009; Shafto et al., 2008; Tenenbaum et al., 2006) generalise this approach using a variety of knowledge structures. These accounts make use of *Bayes nets* – systems of structured, probabilistic links – to represent knowledge about the relations between cat-

egories in specific domains, and the mechanisms by which properties can be transmitted between them. In this framework, CBI is seen as a process of updating beliefs about the distribution of a given property, given evidence about which species have or do not have this property.

Other theories are based instead on what Bright and Feeney (2014b) term *associative* knowledge (e.g. Rogers & McClelland, 2004; Sloman, 1993; Sloutsky & Fisher, 2004a). Similarity, or the presence of shared features, is the simplest form of associative knowledge: things that are similar in ways we know about (they share many features, or properties) are more likely to share other, unobserved features than things that are dissimilar. This idea forms the basis of the Sloman's (1993) Feature Based Induction model, and Rogers and McClelland's (2004) theory of semantic cognition. Sloutsky and Fisher (2004a) also propose a similarity-based account of children's induction. In their account, children's inferences are based on *visual similarity*: things that look alike are judged more likely to share a property than things that do not. Fisher, Godwin, and Matlen (2015) note that similarity-based models of induction in adults are based on overlapping features in our mental representations of various categories (*representational similarity*), while Sloutsky and Fisher's (2004a) developmental account is based on visual *perceptual similarity* between them.

Other forms of associative knowledge used in theories of induction appeal to factors such as temporal contiguity, or co-occurrence (Gluck & Bower, 1988; Rescorla & Furrow, 1977): things often encountered together are more likely to share properties. What these kinds of representation have in common is that they do not explicitly encode the ways in which categories are related, only the association between various categories. Recently, Jackson, Hoffman, Pobric, and Ralph (2015) have proposed a framework combining both representations of the features of categories (i.e. BISCUIT-CRUMB; and thus the similarity, or featural overlap, between categories) and relational associations between categories (i.e. BISCUIT-TEA). According to their proposal, both forms of knowledge are represented in the core semantic system, located in the anterior temporal lobe, superior temporal sulcus, and ventral prefrontal cortex.

As we have seen, the kind of knowledge underlying adults' inductive reason-

ing remains an active area of study. That said, common to all of the above accounts is the idea that induction makes use of some kind of prior knowledge. However, in research on children's inductive reasoning, there remains some debate as to whether young children rely on conceptual knowledge at all, or if they instead draw conclusions based on non-conceptual information, such as perceptual similarity between two entities. While this thesis focuses on adults' reasoning, this developmental debate may have implications for theories of reasoning in adults.

On the one hand, young children possess considerably less conceptual knowledge than adults, and so they may need to rely on external information, most commonly perceptual similarity, to make sense of the world and to reason inductively (Sloutsky, 2003; Sloutsky & Fisher, 2004a). On the other hand, it has been argued that even at the age of four children are capable of reasoning on the basis of conceptual knowledge such as shared category membership (S. A. Gelman, 2004; S. A. Gelman & Meyer, 2011). Experimental data here are somewhat ambiguous. Some experiments suggest that children as young as 4 do make use of conceptual knowledge in their reasoning (S. A. Gelman & Davidson, 2013; S. A. Gelman & Markman, 1986), while others indicate that at this age, children's inferences are primarily driven by visual similarity (Badger & Shapiro, 2012; Sloutsky, Deng, Fisher, & Kloos, 2015; Sloutsky, Kloos, & Fisher, 2007). This suggests another point at which conflict can occur. If some evidence shows that children rely on one form of information, and some evidence that they rely on another, it may be that these two forms of information come into conflict. Of course, this thesis focuses on adults' reasoning, where little attention is given to the role of perceptual cues. However, adults' high level thinking is often based upon simpler cognitive processes acquired while young (Barsalou, Breazeal, & Smith, 2007; Vygotsky, 1980). Therefore, it may be worthwhile to look for known developmental effects in adults' cognition (see, for instance, Falke, Winter, & Spivey, 2013, for signs of A-not-B error, usually thought to be overcome by 12 months, in adults). I discuss this debate in detail in Chapter 3.

1.2.2 A HYBRID MODEL OF INDUCTION

In the theories of induction discussed above, researchers have argued that inductive relies on particular kind of information, be it probabilistic relational networks (Kemp & Tenenbaum, 2009), a taxonomic tree (Osherson et al., 1990), conceptual similarity (Sloman, 1993), or even perceptual similarity (Sloutsky et al., 2007). Recently, however, Bright and Feeney (2014b) have raised the possibility that multiple kinds of information may underlie inductive reasoning. Clearly, if more than one kind of information can be drawn on in induction, there is potential for conflict to occur, and so this idea is an important one for this thesis.

In their *hybrid* theory of induction, Bright and Feeney (2014b; see also Crisp-Bright, 2010) proposed that neither theories of induction based on associative knowledge such as similarity (Sloman, 1993), nor those based on structured relational knowledge (e.g. Kemp & Tenenbaum, 2009; Osherson et al., 1990) were in themselves complete. Instead, they argued that both kinds of knowledge may play a role in induction. In the hybrid model, associative knowledge is thought to be available early in the reasoning process, and require less cognitive resources to utilise than structured knowledge. While usually reliable, inferences from associative knowledge may be fallacious in many circumstances, and associative knowledge is insufficient to capture some of the flexibility and context sensitivity that characterises human reasoning (Heit & Rubinstein, 1994; Murphy & Medin, 1985).

Support for this theory comes from experiments where the strength of the association between categories, and participants' beliefs about the structured relations between them, have been established using pretests. Bright and Feeney (2014b) showed that strength of association predicts ratings of the strength of inductive arguments made under all circumstances. The influence of structured relations such as taxonomic or causal knowledge, however, is diminished when participants are placed under time pressure, or heavy cognitive load. Bright and Feeney (2014b) further showed that when given a property of a base species and asked to name another species likely to share that property, participants under load were more likely to name a species strongly associated with the base, rather than one related according to the appropriate structured knowledge, than were those completing the task without secondary load.

While these findings indicate that both associative and structured knowledge play a role in inductive reasoning, they reveal little about conflict in induction. Luckily, a paradigm exists, primarily used in the developmental literature, which allows researchers to place two kinds of knowledge in direct conflict. In the inductive triad task (S. A. Gelman & Markman, 1986), participants are told a property of one species, and asked which of two other species they believe to be most likely to also have the property (or, alternatively, told properties of two species, and asked which property a third species was most likely to have). In its original form, this task was used to test if young children believed that a target animal would share a property with an animal from a different species that looked alike, or an animal from the same species that looked different.

Typical uses of this paradigm present participants with only trials in which the two kinds of information conflict, and compare participants' performance to chance to claim that participants relied on one or other form of information (see Chapter 3 for a review of these studies). The task has great potential, however, to reveal the presence of conflict between multiple forms of information, if trials in which the two kinds of information conflict are compared to control trials in which they cue the same response. Bright and Feeney (submitted; also Crisp-Bright, 2010, Chapter 5) report such a version of this task, placing associative and structured knowledge in conflict. Figure 1.1 shows a trial from this experiment. Participants were presented with a biological property of a base species, in this case carrots, and asked to chose which of the candidate species, bamboo or rabbits, was more likely to also possess that property. In this case, while structured knowledge would indicate that carrots and bamboo are both plants, and so are more likely to share a biological property, there is a strong association between carrots and the foil option, rabbits, and so participants relying on associative knowledge would give that response instead. On control trials, there was no such inappropriate association between the candidate species and the foil (tigers, in this example). Therefore, an increase in foil responses on conflict trials should be due to the strong association between carrots and rabbits.

While Bright and Feeney's (submitted) participants primarily responded according to structured knowledge, they were less likely to do so if under heavy cog-



Carrots have C5s-cells

Which category is more likely to also have these cells?



Bamboo



Rabbits

Figure 1.1: Screen shot from the triad task used by Bright and Feeney (submitted). Participants learned that carrots had a certain property (they had C5s cells), and where asked what was most likely to also have that property, bamboo, or rabbits.

nitive load, or if lacking in *semantic inhibitory control* (Burgess & Shallice, 1997; Markovits & Doyon, 2004). This last finding, in particular, supports the idea that reasoning in such contexts involves conflict: participants responding associatively were unable to inhibit the association-driven representation in favour of structured knowledge.

The question remains, however, as to how associative and structure knowledge actually interact during reasoning. Later in this chapter, I discuss dual process accounts of reasoning, where responses can be generated by two qualitatively different kinds of process: Type 1, and Type 2 processes. It should be stressed that Bright and Feeney's (2014b) hybrid account is not a dual process theory, in the

sense discussed below. Rather, it proposes that different kinds of knowledge differ in the way in which information is represented, rather than qualitatively in the kinds of process they engage. With that caveat in place, much attention (e.g. Evans, 2007) has been given to the issue of how different processes interact in dual process theories. It is therefore worth considering how existing ideas about how Type 1 and Type 2 processes interact might apply to the interaction of different sources of information in induction.

It may be the case that participants draw on either associative or structured knowledge, for instance, for a given inference. If this is the case, we would not expect to see evidence of conflict between different sources of information on a single trial. Alternatively, it may be that some sources of information are available quickly and easily, and so are accessed by default. In this case, participants may retrieve such information initially, but later inhibit it, or at least attempt to inhibit it, in favour of more useful knowledge, if they realise their initial source of information to be inappropriate. A third possibility is that multiple types of information are brought to bear simultaneously in parallel. However, it is difficult to draw predictions from such an account, and so I will not discuss it further.

This juncture, of course, is not the only point where conflict can arise in reasoning, and in fact the issue of conflict is rarely discussed in the context of induction. However, in recent years the issue of conflict has become the focus of attention elsewhere, by those interested in dual process accounts of reasoning.

1.3 DUAL PROCESSES IN REASONING

From Wason and Evans (1975) onwards, reasoning researchers have considered the possibility that responses on reasoning problems can be generated by one of two qualitatively different kind of process. Versions of this distinction have been proposed many times in many different fields, leading to considerable confusion as to what defines each kind of process (see Evans, 2008; Evans & Stanovich, 2013; Stanovich & Toplak, 2012). Proposed distinctions include those between (from Evans, 2008) input modules and higher cognition (Fodor, 1983), automatic and controlled processes (Schneider & Shiffrin, 1977), heuristic and analytic processes (Evans, 1984, 2006), and associative and rule-based systems (Sloman, 1996), among

many others. More recent accounts (e.g. Evans & Stanovich, 2013) offer a distinction that somewhat simplifies the matter: Type 1 processes are a heterogeneous set of processes, some, but not all, of which are dependent on specific neural subsystems, which operate automatically in the presence of their enabling conditions (that is, they are *autonomous*; Stanovich, 2009a), and do not require conscious effort. Type 2 processes, in contrast, are mental operations that require a) the formation, manipulation, and updating of an abstract representation in working memory, and b) the *decoupling* of this representation from Type 1 processes (see Stanovich & Toplak, 2012). This account can be further simplified without much loss in the central message: Type 2 processes are those which require working memory, while Type 1 processes are those that do not.

It should be noted, however, that dual process accounts are not without controversy. There has been extensive criticism of such accounts both in the reasoning literature (Gigerenzer & Regier, 1996; Kruglanski, 2013; Kruglanski & Gigerenzer, 2011; Osman, 2004) and more generally (e.g. B. R. Newell & Shanks, 2014; Tunney & Shanks, 2003). In particular, Kruglanski and Gigerenzer (2011, also Gigerenzer & Regier, 1996) argue that Type 1 and Type 2 processes both reflect the application of cognitive rules, and that these rules differ only quantitatively in how much mental effect they require, and how consciously accessible they are.

1.3.1 CONFLICT IN DUAL PROCESS THEORIES

My interest in dual process theories in this thesis centres on their potential to explain conflict during reasoning. How and when this conflict occurs depends on how Type 1 and 2 processes interact. A common thread throughout dual processes accounts is that Type 1 processes in many contexts produce responses that are heuristic, biased, and inappropriate, whereas Type 2 processes, when used, typically lead to normatively correct responses. In this sense, dual process theories build on the earlier Heuristics and Biases research tradition (Gilovich, Griffin, & Kahneman, 2002; Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1974). This tradition focused on situations where people often rely on mental shortcuts, or *heuristics* instead of engaging in more effortful decision making, leading to systematic biases in situations where the heuristic does not provide a

useful solution to the problem at hand. More recent treatments of the heuristics and biases literature (Kahneman, 2011; Kahneman & Frederick, 2002, 2005) have explicitly described heuristics as the result of Type 1 processes, while the *reflective* thinking needed to resist these heuristic responses (Frederick, 2005) is the consequence of Type 2 processes (although, again, see Gigerenzer & Gaissmaier, 2011; Kruglanski & Gigerenzer, 2011, for contrasting views).

Evans (2007; see also Gilbert, 1999) argues that there are three possible ways in which Type 1 and Type 2 processes could interact. He labels these *pre-emptive conflict resolution*, *default-interventionist*, and *parallel-competitive* architectures (perhaps more intuitively referred to as *selective*, *corrective*, and *competitive* designs by Gilbert, 1999).

Simplest are selective, pre-emptive conflict resolution models (Chaiken, 1987; Klaczynski, 2000; Klaczynski & Cottrell, 2004; Petty & Cacioppo, 1986), where a given decision cues the activation of either Type 1 or Type 2 processes. Clearly, there is little role for conflict in such accounts, as only one or other kind of process should be active at any one time. Gilbert (1999) notes that such models are popular with researchers interested in predicting overt behaviour rather than inferring underlying processes, and so have proven resilient within social psychology, often as models of persuasion (Chaiken, 1987; Petty & Cacioppo, 1986). Klaczynski (2000; Klaczynski & Cottrell, 2004) proposes an analogous model in the reasoning literature. Although they do not account for conflict during reasoning, selective dual process models provide a parsimonious account of participants' responses on many tasks. In early dual process research, it was these responses that were the focus of analysis: researchers primarily analysed only the choices participants made, such as numerical judgements (e.g. Tversky & Kahneman, 1981), ratings of the persuasiveness of arguments (e.g. Chaiken, 1987), or judgements of validity of logical syllogisms (e.g. Evans, Barston, & Pollard, 1983). Many theories that initially proposed selective activation, however, have subsequently been updated to account for results that suggest conflict or interference, and so have developed into default-interventionist or parallel-competitive accounts. Evans (2007) highlights an additional limitation of these models: they require, but often fail to provide, some means of deciding which process should be engaged.

Corrective, default-interventionist models (Evans, 2006; Kahneman & Frederick, 2002, 2005) provide an alternative account. According to these models — typically the best known in the literature, particularly following the publication of Kahneman's (2011) *Thinking, Fast and Slow* — Type 1 processes are constantly active, cuing responses, generating mental representations, and processing incoming stimuli. The main role of Type 2 processes is to monitor and inspect these intuitive beliefs and responses, and to inhibit or override them if they are found to be inadequate. If the output of a Type 1 process is deemed inappropriate, it must be inhibited, and possibly replaced with the product of more effortful Type 2 thinking. Most theories hold this monitoring processes is typically lax — people are *cognitive misers* (Fiske & Taylor, 1991) — and so under most circumstances we give intuitively-produced responses and engage in intuitively-cued behaviours (Gilovich et al., 2002; Kahneman & Frederick, 2002, 2005; Kahneman et al., 1982; Kahneman & Tversky, 1973; Stanovich & West, 1999; Tversky & Kahneman, 1973, 1974). Conflict, however, occurs when people detect that their intuitive responses are inadequate, and attempt to inhibit or override them.

In competitive, or parallel-competitive models (Darlow & Sloman, 2010; Sloman, 1996, 2002, 2014a, 2014b), both Type 1 and Type 2 processes are thought to be activated simultaneously. An important aspect of Sloman's (1996) model, although also consistent with default-interventionist accounts, is that parallel activation of both kinds of process would allow people to simultaneously hold two potentially contradictory beliefs, one generated by each kind of process (Criterion S; Sloman, 1996). Parallel models differ from their interventionist counterparts in that they would predict conflict to be much more common in cognition. If both Type 1 and Type 2 processes are activated during reasoning, then both may simultaneously cue conflicting responses. When this happens, participants should be conflicted regardless of what response they actually give — they will be partially drawn to the Type 1 response if giving the Type 2 one, and vice versa.

Although not included in Evans' (2007) review, more recent accounts, while still operating within a dual process framework, have questioned the notion that Type 1 processes are always biased, and that Type 2 processes are essential for sound reasoning. First, a number of experiments (e.g. De Neys, Cromheeke, & Osman,

2011; De Neys & Glumicic, 2008; De Neys, Moyens, & Vansteenwegen, 2010; De Neys, Rossi, & Houdé, 2013; De Neys, Vartanian, & Goel, 2008; De Neys et al., 2008; Morsanyi & Handley, 2012) have compared problems where participants predominantly give the incorrect, heuristic response to no-conflict versions where the heuristic response is the correct one. In these experiments, participants rarely explicitly report that they realise their heuristic responses are suspect (De Neys & Glumicic, 2008). However, using more subtle implicit measures, these experiments have shown that participants are sensitive to the conflict between their heuristic responses and normative principles; for instance, on conflict trials participants are slower to respond (De Neys et al., 2008), are less confident in their responses (De Neys et al., 2011, 2013), sweat more (De Neys et al., 2010), show greater activation in the anterior cingulate cortex, a neural region implicated in the detection of conflict (De Neys et al., 2008), and even report “liking” logically valid syllogisms more than equivalent invalid ones (Morsanyi & Handley, 2012). For clarity, I will refer to De Neys’ (e.g. De Neys & Glumicic, 2008) proposal that reasoners who give heuristic responses detect some conflict while doing so as the *dual process conflict monitoring theory*. This account draws on theories of conflict monitoring in simpler cognitive domains (M. M. Botvinick, Braver, Barch, Carter, & Cohen, 2001).

More recent work (i.e. De Neys, 2012, 2014) proposes a stronger interpretation of these results: that Type 1 processes can simultaneously cue both the sometimes incorrect heuristic response and the logically correct response. Clearly, this is a stronger claim than that made by the dual process conflict monitoring theory, and I will reserve the name *intuitive logic* theory for this account. According to this stronger account, participants predominantly give the heuristic response because it is cued more strongly by Type 1 processes than the correct one (i.e. it is *prepotent*). Type 2 processes, in this account, are activated when participants detect a conflict between multiple responses generated by Type 1 processes, and must inhibit the prepotent heuristic response in order for participants’ to produce the correct response. However, according to De Neys and Bonnefon (2013), while participants will usually detect this conflict, and attempt to engage Type 2 processes, they often fail to inhibit the heuristic response, leading to biased reasoning. A similar account is offered by Handley and Trippas (2015), who in addition to arguing

Table 1.1: When do different dual process theories predict that conflict should occur? Note: T1 = Type 1 processes; T2 = Type 2 processes.

	T1 responses	T2 responses
Selective	No	No
Default-Interventionist	No	Yes (inhibit T1 response)
Parallel-Competitive	Yes (T2 interference)	Yes (T1 interference)
Intuitive logic	Yes (multiple T1 responses)	Yes (inhibiting prepotent T1 response)

that Type 1 processes can produce both heuristic and correct responses, claim that Type 2 processes can also be responsible for the generation of biased responses. Finally, Pennycook, Fugelsang, and Koehler (2015) offer a framework that formalises the interaction of multiple Type 1 responses proposed by De Neys (2012, 2014). For the purposes of this thesis, the differences between these theories are not particularly germane. Therefore, I will restrict my discussion for the most part to the intuitive logic theory (De Neys, 2012, 2014), with the understanding that points applicable to it are also relevant to the accounts put forward by Handley and Trippas (2015) and Pennycook, Fugelsang, and Koehler (2015).

Intuitive logic theories, like parallel-competitive accounts, predict that conflict should be common in reasoning. Both of these accounts predict that participants should experience conflict even when they give a heuristic response. The intuitive logic account holds that this happens because both responses are cued by Type 1 processes, and the role of Type 2 processes is to select between candidate responses, and to engage and direct the inhibitory processes needed to hold back the prepotent heuristic response. The parallel-competitive account, on the other hand, holds that Type 2 processes cue the logical response at the same time as Type 1 processes cue the heuristic one.

1.3.2 FACTORS DICTATING THE GENERATION OF TYPE 1/TYPE 2 RESPONSES

So far, I have discussed the circumstances under which various dual process theories would predict that conflict should occur. However, one important question remains: in situations where these processes do conflict, how is this conflict resolved? In this section, I introduce a number of recent frameworks proposed to make sense

of why people sometimes produce responses, generated by Type 1 processes, and other times correct responses, generated by Type 2 processes (but again, see De Neys, 2012; Handley & Trippas, 2015).

Stanovich and West (2008) provide one such framework, in their discussion of individual differences in susceptibility to various heuristics and biases, which is equally applicable to the question of conflict in reasoning. Their framework is based on default-interventionist dual process accounts, where the heuristic, Type 1 response is given unless overridden by Type 2 processes. Firstly, if participants do not know the logical rules required to produce the correct response (the *Mindware*, in the terms of Stanovich & West, 2008), they will not be able to do so. After this, participants must realise that their heuristic response is inadequate, something which some theories (De Neys, 2012) hold to happen automatically, while others (Evans, 2006; Evans & Stanovich, 2013; Kahneman, 2011) believe this requires motivated attention and the engagement of Type 2 processes. Thirdly, some tasks require executive processes to inhibit the heuristic response, and to form and manipulate the necessary *decoupled* representation in working memory in order to reach the logically valid response. On tasks for which this is not necessary, it should be trivially easy for participants who understand the logical principles, and realise their intuitions are mistaken, to produce the logical response. Where it is necessary, however, participants must engage these executive processes in order to respond correctly, and so it is at this point where conflict should occur. Figure 1.2 shows a schematic of this framework.

De Neys and Bonnefon (2013) discuss a similar issue, and highlight three possible “failures” that lead participants to give incorrect, heuristic responses. Their *storage* failure is analogous to Stanovich and West’s (2008) Mindware gap: participants who do not possess the knowledge necessary to respond correctly — the rules of formal logic, or probability theory, for instance — will respond heuristically. Their *monitoring* failure occurs when participants, although possessing the correct knowledge, fail to realise that their heuristic responses are inconsistent with that knowledge. Finally, their *inhibition* failure occurs when participants do detect the inadequacy of their heuristic response, but fail to properly inhibit it.

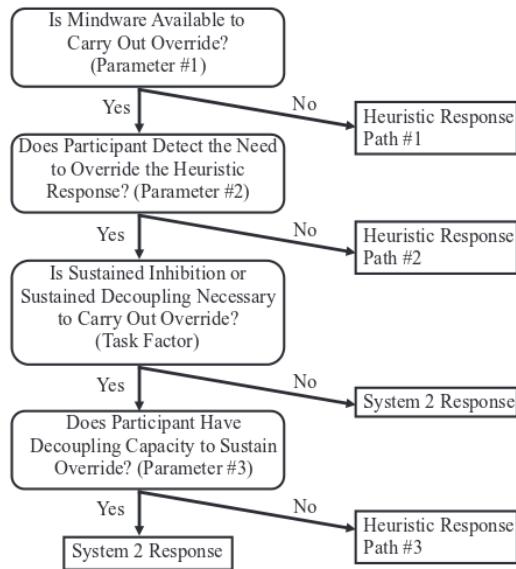


Figure 1.2: Stanovich and West's (2008) framework for individual differences in reasoning. From Stanovich and West (2008).

De Neys and Bonnefon (2013) also draw attention to the issue of *when* individual differences between heuristic and rational reasoners should emerge. Differences due to storage failures should emerge at the very onset of reasoning: participants without the correct knowledge will be biased from the start; those with that knowledge will reason correctly. Such a failure may be analogous to a selective dual process account (Klaczynski & Cottrell, 2004). Differences due to monitoring failures should emerge slightly later, as some participants realise that their heuristic responses, once generated, are incorrect, while others do not. This corresponds to the classical default-interventionist account (Evans, 2006). Lastly, differences due to inhibition failures should emerge latest of all: participants realise that their heuristic responses are inadequate, and attempt to inhibit them, but only some manage to complete this inhibition. This possibility accords most closely with the intuitive logic account (De Neys, 2012).

Finally, Pennycook, Fugelsang, and Koehler (2015) put forward a dual process framework that consolidates aspects of default-interventionist dual process accounts (e.g. Evans, 2006) and the intuitive logic theory (De Neys, 2012). This framework is illustrated in Figure 1.3. In line with the intuitive logic theory, they propose that, when faced with a problem or cue, Type 1 processes can generate

multiple responses, often including the correct one. If participants fail to detect a conflict between these multiple responses (or, presumably, if they only generate one response), then they will give the prepotent, or heuristic response, which is incorrect in many tasks studied in the lab. If they do detect a conflict, participants activate Type 2 processes. With these, they can either a) engage in *rationalisation*, attempting to justify the heuristic response already cued by Type 1 processes, b) inhibit this prepotent response and give one of the other intuitively-produced responses, or c) inhibit the prepotent response and explicitly work out a new response, not cued by Type 1 processes.

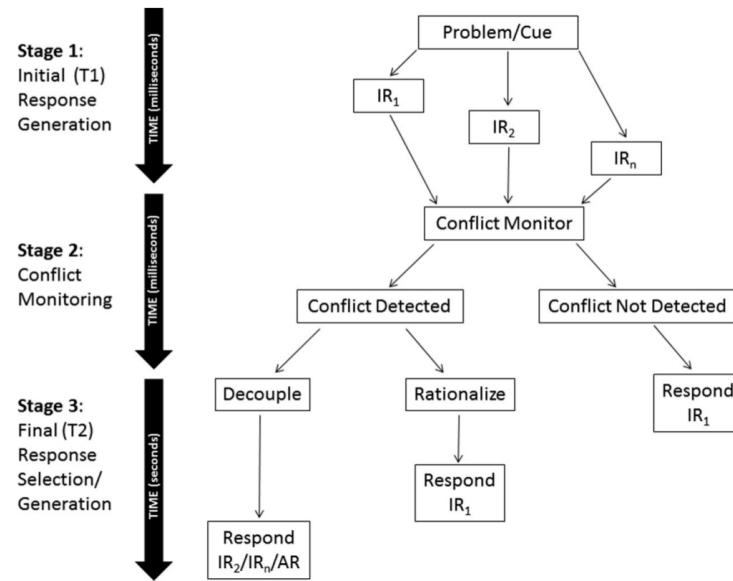


Figure 1.3: The dual process framework proposed by Pennycook, Fugelsang, and Koehler (2015). From top to bottom: once participants are presented with a cue, they generate a number of intuitive responses (IR_1 , IR_2 , IR_n) in parallel, using Type 1 processes. If they do not detect conflict between these responses, they will proceed to give the prepotent response, IR_1 . If they do detect a conflict, they will attempt to engage Type 2 processes, and either seek to justify their strongest intuitive response IR_1 (that is, they *rationalise* it), or inhibit IR_1 in favour of one of the other intuitive responses, or alternatively reach a new response (AR) through Type 2 reasoning. From Pennycook, Fugelsang, and Koehler (2015).

From the above, it is clear that conflict has come to play an important role in dual process theories of reasoning. Later in this thesis, I present experiments where I use the mouse tracking paradigm to reveal more about conflict between Type 1 and Type 2 processes. If we are to study this conflict, however, we need a means of investigating it empirically in the lab. In the next section, I discuss how this can be achieved in general, and in particular how I achieve it in this thesis.

1.4 MEASURING CONFLICT

As we have seen, there are a number of points at which we might expect conflict to occur during high-level reasoning. If we are to study this conflict, however, the data we collect must reflect it. This is not always straightforward. While recent technological advances have made it possible to know whether or not neurons in a given part of the brain, or under some circumstance even specific neurons, are firing, we remain some distance from being able to describe complex cognitive phenomena in terms of these underlying, observable, biological processes. Instead, experimental cognitive psychologists make use of an approach inherited from the Behaviourist tradition: we present our participants with experimentally controlled stimuli, and we record what occurs as a result, either in terms of overt responses, or more subtle measures that shine a light on underlying processes.

In this section, I review methods previously used in the study of both reasoning, and of conflict more broadly. These range from analyses of participants' discrete responses, to response latencies, to subtler paradigms that reveal something of underlying processes, such as fMRI and eye tracking. I also introduce the mouse tracking paradigm, which forms the basis of the experiments reported later in this thesis.

1.4.1 RESPONSES

The basic experimental paradigm used by cognitive psychologists has changed little over the last 50 years: participants are presented with stimuli, ranging from a simple perceptual probes to complex verbal problems, and asked to provide a discrete response, either verbally, numerically, or by pressing a button. Thus, most psychological research rests on the analysis of only these single responses, collected after the process of interest has terminated. More complex designs, incorporating confidence ratings or list sorting, for instance, add some nuance to the analysis, but remain focused on the end product of the cognitive process. Collecting only these responses in a single condition, or in simple experiments, it is not possible to detect conflict directly, but such data can reveal what factors drive decisions. For instance, as discussed above, S. A. Gelman and Markman (1986) presented

children with a forced choice, asking them to generalise properties either between entities that belong to the same category, but look different, or between entities that are visually similar, but belong to different categories. By demonstrating that children overwhelmingly choose to generalise based on category membership, they concluded that category knowledge drives inductive inference in children.

True experiments, in which the factor of interest is manipulated between conditions, provide a better test of what influences final choices. Evans et al. (1983) asked participants to decide if presented syllogisms were valid or invalid, and manipulated if the arguments' conclusions were believable, or unbelievable. Finding that participants were more likely to class arguments with believable conclusions as valid, both for valid and invalid items, they concluded that belief and logic conflicted on the task. Response data can also be combined with a range of information about individual differences between participants in correlational analyses. Stanovich and West (1999, 2000, 2008), for instance, showed that for a number of tasks in which intuition and logic are thought to conflict, the number of logically correct responses a participant produces can be variously predicted by their IQ, or by personality measures.

However, terminal responses remain only an indirect indicator of conflict: it is difficult, for instance, to differentiate between a manipulation that elicits conflict between multiple processes, and one that merely changes the kind of processing participants engage in. There are many measures we can collect that reveal more about underlying processes, but the most popular of these are response latencies.

1.4.2 RESPONSE LATENCY

With the use of personal computers to collect data, it is easier than ever to record not only the responses participants give, but how quickly they give them — their *response latencies*. This *chronometric* approach has a long history, particularly in the individual differences, or *psychometrics* tradition (see Meyer, Osman, Irwin, & Yantis, 1988; Posner, 1978, for reviews). Response latencies have obvious advantages over simple responses in the detection of conflict during cognition. Although conflict may not affect the responses participants give, slower responses are indicative of greater difficulty in generating these responses, often as a result of the need

to resolve or inhibit conflict. In the Stroop task (Stroop, 1935), participants are asked to verbally identify the colour in which words are printed. Doing so may involve conflict when the words themselves are colour names and the printed word and ink colour differ: participants must inhibit an automatic tendency to read the colour name aloud in order to properly perform the task. In non-clinical populations, error rates on this task are typically very low, but conflict, and individual differences in the ability to resolve this conflict, can be inferred by comparing response latencies for congruent and incongruent colour names. The same principle has been applied to forced choice tasks, including the Eriksen flanker task (Eriksen & Eriksen, 1974), in which participants are required to respond on the basis of a centrally located probe, ignoring flanking stimuli on either side which may be congruent, incongruent, or neutral. Similarly, in the Simon task (J. R. Simon & Wolf, 1963), stimuli can be presented centrally on a display, or lateralised to either side. Responses with the right hand are slower to stimuli presented in the left visual field, and vice versa.

One issue that arises in the analysis of response latencies is the relationship between response time and choices. On tasks in which more than one response is possible, and there is variance in the choices participants do make, there is often a *speed-accuracy trade-off*, with participants striking a balance between maximising speed, and thus sacrificing accuracy, or vice versa (Garrett, 1922). A popular solution to this problem has been to fit models of the underlying decision process that account for both the participants' choices and their corresponding response times. *Sequential sampling models* (Busemeyer & Townsend, 1993; Hawkins, Hayes, & Heit, 2015; Ratcliff, 1978; Ratcliff & McKoon, 2008) are commonly used for this purpose. These model both participants' responses and their response latencies as a process of accumulating evidence over time. However, despite the strengths of recent modelling-based approaches in inferring underlying processes from responses latencies, the data upon which these models are based remain after-the-fact products of the processes we are interested in. Greater sensitivity still comes from a family of methods known as “*process tracing*” that seek to monitor these cognitive processes as they happen.

1.4.3 PROCESS TRACING

Process tracing paradigms are those where cognitive processes, or their correlates, are recorded as they unfold. These paradigms range from asking participants to think aloud, to measurement of neural and biological states, including fMRI and EEG, to recording of eye gaze in order to infer attention during cognition, and to mouse tracking, the method which will form the core of this thesis.

Protocol analysis (Ericsson & Simon, 1980; Ranyard & Ola, 2010), where participants are asked to explicitly “think aloud” while performing a task, is perhaps the most straightforward such method. Often, this involves experimenters specifying, *a priori*, the possible strategies participants may use, or the information they may make use of, and identifying from participants’ verbal reports which of these were actually used. In reasoning research, Evans et al. (1983) make use of this approach in conjunction with experimental data, showing that when asked to evaluate the logical validity of syllogisms, participants were influenced by whether or not the conclusion of the syllogism was believable, and that participants explicitly reported taking believability into account when completing the task. Of course, as noted by Nisbett and Wilson (1977), not all cognitive processes are available to introspection. Therefore, this paradigm works best as a measure of the steps participants are explicitly aware of taking, or the information they were explicitly aware of using, during a task. For instance, De Neys and Glumicic (2008) showed that while implicit measures showed that participants were sensitive to conflicts between their biased inferences and statistical information, participants rarely made reference to this statistical information, suggesting that this conflict detection was achieved implicitly.

Other methods go beyond the limitations of relying on participants’ self reported mental states. There has been a huge body of research, across many fields, on the neural and biological correlates of cognition and their measurement, which I can only briefly list here. Functional Magnetic Resonance Imaging (fMRI; Huettel, Song, & McCarthy, 2004) is perhaps the best-known of these techniques, and reveals the firing rate of populations of neurons by monitoring the oxygen content of inflowing blood. This provides excellent spatial resolution, but at a substantial temporal delay. In contrast, electroencephalography (EEG; Niedermeyer &

da Silva, 2005) and less commonly magnetoencephalography (MEG; Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993) record the weak electrical and magnetic fields generated by active neurons, allowing for extreme temporal accuracy in inferring neural activation, but with limited spatial acuity.

The usefulness of these methods in studying cognitive tasks relies on the assumption that cognitive processes can be directly mapped onto the activation of specific neural regions, or to characteristic signals time-locked with the onset of the relevant stimuli (Event Related Potentials). Fortunately, although this assumption has been questioned in a number of domains (see Coltheart, 2013), there appear to be clear neuroanatomical mappings in the case of conflict in cognition. Specially, the detection of conflict between processes is known to be strongly correlated with activation of the ACC (M. Botvinick & Plaut, 2004), while the processes of inhibiting conflicting processes engages the prefrontal cortex (Aron, Robbins, & Poldrack, 2004; E. K. Miller & Cohen, 2001). Thus, the logic of studying conflict using these measures is clear: a manipulation can be claimed to induce conflict if it leads to greater activation of the ACC, and participants who show greater PFC activation are engaged in the inhibition of a conflicting process (see, i.e. De Neys & Glumicic, 2008).

Other biological measures exist that do not directly reflect neural activity, including galvanic skin conductance (De Neys et al., 2010; Figner & Murphy, 2010), and pupil dilation (Fiedler & Glockner, 2012; Kahneman & Beatty, 1966; J. e. Wang, 2011). All of these measures reflect arousal of the autonomic nervous system, which in turn is influenced by *cognitive load* (Kahneman & Beatty, 1966). However, like the neural measures reviewed above, the processes measured in these paradigms are not the cognitive phenomena in which we are primarily interested, but epiphenomenal by-products, or at best physical correlates of the mental processes.

An alternative to analysing these by-products is to look to a process that we know to precede cognition: attention. Although some work (i.e. Evans, 1996) has asked participants to explicitly indicate the locus of their attention, the most commonly used measure of attentional focus is the eye tracking paradigm. Eye tracking has a long history in psychology (Ball, 2014; Huey, 1908; Mele & Federici, 2012; Yarbus, Haigh, & Rigss, 1967; see Duchowski, 2007 for a review) and in its

modern form typically relies on using a camera to record the reflection of infrared light off different parts of the eye. Popular uses of the method include measuring attention as part of complex behaviours such as social interaction (i.e. Hanley et al., 2014), decision making (Krajbich & Rangel, 2011) and reasoning (Ball, 2014), as well as in scenarios in which the low-level processes driving gaze fixation themselves are of interest, such as during reading (Rayner, 1998) and visual search (Treisman & Gelade, 1980). The strengths of the eye tracking paradigm are clear. It provides an unobtrusive measure of participants' attention, with impressive spatial and temporal resolution, across almost any cognitive task.

As a paradigm, eye tracking is also extremely flexible, and eye tracking data can be analysed in a number of ways. Perhaps reflecting the popularity of eye tracking in psycholinguistics, many experiments use eye tracking data to infer how long participants spend reading particular passages of text, reflecting the time spent processing particular information, or its unexpectedness. It is additionally possible to differentiate between effects which manifest quickly (which affect participants' first pass over the text in question) and those which occur later (which affect processing of subsequent text, or regressions back into the text in question). Haigh, Ferguson, and Stewart (2014), for instance, presented participants with vignettes where a particular inference could be drawn early in the text, and showed that participants slowed down on their first pass through sentences which were inconsistent with that inference, indicating that the inference was made quickly and automatically while reading. Alternatively, participants' gaze can reflect the extent to which they are considering certain options. Ball, Lucas, Miles, and Gale (2003) used eye tracking in this way, analysing participants' *inspection times* for each card on the Wason (1968) selection task. They found that participants were more likely to initially inspect cards cued by Type 2 processes (those mentioned by name in the instructions), and subsequently more likely to select cards if they inspected them initially. Reanalysis of this data by Evans and Ball (2010) showed that Type 2 processes also played a role, as participants were more likely to select a card after initially inspecting it when doing so was the logically correct thing to do.

Eye tracking is also a useful tool when studying conflict. In the popular Visual

World paradigm (Allopenna, Magnuson, & Tanenhaus, 1998; Huettig, Rommers, & Meyer, 2011; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), for instance, participants are instructed over headphones to interact with a display in front of them, or on a computer screen. On some trials, there was syntactic ambiguity in these instructions — for example, they begin “place the apple on the towel ...” when the display includes an apple, a towel, and a second apple on a second towel; it is not clear if the task is to place the first apple onto the towel, or move the apple that is already on the towel elsewhere. In these cases, participants look to both possible referent, before later information (e.g. “... into the box”) disambiguates the referents. In a similar vein, McMurray, Aslin, Tanenhaus, Spivey, and Subik (2008) asked participants to categorise phonemes that were intermediate between simple consonants such as “[d]” vs. “[g]”. Although participants typically report *categorical perception* of one or other consonant, they looked more often at the alternative response option as the stimuli became more ambiguous, indicating conflict. Intriguingly, Richardson and Spivey (2000) demonstrated that eye tracking can also reveal retrieval from memory, as participants look at regions of the screen where information was presented when recalling that information. Finally, while most analyses of eye tracking experiments reduce the data down to single summary measures — for instance, time spent looking at each response option — it is also possible to analyse the data as a continuous time series. Recently, several psycholinguistics researchers have explored the use of multilevel regression techniques to make sense of eye tracking data in such a way (e.g. Barr, 2008; Mirman, 2014a; Mirman, Dixon, & Magnuson, 2008). I will return to these methods in Chapter 2.

One notable shortcoming of the eye-tracking paradigm in studying high-level cognition, however, is that the connection between attention and the underlying processes of interest cannot always be taken for granted. If we find that participants look more to one option than another, for instance, it is difficult to tell if they look to this option because they wish to choose it, they wish to choose this option because they have looked at it, or some interaction of the two (see Krajbich & Rangel, 2011). Furthermore, there is always a possibility, difficult to eliminate, that participants look to something not because they wish to choose it, but be-

cause they wish to rule it out. Beyond this, participants can obviously only look at one thing at a time, and eye movements typically consist of a series of *saccades*: rapid, discrete movements between one fixation point and another, in a straight line. Generally, researchers will circumvent this issue by aggregating participants' fixations: if they look to option A half the time, and option B the remainder, then they are judged to be equally drawn towards both option. The issue remains, however, that researchers are required to construct often complex theoretical models as a bridge between hypothesised psychological processes and observed eye tracking data.

1.4.4 MOUSE TRACKING

The mouse tracking paradigm, introduced by Spivey et al. (2005), goes some way towards overcoming some of these problems inherent to eye tracking. In this, participants' mouse cursor movements are recorded as they complete forced-choice tasks, and analysed to infer the extent to which participants are drawn towards each response option over the course of a trial. Therefore, this paradigm appears to provide a window into the temporal dynamics of participants' developing beliefs over the course of a trial. It is this paradigm that I will use throughout this thesis, and so I will provide a general introduction to it in the current chapter, before discussing technical aspects of the paradigm in detail in Chapter 2.

In Spivey et al.'s (2005) original experiment, participants saw a computer display containing images of common objects in the top corners (see Figure 1.4). Over headphones, they heard the name of one or other object, and were required to click on the relevant image. For some trials, the two objects were *phonological competitors*, meaning that their names shared onset phonemes, such as “*picture*” and “*pickle*”, while for others they did not. On these conflict trials, participants' mouse cursor trajectories displayed what has been termed *continuous attraction*, moving initially towards a point between the two images, before eventually homing in on one or other option. A classical view of language comprehension (Fodor, 1983) holds that speech recognition is handled by a *perceptual module*, which processes its input, and then passes on its output. However, these results indicate that the speech recognition instead feeds continuously into other processes, includ-

ing those executed by the motor system, in real time. Spivey et al. (2005) argue that these data support a *continuous dynamic* model of language comprehension, where spoken words initially partially activate representations of multiple words consistent with the stimuli, before this competition is resolved in favour of a single candidate word (see also Spivey, 2007). While the issue of language comprehension will not be the focus of this thesis, this first application of the mouse tracking paradigm demonstrates some of its value. Based on their mouse data, Spivey et al. (2005) were able to demonstrate two things. First, they showed that conflict takes place during their task. Second, they showed that this conflict is continuous in nature (participants were simultaneously drawn towards both response options), rather than discrete (initially be drawn towards one response, before changing their mind mid-trial).

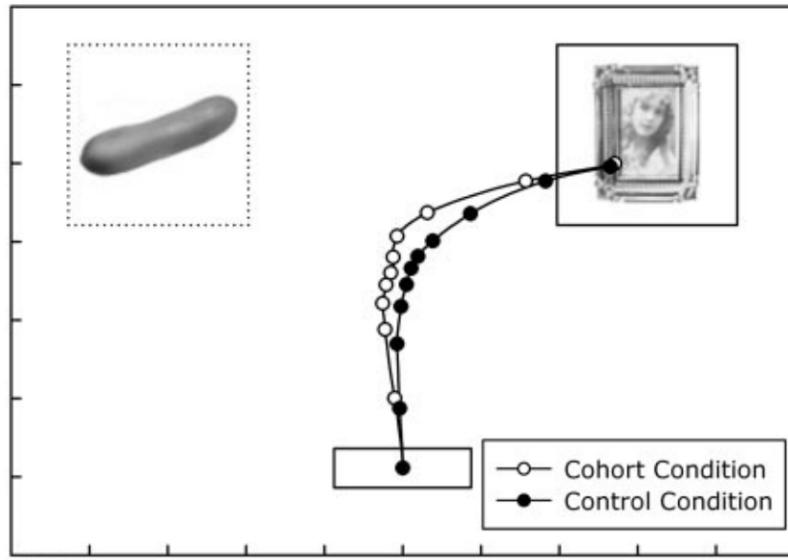


Figure 1.4: Spivey et al.'s (2005) mouse tracking paradigm. Participants were told aurally to click on the “picture”. Their mouse cursor trajectories showed a greater, graded attraction towards the alternative icon in the cohort (conflict) condition, where it showed a phonological competitor (PICKLE), than in the control condition, where it did not.

Therefore, mouse tracking can reveal more than simply the presence of conflict. For instance, Spivey et al.'s (2005) original mouse tracking experiment, and many since, have reported continuous attraction effects — participants were partially and simultaneously drawn towards both response options at the same time, resulting in a curved trajectory. Other studies, however, (e.g. Barca & Pezzulo, 2015; Dale & Duran, 2011; Freeman, 2014; Resulaj, Kiani, Wolpert, & Shadlen, 2009;

Tomlinson, Bailey, & Bott, 2013) have revealed more discrete effects, as participants move directly towards one or other option, but sometimes change direction mid-flight. In this way, mouse tracking can show us something of the qualitative nature of conflict during a task. In Appendix A, I have compiled a table showing some of the mouse tracking studies that have found graded, continuous attraction effects, and some of those that have shown discrete reversals, or changes of mind. I return to the issue of reversals in cursor trajectories from a practical point of view in Chapter 2, and again from a theoretical point of view in Chapter 7.

Mouse tracking may also reveal something of *when* conflict occurs, or at what point in time various factors influence participants' cursor movements. For example, Freeman and Ambady (2011b) asked participants to categorise faces according to either their gender, or their age (young or old), and manipulated both the pigmentation and the shape of faces to be either category-typical or category-atypical. They found that face pigmentation affected cursor trajectories for gender categorisation 50 ms before face shape did, and that the effect of pigmentation manifested 100 ms earlier for gender categorisation than for age categorisation.

At this point, it is worth highlighting that this procedure is not the only known "mouse tracking" paradigm. Analysis of computer mouse cursor locations have been used for some time as a measure of participants' information search, or attention, during cognitive tasks. Information search experiments are typically implemented in two ways, using either the "flashlight" procedure (Schulte-Mecklenbeck, Kühberger, & Ranyard, 2011; Yamauchi, Kohn, & Yu, 2007), in which regions of the screen are blurred or dark, and the mouse must be pointed directly at them in order to see the relevant stimuli, or the MouseLab procedure (Willemse & Johnson, 2011), in which information is presented in tables with occluded cells, that participants must click or hover over to temporarily reveal. In a related paradigm, participants' patterns of attention can be monitored by explicitly asking them to point the cursor at the focus of their attention (Evans & Ball, 2010; Evans & Over, 1996; Stupple & Ball, 2008). This method likely provides a low-cost proxy for more difficult eye tracking methods.

Another field of research making use of mouse cursor data is work by those interested in human computer interaction. As the mouse cursor is the natural means

of interacting with computer interfaces, such as web pages, analysing mouse movements provides a naturalistic measure of the effects of manipulating their layout and content (e.g. Cox & Silva, 2006; Huang, White, & Dumais, 2011). Work combining mouse and eye tracking (M. C. Chen, Anderson, & Sohn, 2001; Rodden, Fu, Aula, & Spiro, 2008) has shown that while mouse movements correlate with gaze, they also are used in more task-specific ways, such as hovering near potential selections as a marker while gaze is used to explore other less likely candidates. I will return to these alternative notions of mouse tracking in Chapter 6, where a novel mouse tracking paradigm is introduced. However, for the first three experimental chapters of this thesis, the paradigm used will be essentially that used by Spivey et al. (2005).

1.5 THESIS OVERVIEW

This thesis is about conflict in reasoning. Many previous investigations of such conflict have used methodologies that focus on the end product of the reasoning processes, rather than the unfolding process itself. In this thesis, I apply a method new to the reasoning literature, the mouse tracking paradigm, which allows not only analysis of participants' responses, but also their movements on the way to giving these responses. In doing so, the experiments reported herein reveal much about when conflict occurs during reasoning, as well as helping to answer questions about the qualitative nature of this conflict.

The objectives of this thesis are threefold. First, by applying the mouse tracking paradigm to a number of reasoning tasks, I will show directly when conflict occurs during reasoning. Second, by analysing the time course of participants' mouse cursor trajectories, it is possible to infer at what point in time participants are driven by various competing influences in reasoning. Finally, the paths followed by these mouse cursors can reveal not only the presence of conflict, but something of the qualitative nature of this conflict and the interaction between the various processes.

1.5.1 THESIS STRUCTURE

The remainder of thesis is organised as follows. In Chapter 2, I discuss the technical details of the mouse tracking paradigm used in subsequent experiments, and the analyses used to explore the data it generates.

Following this, I report the results of six experiments, across four chapters. Chapters 3 and 4 concern conflict between different sources of information during inductive reasoning. In Chapter 3, I present versions of the inductive triad task that pit information present in perceptual cues against more abstract conceptual knowledge. This is done both for real categories in the natural world (Experiment 1) and for artificial categories that participants learned in the lab (Experiment 2).

Chapter 4, in a similar way, pits associative knowledge against structured, relational knowledge, adapting the triad task presented by Bright and Feeney (submitted). Experiment 3 directly adapts Bright and Feeney's (submitted) stimuli for the mouse tracking paradigm, while Experiment 4 replicates Experiment 3 using different stimuli, and a design that allows for a more fine-grained analysis of how the two forms of knowledge interact.

Chapters 5 and 6 deal with conflict in reasoning from a dual process perspective, and pit fast, automatic Type 1 processes against slow, effortful Type 2 processes. In Chapter 5 (Experiment 5), I present a mouse tracking version of one of the most popular tasks in the dual process/heuristics and biases literature: the base rate neglect paradigm, where participants make judgements that can be influenced by stereotypes about social groups, thought to be mediated by Type 1 processes, and statistical information, traditionally thought to require Type 2 processes to process.

Chapter 6 (Experiment 6) departs somewhat from the previous work, as I present a four-option multiple choice version of the Cognitive Reflection Test (Frederick, 2005), a paradigmatic dual process/heuristics and biases task that pits incorrect, intuitively appealing, heuristic Type 1 responses against the less obvious correct response, reached using Type 2 processes. Unlike the previous experiments, this involves analysis of participants' mouse movements over up to 60 seconds, as the meandering of the cursor during complex reasoning reveals participants tentative beliefs about each response.

This thesis ends, in Chapter 7, with a general discussion, in which I talk about what has been learned over the course of the previous six experiments, and discuss implications for theories of reasoning, accounts of conflict in cognition, and users of the mouse tracking paradigm.

CHAPTER 2

Methods

The six experiments that follow apply the mouse tracking paradigm to a number of reasoning tasks. Many details of this paradigm are constant over the first five of these experiments, and so in this chapter I discuss the procedure used, and the analysis of the resulting data, in more detail. This is hoped to serve both as a primer on the paradigm for readers unfamiliar with it, and to cover the technical details of these experiments. There is of course some variation between experiments, and so this chapter will mostly cover those details which would become tedious if repeated for each experiment. This chapter also introduces some of the statistical analyses used later in this thesis for readers unfamiliar with them.

2.1 MOUSE TRACKING

The mouse tracking paradigm was introduced by Spivey et al. (2005). Participants saw a computer display with images of common objects in the top corners (Figure 2.1), and told over headphones to click on one or other image. When the objects pictured had names that were phonological competitors (e.g. “picture” and “pickle”) participants’ cursor trajectories to the object named were less direct, compared to when the foil was unrelated (“picture” and “candy”).

In the years since its introduction, this paradigm has been adapted for a range of problems. Unsurprisingly, there has been considerable diversity in these subsequent mouse tracking studies, in terms of specific details of the experiment itself, the processing and analysis of dependent variables, and the nature of the results

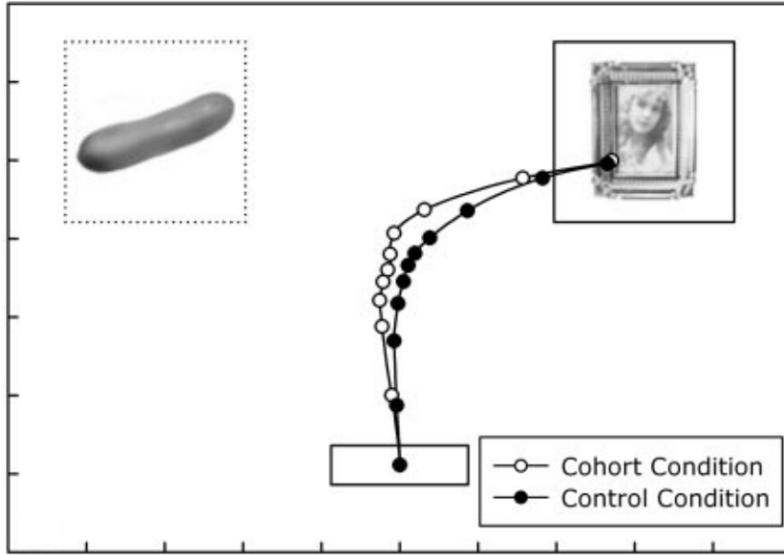


Figure 2.1: Spivey et al.’s (2005) mouse tracking paradigm. Participants were told aurally to click on the “picture”. Their mouse cursor trajectories showed a greater, graded attraction towards the alternative icon in the cohort (conflict) condition, where it showed a phonological competitor (a pickle), than in the control condition, where it did not.

obtained. In the next section, I review previous mouse tracking work along these lines, and outline the procedure as used in the later experimental chapters of this thesis. By covering this material here, we are saved from repeating many of the technical details for each experiment.

As discussed in Chapter 1, this is not the only mouse tracking paradigm used in behavioural research. However, in this review, I will focus on the two-option forced choice mouse tracking design introduced by Spivey et al. (2005).

2.1.1 PROCEDURE

The canonical mouse tracking procedure differs little from that used by Spivey et al. (2005). Two response options are located in the top corners of the screen, and participants click on a start button located in the bottom of the screen to begin a trial. After this, and usually following a fixation cross, stimuli are presented, either visually or aurally, and participants move the mouse cursor from its starting point to click on one or other response option.

Naturally, for mouse tracking to be useful, participants must begin moving the cursor before they have fully made up their minds on each trial. Following Freeman and Ambady (2009), it is common practice on trials where the mouse cursor

isn't moved within a given time from stimulus onset (400 msec in Freeman & Ambady, 2009) to show participants a message asking them to start moving earlier, even if they weren't fully certain of their response. The length of this window obviously varies between tasks, as participants may simply refuse to move too early for more difficult decisions. In the experiments reported here, the maximum initiation window for each experiment was decided based on pilot testing. Scherbaum, Dshemuchadse, Fischer, and Goschke (2010) offer a different solution to this problem by programming their experiment to only reveal the stimuli once participants began their mouse movement. However, this technique does not appear suitable for the more complex experiments reported in this thesis.

In designing the experiments here, I have developed a technique which may prove useful for readers interested in adopting these methods, of encouraging mouse movements during the decision process. Through pilot testing, it is possible to estimate a reasonable amount of time for participants to be able to initiate their response movements for a given experiment. Having identified a desired initiation time (for instance, 500 msec), the duration of the fixation cross is set to three times this, that is, a blank screen for 500 msec, a cross for 500 msec, and a blank screen for 500 msec, followed by onset of the stimuli. This creates a ready-set-go effect, and encourages participants to attempt to synchronise the onset of their own movement to this interval.

Another variable is the maximum amount of time allowed to respond. For many simple experiments, such a limit is unnecessary, as participants respond within 2 to 3 seconds without extra prompting. For more complex experiments, and certainly for reasoning, however, self-paced participants would be likely to take considerably longer than this to respond, and so their decisions would not be captured in their mouse movements. In addition, to require early movement initiation, without speeded responses, would incentivise participants to move the cursor around the screen as they made their decisions an in undesirable fashion. Therefore, most of the experiments reported here required participants to respond within a certain time window, the length of which was again decided upon after pilot testing.

A fine detail in these experiments is the location of the cursor on stimulus onset. In the majority of experiments, this is controlled by the software: after clicking on

the start button, a fixation cross is typically displayed before stimulus onset, and the cursor position is reset to the bottom centre of the screen if it has moved during the fixation period. In the experiments reported herein, two different sets of software are used for different experiments, one written in python, and one run in the web browser (see below). At present, it is not possible to manipulate the position of a mouse cursor in the browser, and so in the experiments which used this software it was not possible to reset the cursor position. This issue was circumvented in these particular experiments by using very short fixations, or no fixation at all.

In mouse tracking experiments, attention must also be given to the way in which information is presented to participants during a trial. Most experiments use simple stimuli, such as images or text, which can be presented to participants simultaneously. In reasoning experiments, however, it is often necessary to present participants with much more information. My initial attempts to present entire reasoning problems at once revealed that participants would first processes all of the stimuli, over several seconds, and only after making their decision move the mouse cursor. Therefore, great care was taken to present participants with stimuli on each trial in such a way that they had sufficient time to process all of the contextualising information. Trials were structured so that it was not possible to begin making a decision until the onset of the final stimuli, so that the movement of the mouse cursor could capture this process. For instance, in Experiment 3, an inductive reasoning task, participants were presented with a property, told that it can be found in one of two species, and shown each of the candidate responses in their locations in the top corners of the screen, one at a time. Only after this information was presented could participants click on the start button, which caused the base species, which was known to have the property in question, to be shown in the centre of the screen. At this point, participants used this new information to decide which of the candidate species was more likely to possess the property.

2.1.2 STIMULI

Typically, mouse tracking experiments use one of two manipulations. One option is to manipulate, between conditions, the probe to which participants respond. In

Freeman, Ambady, Rule, and Johnson (2008), participants were shown computer-generated male and female faces which were either gender typical or typical, while the response icons remained constant: MALE, and FEMALE. The alternative is to manipulate the possible responses. For instance, Spivey et al. (2005) presented the probe word “pickle” twice: once when the foil response was a phonological competitor (a PICTURE) and once when it was not (a CANDY).

This distinction may be important, as in the former case the experimental manipulation changes the relationship between the probe and the chosen response, and so it is difficult to disentangle an effect driven by conflict, and an effect driven by reduced confidence in the response. In the latter case, the link between the probe and the response remain the same, and so any changes in motor output must be a result of an attraction towards the distractor response. Dale, Kehoe, and Spivey (2007) demonstrate the value of combining these approaches. In their first experiment, they showed that as participants categorised animals into their super-ordinate categories (i.e. WHALES as MAMMALS), they showed greater mouse curvature for atypical category members (WHALES) than typical ones (COWS). However, this effect could be simply due to reduced confidence in the chosen response, rather than conflict per se. To show that this was a conflict effect, they ran a follow-up experiment showing that the effect was diminished when the distractor was not associated with the probe item (i.e. BIRDS, instead of FISH, for WHALES). Another possible implication of this distinction, which is yet to be fully investigated, is that when the distractor response is the independent variable participants could plausibly be faster to execute some aspects of their responses under conflict. Although their reasons for choosing the dominant response remain the same, the conflict condition in such experiments provides an additional reason to initiate a response, and so movement initiation may paradoxically be faster under conflict. In this thesis, I use both kinds of manipulation.

2.1.3 SOFTWARE

A number of tools are available for running mouse tracking experiments, and processing the resulting data. MouseTracker (www.MouseTracker.org; Freeman, Pauker, Apfelbaum, & Ambady, 2010), the most popular option, provides a graphical ex-

periment builder, and a point-and-click interface for calculating relevant summaries of trajectories (discussed below) and exporting them for analysis. While MouseTracker was initially used in this research, and should be recommended for researchers interested in exploring mouse tracking, it sacrifices a certain amount of flexibility for ease of use, and so proved to be inadequate for the complex reasoning paradigms used in this thesis.

I developed two tools to solve this problem. First, I implemented the paradigm as a script in the python programming language, and ran it using the open source OpenSesame experiment builder (Mathôt & Theeuwes, 2011). This implementation, along with detailed instructions, are available online at <https://github.com/EoinTravers/QuickstartMousetracking>. I also created a second implementation of the paradigm as part of PsychScript (<https://github.com/EoinTravers/PsychScript>), a set of tools I have written in HTML, JavaScript, and PHP for running behavioural experiments online in the web browser. Both implementations of the mouse tracking paradigm were extremely similar, and differed principally in that it was not possible to manipulate the position of the mouse cursor using PsychScript, and so its location was not reset on experiments implemented using this platform (see above). Although it was possible to run experiments programmed using PsychScript outside the laboratory (that is, online), this capability was only used for pilot testing due to concerns about the quality of data collected in this way. Instead, this implementation was primarily used to collect data in undergraduate computer laboratory classrooms, as it made it possible to recruit large numbers of participants at once without the necessity of installing OpenSesame on each computer.

MouseTracker automatically calculates a number of statistics for describing mouse cursor trajectories to be used in later analyses (see below). Using my own implementations of the paradigm, it was also necessary to write code to calculate these same measures. This was done in the python programming language, and released as the python package Squeak (<https://github.com/EoinTravers/Squeak>). A guide to using Squeak can be found alongside my python implementation of the mouse tracking procedure, at <https://github.com/EoinTravers/QuickstartMousetracking>.

2.2 ANALYSING MOUSE DATA

Even for simple experiments, mouse tracking creates a voluminous amount of data. As a result, there are many ways in which data from these experiments can be analysed. In general, however, two strategies are popular: each trials' trajectory can be reduced to a single summary measure, or the actual time course of trajectories across conditions can be analysed. While space precludes a review of all measures used in this kind of research, in this next section I will outline the measures used in the present studies, and the motivation behind their choice.

The simplest summary measure, of course, is response latency, or response time (RT), which works the same way for mouse tracking as it does in other paradigms: the time between stimulus onset and a response. This can be accompanied by a measure of movement initiation time (IT), reflecting the time from stimulus onset to the cursor beginning to move. In the current studies, IT was defined as the time taken for the cursor to move 1% of its total distance travelled vertically.

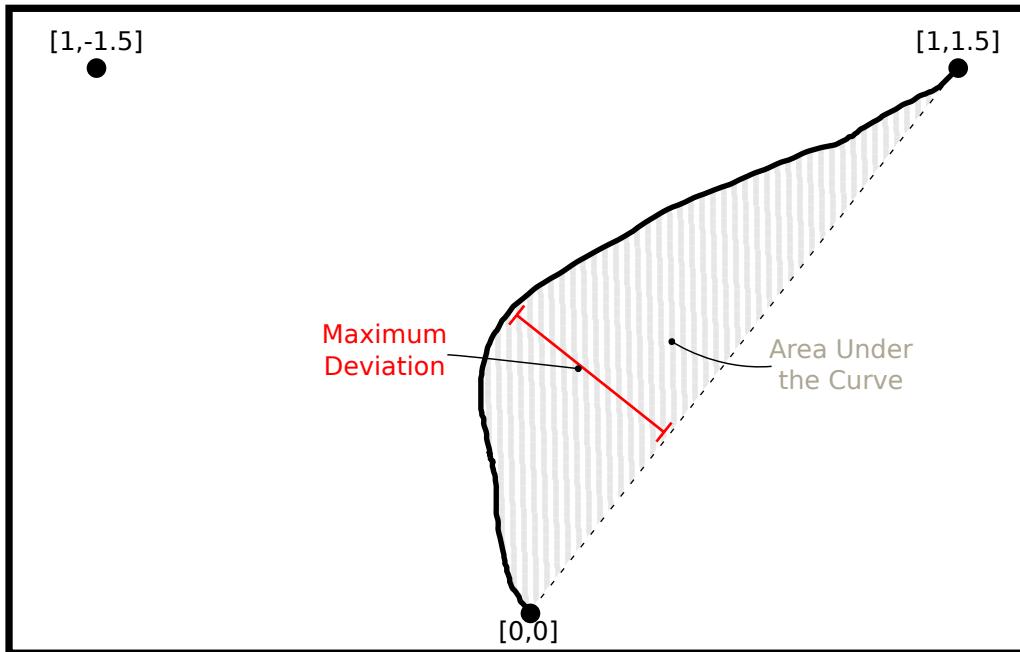


Figure 2.2: For analysis, mouse cursor trajectories are remapped to a standard co-ordinate space, beginning at [0, 0] and ending at [1, 1.5]. Area Under the Curve reflects the size of the region bounded by this trajectory and an ideal straight line from start to finish, in the dimensions of the standard co-ordinate space. Maximum Deviation is the greatest perpendicular distance between the ideal straight line and the actual trajectory.

More interesting, however, are measures derived from the shape of the trajectory

followed by the cursor. For these, as in all subsequent analyses of cursor movements, each trajectory is mapped on to a common co-ordinate system, such that they begin at point [0, 0] (the bottom centre), and end at point [1, 1.5] (the top right corner; see Figure 2.2). Trajectories which finished in the top left corner are usually reflected on the y axis so that all trials end at the same point, although in analyses including both correct and incorrect responses the original endpoint of [1, -1.5] may be used as the location of the incorrect response option. Spivey et al. (2005) quantified the curvature of their cursor trajectories using a measure of Area Under the Curve (AUC), corresponding to the area bounded between the actual trajectory and a straight line between its start and end points (again, see Figure 2.2). Complications arise, however, when we attempt to calculate AUC for trajectories which contain loops, or those which pass on the other side of the idealised straight line (closer to the edge of the screen), as the treatment of such cases varies depending on the algorithm used, and no standard has been established among mouse tracking researchers. While such trajectories were apparently rare in studies of simple perceptual judgements and could be excluded from analyses, they have proven more common in my work on reasoning, and so I do not make use of AUC in this thesis.

A commonly used alternative to AUC is the Maximum Deviation (MD) of a trajectory, the greatest distance achieved between the trajectory and the idealised straight line (again, see Figure 2.2). In conversation with other researchers interested in mouse tracking, it has also become apparent that there is more than one way to calculate this. For clarity, MD is calculated in this thesis by rotating a trajectory by $\tan^{-1}(\frac{1.5}{1})$ (approximately 56°) clockwise so that its ideal straight line follows the x axis. MD is then simply the greatest height of the rotated trajectory on the y axis. MD can be negative: trajectories which stray further from the straight line to the right (towards the edge of the screen) than to the left (towards the alternative response) will have MDs of less than 0. Some variations have been proposed on this measure, including Maximum Absolute Deviation, in which these negative values are treated as positive, and Average Absolute Deviation, which takes the mean of the absolute deviation across the time course of trial (Koop & Johnson, 2013). However, I have opted for MD as my preferred measure, largely

because it is the most commonly used measure of deviation.

Other measures reflect the complexity of cursor movements, indicating either conflict, or at very least uncertainty (Dale & Duran, 2011; Wojnowicz, Ferguson, Dale, & Spivey, 2009). X-flips (changes of direction on the x axis) over the course of a trajectory, provide one such measure, although it is not always clear what this statistic reflects, as zig-zags across the entire screen and wobbles over a few pixels both constitute changes of direction. Alternatively, it is possible to calculate the number of acceleration components – changes from acceleration to deceleration, and vice versa – during a trial as a measure of movement complexity (Dale & Duran, 2011). An interesting alternative measure of complexity is sample entropy, (Dale et al., 2007; Richman & Moorman, 2000) which reflects the amount of extra information imparted at every time point in a trajectory. However, I do not make use of these measures in this thesis.

2.2.1 CONTINUOUS AND DISCRETE EFFECTS

Unlike eye movements, which are ballistic in nature, and so (almost) always move in a straight line from their starting point to their destination, hand movements can follow any number of trajectories. Thus, the mouse tracking paradigm can also reveal something of the qualitative nature of conflict as it unfolds. In Chapter 1, I discussed two ways conflict has been seen in mouse tracking. Most studies have found continuous attraction effects, where participants are initially drawn towards both response options, but gradually home in one or other of them, following a graded curve of the type seen in Figure 2.2. Some studies, however, have found discrete effects, where participants sometimes move directly to their response, but other times initially move in one direction, but change their mind, and change direction, mid-trial. In Appendix A, I list studies in which each kind of effect has been found.

To tease these possibilities apart, it is necessary analyse the distribution of trajectories in an experiment (Freeman & Dale, 2012; Spivey et al., 2005). While these points also apply to the trajectories themselves, for practical purposes these analyses usually focus on summary measures, such as AUC or MD. In cases of continuous attraction, according to this logic, we should find that most trajectories

are slightly curved, and that the average curvature is greater for conflict trials. Statistically, this equates to data where MD is normally distributed within each condition, and the mean of this distribution is higher for conflict trials. In cases of discrete attraction, however, there are some trajectories which go straight to an option, with very little curvature, and some which initially go one way, and then change direction, with high curvature. Statistically, this leads to MD being bimodally distributed, with the direct trajectories forming a sub-population with MD close to 0, and the reversals, or changes of mind, forming a sub-population with a much higher mean MD (Figure 2.3).

To test for bimodality in their data, Spivey et al. (2005) calculated the Bimodality Coefficient (BC; SAS Institute, 1990):

$$BC = \frac{m_3^2 + 1}{m_4 + 3 \times \frac{(n-1)^2}{(n-2)(n-3)}} \quad (2.1)$$

where n is the sample size, m_3 the skewness, and m_4 the excess kurtosis (see Freeman & Dale, 2012; Pfister, Schwarz, Janczyk, Dale, & Freeman, 2013). A uniform distribution has BC of .555, and so values higher than this constitute evidence of bimodality. Spivey et al. (2005) also supported their interpretation, although this method has not been widely adopted, by comparing the distributions of the conflict and control conditions using the Kolmogorov–Smirnov test. Finding that there was no significant difference in the shape of the distributions between conditions, they concluded that the greater AUC under conflict was not the result of a subset of high AUC trials. Freeman and Dale (2012) provided a more in-depth analysis of tools for distinguishing between unimodal and bimodal distributions, and recommend, based on simulation results, the use of Hartigan’s Dip Statistic (Hartigan & Hartigan, 1985). This measure tests the null hypothesis that the data is unimodally distributed, and so $p < .05$ indicates that the data is significantly bimodal. In this thesis, I make use of both BC and Hartigan’s Dip Statistic when analysing distributions.

In all of the experiments reported in Chapters 3, 4, and 5, I find evidence of discrete attraction: some trials move directly towards the correct response option, some initial move towards the wrong option before redirecting, while some give

the wrong response (see Appendix B). It is difficult, from a statistical perspective, to analyse such data, as most analyses require normally distributed residuals, and unlike in the case of skewed data such as response latencies, there is no transformation to map bimodal data onto a normal distribution. Therefore, I instead class each trajectory as being either direct (from the low-MD sub-population) or a reversal (from the high-MD sub-population), and use the proportion of reversals in each condition as my dependent variable.

To classify trajectories in this way, I use two-sample finite mixture models, modelling the MD distribution as two normally distributed sub-populations, and derive from this a cut-off value. MDs below this value are classed as direct movements, and those above it as reversals. I have chosen this approach, rather than others such as the k-means algorithm, as unlike k-means, finite mixture models allow for differences in variability between the two sub-populations: we would expect less variability in direct movements to the response than in reversals, and this is confirmed in the data. Figure 2.3 shows this procedure as conducted for Experiment 2. Appendix B shows details of this procedure for each of Experiments 1 to 5.

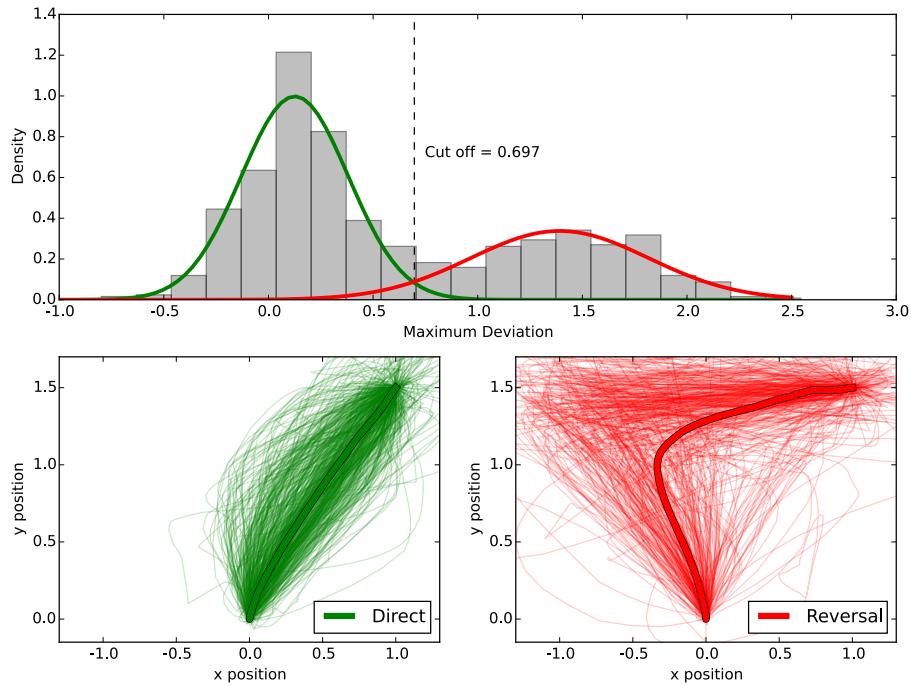


Figure 2.3: Trajectories classed as direct and reversals in Experiment 2.

2.2.2 TYPES OF TRAJECTORY

In these mouse tracking experiments, participants chose either the correct option or the foil, and either traced a direct path, or changed direction while doing so. In this sense, setting aside questions of *when* each movement occurs, there are four kinds of responses, or trajectories, possible in such experiments: direct movements to the correct response, reversals ending with the correct response, direct movements to the foil, and reversals ending with the foil (see Figure 2.4). It is illustrative to analyse how the proportion of each trajectory changes between conditions in an experiment, and I do so for Experiments 1 to 5.

Perhaps a more interesting analysis, however, is to look to the decisions participants could make over the course of a single trial. I assume that participants make two decisions on each trial. First, at the onset, they must decide to move either towards the correct option, the probability of which I will denote as α , or to move towards the foil, with probability $1 - \alpha$. If they do initially move towards the correct option, they may subsequently select this option, with probability β , or change direction and select the foil option instead, with probability $1 - \beta$. Finally, if they initially move towards the foil, they may still change direction and chose the correct option, with probability γ , or otherwise persevere to the foil option, with probability $1 - \gamma$. These probabilities are illustrated in Figure 2.4. Note that for each decision, the labelled probabilities (α , β , and γ) denote the probability of moving towards or selecting the correct option, while their compliments ($1 - \alpha$, $1 - \beta$, and $1 - \gamma$) denote the probability of moving towards or selecting the foil. Equation 2.2 summarises the definitions of each of these probabilities.

$$\begin{aligned}\alpha &= P(\text{Initially correct}) \\ \beta &= P(\text{Correct response} | \text{Initially correct}) \\ \gamma &= P(\text{Correct response} | \text{Initially incorrect})\end{aligned}\tag{2.2}$$

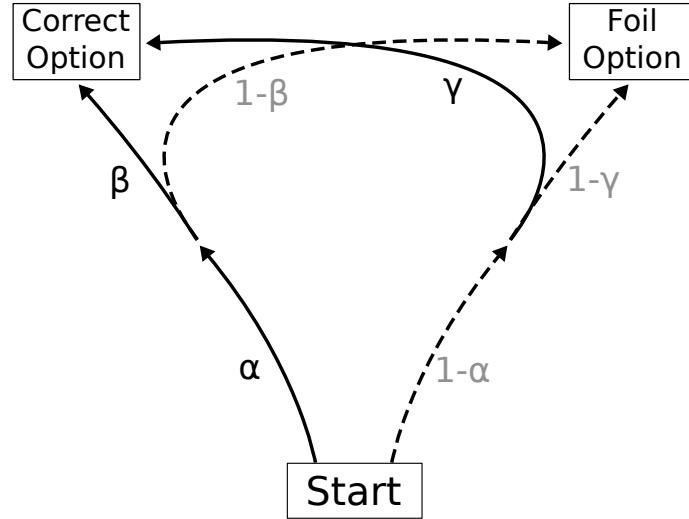


Figure 2.4: Mouse trajectories in a given condition can be described in terms of the probability of initially moving toward the correct option (α), the probability of choosing the correct option after initially moving towards it (β), and the probability of choosing the correct option despite initially moving towards the foil (γ).

2.2.3 TIME COURSE

Beyond the shape of cursor trajectories, we can analyse their time course. Unsurprisingly, there has also been considerable variation in how these time series data are used in the literature. Spivey et al. (2005) introduced the convention, still used in most studies, of standardising the time course of trajectories of different durations by interpolating them to 101 discrete time points, corresponding to the location of the cursor from 0% to 100% of the way through the trial. While this approach somewhat simplifies further analyses, it is problematic in the current research; when there is considerable variation in response latencies, the 50% point may be significantly later in one condition than another, and so it would be misleading to compare the location of the cursor at this point between conditions.

An alternative approach, which I use instead, is to use the real time information, specifically the location of the cursor every 20 msec through the trial. As standard computer mice typically update their position every 10-15 msec, this frequency strikes a balance between temporal precision and oversampling of the data.

Most mouse tracking studies include some form of plot showing the average cursor trajectories in each condition. However, for the experiments reported here — every one of which revealed discrete attraction effects — to do so would be mis-

leading. As noted by Spivey et al. (2005), if we average trajectories from conditions where some trials initially go towards the alternative option and others go straight to the correct one, or if we average trajectories from conditions where all trials were moderately curved, in both cases we end up with a mean trajectory that follows a smooth curve, like those shown in Figures 2.1 and 2.2. However, such an average trajectory would be a misleading description of the kind of data I find in this thesis, as few trials actually follow such a gently curving trajectory (see Figure 2.5).

Instead, I opted to match these discrete attraction effects with a discrete analysis. For every 20 msec window, from every trial, I classed the cursor as being either in its starting position, on the side of the screen containing the correct option, or on the side containing the foil. I then investigated the proportion of trials on each side of the screen, over time.

Beyond visualisation, researchers usually wish to perform some kind of inference with these data. A crude, but simple method of assessing when conflict occurs, again introduced by Spivey et al. (2005), is to conduct a series of t tests, usually on the 101 normalised time steps, but also possibly on the real-time data, comparing the x axis position between conditions at each time step, and noting the window for which the difference is significant. Although this approach does not provide a valid significance test for comparing conditions in general, it does provide an easily interpretable indication of *when* participants are drawn towards a competing response. It is possible, but rarely done in practice, to derive by means of simulations the likelihood of achieving a run of significant differences of length n under the null hypothesis (Dale et al., 2007), and thus calculate a valid p value for the difference between conditions overall. A similar analysis has also been reported calculating at each time step the distance between the cursor and the non-chosen response, and comparing this between conditions (Falke et al., 2013).

Again, though, the experiments reported here reveal discrete movements towards one or other response option, with reversals on some trials, rather than the graded, curved trajectories seen in much previous work. Therefore, an averaged trajectory is an inappropriate measure, as it falls somewhere between the direct trajectories and the reversals (see Figure 2.5). For this reason, in Experiments 1

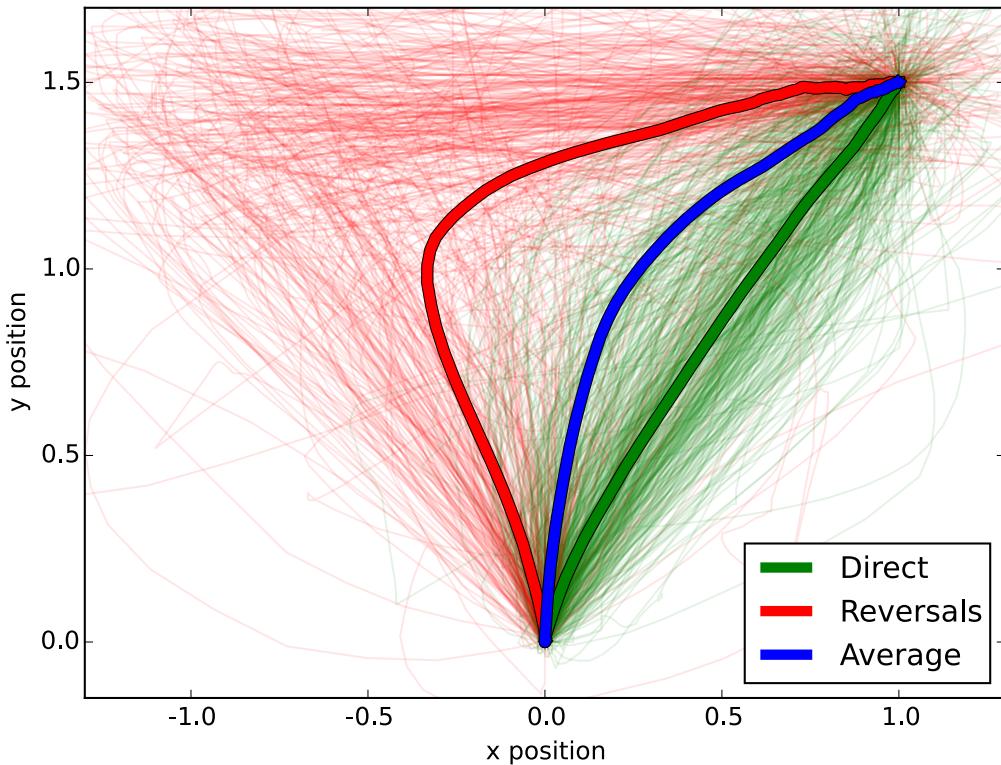


Figure 2.5: Averaging direct trajectories (green) and reversals (red) produces an aggregate trajectory that appears to curve gently. However, few of the actual trajectories curve in this way. These data come from Experiment 2 (Chapter 3).

to 5 I instead code the location of the cursor over time on each trial as being either on the side of the screen corresponding to the correct option, or the side corresponding to the foil (excluding cases where the cursor is yet to be moved). I then compared, in every 20 msec window, the probability of being on either side of the screen to chance, or the probability of being on one or other side between conditions, using logistic mixed effects models (see below). Attraction towards the foil option is therefore revealed by an increase in the proportion of trials where the cursor is on that side of the screen.

In line with previous mouse tracking studies, for each time course analysis I run a series of comparisons, one on each 20 msec time slice, and report the times from which significant effects emerge and disappear. Instead of t tests, however, I use logistic mixed models, for reasons outlined below.

Finally, an alternative to running a series of comparisons is to fit polynomial regression models, or *growth curves*, to the time course data (Falke et al., 2013;

see Mirman, 2014b, for an introduction to this method in eye tracking research). These are regression models in which time, and often polynomial terms such as time^2 and time^3 , are included as predictors. Such models are often used to model changes in subjects over time, such as the change in height of a child over a number of months, hence their name. Growth curves are discussed in detail in Chapter 6.

2.3 STATISTICAL ANALYSES

Finally, I will use the remainder of this chapter to outline some of the statistical methods used in this thesis. Again, by covering this material here, I hope to avoid repeating it throughout the later chapters.

2.3.1 MIXED EFFECTS MODELS

Throughout this thesis, I use *mixed effects* regression models (Baayen, Davidson, & Bates, 2008; A. Gelman & Hill, 2007) — otherwise known as linear or logistic mixed models (LMMs), hierarchical regression, multilevel regression, and random effects regression — to analyse my data. I will not provide a full tutorial on these methods here (see instead Baayen et al., 2008; Bates, Mächler, Bolker, & Walker, 2015; A. Gelman & Hill, 2007; Quené & van den Bergh, 2008). However, in this section I will explain my motivation in using these approaches, and provide sufficient introduction so that readers can follow the results presented later. Unless otherwise stated, all analyses were conducted using the R programming language (R Core Team, 2015), and mixed effects models were fit using the lme4 package (Bates, Mächler, et al., 2015).

A typical psychology experiment records several data points from each participant, one per trial. For illustration, I will use the case of an experiment where participants p_1, p_2, \dots, p_P perform a task in two conditions, c_1 and c_2 , and their response times rt are recorded on each trial. In a classical analyses, we would calculate the mean rt for each participant in each condition, and compare each participants' means from c_1 to those from c_2 using Student's t test. Furthermore, there will be individual differences in rt in general, as some participants generally re-

spond faster than others — this data is clustered within participants, in other words. Therefore, we can improve our statistical power by using a paired t test, which accounts for these individual differences in baseline rt .

However, aggregating data in this way is not always appropriate (see Baayen et al., 2008). Firstly, it discards a great deal of information, and as a result the statistical power of our tests can be diminished. Secondly, in averaging over multiple stimuli, we ignore any systematic clustering by stimuli (or items) in our data: some stimuli may produce faster responses than others in general, for example. In linguistics, where items usually correspond to specific words or phrases, this issue is traditionally addressed by running two analyses, one aggregating by participant, and one by item, and reporting the results of both (Clark, 1973).

A more serious problem with aggregation is that it can involve unbalanced data. In many analyses, we only include trials which meet certain criteria, for instance only analysing response times for correct responses. In reasoning research, where participants can often produce many different inferences, this is particularly common. If one participant gives few correct responses, their average response time will be estimated from only a few trials, and so will be a noisy measure. In the classical analysis, however, this datum is given as much weighting as an average response time from a participant who responded correctly on every trial. When, as often is the case, there is some relationship between, for instance, accuracy and response times, this can cause systematic biases in our analysis, hiding true effects and potentially creating spurious ones (Baayen et al., 2008).

Aggregation is also inappropriate when the dependent variable is not a continuous quantity such as rt . Most notably, when analysing proportions and percentages (e.g. accuracy), aggregate data systematically violate the assumptions of t tests/ANOVA, in a way which cannot be remedied by transforming the data (Warton & Hui, 2010). These issues with aggregation can be avoided, however, by fitting regression models to the data from each trial, rather than the aggregates.

As every undergraduate psychology student is taught, the standard (unpaired) t test, like the between-subjects ANOVA and many other classical tests, is a special case of linear regression. The regression equivalent of the unpaired t test for this

experiment is given as

$$rt_i = \alpha + \beta * Condition_i + \epsilon \quad (2.3)$$

rt_i is the rt on trial i , and is the sum of α (the mean rt in baseline condition c_1), $Condition_i$ (a dummy variable representing the condition on trial i , set to 0 to represent c_1 and 1 to represent c_2), β (the difference in rt between the conditions), and the normally distributed error term ϵ .

However, like the unpaired t test, this model does not account for individual differences in rt . To do this, we can extend Equation 2.3 so that the intercept α is allowed to differ for each participant. This is, in effect, a within-subjects regression model. The result is a *mixed effects model*, in that it includes both *fixed effects* — the overall intercept term α , and the overall effect of condition β — and *random effects* — the individual differences around α . Such a model is said to have *random intercepts for each participant*.

Mixed models are even more flexible than this, however. We may also want to allow for differences between each item in rt . This can be done by including *crossed random intercepts* (Baayen et al., 2008) for both participants and items, allowing α to vary for each participant, and for each item, while still calculating the overall intercept term. These random effects are not restricted to the intercept either. The β term reflects the difference between the two conditions. It may be that this difference is greater for some participants than others, or greater for some items than others, if the same items are used in both conditions. Therefore, we can also add random effects of condition for each participant, for each condition, or for both, as desired.

2.3.2 WHAT RANDOM EFFECTS?

At this point, it should be noted that there is still active debate as to what random effects should be included in a model. Barr, Levy, Scheepers, and Tily (2013) argue for the use of *maximal* random effects structures, where random effects are included wherever the design of the experiment allows. However, even with simple designs such a model cannot always be fit in practice, as the amount of data

required increases exponentially with the number of parameters to be fit. Instead, Bates, Kliegl, Vasishth, and Baayen (2015) advise using *parsimonious* models — fitting the maximal model, and then removing the parameters that account for least variation as necessary until the model can be fit. It is this approach I have taken in this thesis, and in each results section I specify the random effects structure used for the analysis.

2.3.3 DATA TRANSFORMATIONS AND BINARY OUTCOMES

Being regression models, mixed effects models can also make full use of the data transformations and link functions used in regression modelling. In this thesis, I use two such tools. First, it is expected that latency data, such as response times, movement initiation times, and in Chapter 5, reading times, will have a log-normal distribution: a distribution like a normal distribution that has been log-transformed. Therefore, to analyse these measures I simply log-transform them, and apply a standard linear mixed effects model. Second, several of the measures I analyse, including reasoning accuracy, and the number of reversal trajectories per condition, are binary proportions. These can be analysed by using logistic, rather than linear, mixed effects models.

2.3.4 CATEGORICAL PREDICTORS

Just as the ANOVA is a special case of linear regression, mixed effects models can be used to evaluate the main effect of a categorical predictor in these complex repeated-measures designs. In the classical ANOVA, we use the Sum of Squares¹ of models fitted with and without that predictor to compute an F statistic, which reflects the main effect of that predictor. However, with mixed effects the degrees of freedom (DFs) for this F statistic are not easily defined, and so it is difficult to calculate the appropriate p value. In mixed effects models, we therefore compare the deviance of models ($-2 \times$ their log-likelihoods) fitted with and without the predictor, and use a chi-squared test, with DF corresponding to the number of terms added by larger model (adding a categorical predictor with three levels adds 2 DF) to find a p value for that predictor's main effect.

¹More accurately, the sum of the squared deviations from the model fit.

2.3.5 REPORTING

Of course, the output of mixed effects regression models differs from that of a simple t test. In most analyses, I will be investigating the effect of one or more binary variables, such as condition, on the dependent variable. This is quantified by the regression slope (β) for that variable, and so for each comparison I will typically report the means and standard deviations in each condition, β and its 95% confidence intervals, reflecting the estimated difference between the conditions, and finally a t test comparing β to 0.²

For analyses where the dependent variable was log-transformed (e.g. response times), the exponentiated regression weight e^β is reported instead. This indicates the percentage change in the dependent variable between the conditions. For instance, if the mean response time in condition c_1 is 1,000 msec, and the effect of condition is $e^\beta = 110\%$, response times in condition c_2 are 10% greater than those of condition c_1 , for an average of 1,100 msec. e^β is also reported for the logistic analyses, where it reflects the multiplicative change in the odds of the outcome in question. For example, if participants gave the wrong response on 20% of trials in condition c_1 , the odds ratio of wrong to right responses in this condition is $\frac{20\%}{80\%} = 0.25$. An effect of $e^\beta = 1.5$ would mean the odds were one and a half times greater of doing so in condition c_2 ; odds of $0.25 * 1.5 = 0.375$, which correspond to a probability of $\frac{0.375}{1+0.375} = 27.3\%$.

² In linear regression, the DF for this test would simply be the number of data points minus the number of variables. When random effects are included, it is not possible to define the DF in this way. In this thesis, I take the conservative approach of using Satterthwaite's approximation (Satterthwaite, 1946) for the DF, as implemented in the lmerTest (Kuznetsova, Brockhoff, & Christensen, 2015) package for R. For logistic mixed models, the t test is replaced by a z test, and so it is not necessary to calculate DF.

CHAPTER 3

Perceptual Similarity vs. Conceptual Knowledge in Induction

3.1 INTRODUCTION

In Chapter 1, I discussed how people can draw on different sources of information in induction. In this chapter, I report a pair of experiments in which two such sources — perceptual similarity, and conceptual knowledge in the form of category membership — are placed into conflict. This was done using a mouse tracking version of S. A. Gelman and Markman’s (1986) inductive triad task. In Experiment 1, participants learned a biological property of different species in the natural world, and were asked to project each property to one of two species, one of which belonged to the same taxonomic group as the base species. In the conflict condition, but not the control condition, the base and the foil species looked alike, and so perceptual cues conflicted with conceptual knowledge. In Experiment 2, participants completed a similar task, but using artificial categories, learned at the start of the experiment. I also manipulated the nature of the properties participants reasoned about in Experiment 2, between participants, to reveal how this modulated participants’ inferences.

As discussed in Chapter 1, theories of inductive reasoning can be organised into two groups. One class of theory relies on conceptual knowledge about the categories to which different entities belong, and the relations between these categories. Category membership is perhaps the most fundamental such form of con-

ceptual knowledge that we can use as the basis for induction. According to contemporary accounts of categorisation (see Murphy, 2004; Rosch, 1988), a category consists of a set of entities that have attributes in common. Therefore, if we know the category to which something belongs, we may believe that it has attributes that we know are found in other members of that category (Murphy, 2004; Osherson et al., 1990). That adults can reason in this way has never been in question, and indeed this idea is central to the very notion of categorisation (S. A. Gelman & Markman, 1986; Mill, 1856). This principle applies equally to inferences about people, where it is better known as *stereotyping* (Greenwald & Banaji, 1995; Oakes, Alexander, & Turner, 1994). Strikingly, people rely on categorical information even when it is disadvantageous to do so. Murphy and Ross (Malt, Ross, & Murphy, 1995; Murphy, Chen, & Ross, 2012; Murphy & Ross, 1994, 2010) showed, for instance, that when it is not certain to which category an entity belongs, participants nevertheless reason on the basis of it belonging to the most likely category. Mishra and Mishra (2010) also report a *border bias* phenomena: distant locations in the same state (for instance New York City, and Buffalo, New York) are perceived as physically closer, and more likely to share properties, than closer locations divided by state lines (New York City and Princeton, New Jersey).

Category membership is also at the core of many more sophisticated theories of induction (i.e. Griffiths & Tenenbaum, 2009; Kemp & Tenenbaum, 2009; Osherson et al., 1990), which combine information about category membership with knowledge about the relationships between various categories.

Other theories of induction are not based on category membership. Similarity between entities is the most common kind of non-categorical information proposed to underlie induction: entities that are similar (i.e. that share many features) are more likely to also share a new feature than things that are dissimilar. This principle forms the basis of many feature-based accounts of induction (Fisher et al., 2015; Rogers & McClelland, 2004; Sloman, 1993; Sloutsky & Fisher, 2004a), discussed in Chapter 1. Fisher et al. (2015) draw a distinction between two kinds of similarity, based on either overlapping perceptual cues, or on shared features in our mental representations. With regard to adults' inferences, the focus has been on the latter, *representational* similarity. While it is usually assumed that adults

are not swayed by inappropriate perceptual cues, overlapping features in our mental representations of entities form the basis of both Sloman's (1993) Feature Based Induction model, and Rogers and McClelland's (2004) Semantic Cognition accounts of induction. Sloutsky and Fisher (2004a, 2008) propose an analogous model to explain children's inferences, based on perceptual similarity.

Given the importance of conceptual knowledge such as category membership in induction, why then might perceptual similarity play a role? One reason is that similarity is a useful proxy for shared category membership: categories are collections of things that share properties, including visible features, and things which look alike tend to belong to the same category. Furthermore, regardless of category membership, attributes tend to be correlated in the real world: things that share properties we know of are more likely to also share novel properties (see, e.g., Kemp, Shafto, & Tenenbaum, 2012). For both of these reasons, under many circumstances similarity — even perceptual similarity — and conceptual knowledge will support the same inferences. However, there do exist problems for which these two kinds of information disagree, most notably when reasoning about entities that look more like members of a different category than members of the category to which they belong (that is, visually *atypical* category members). The archetypal examples of such atypical entities are whales, which are mammals, but bear closer resemblance to fish than to other mammals.

This chapter, like the rest of this thesis, focuses on adults' reasoning. However, there has been little prior research on the role of simple perceptual cues in adults' inductive reasoning, perhaps because it would appear self-evident that adults rely on conceptual knowledge, or at least representational similarity. There has been extensive debate, however, as to whether young children make use of conceptual knowledge at all, or if their inferences are simply driven by perceptual similarity.

There are a number of reasons to believe that children may rely on perceptual cues, rather than conceptual knowledge, in induction. Conceptual knowledge must be acquired through instruction and experience, and so early inferences, by which infants and children make sense of the world, must be driven by a simpler, associative process like visual similarity (French, Mareschal, Mermilliod, & Quinn, 2004; Westermann & Mareschal, 2013). In recent years, it has also become apparent that

simple associative mechanisms can produce inferences that are surprisingly complex and flexible (Hinton & Anderson, 2014; Sloutsky & Fisher, 2008). Indeed, *deep neural networks* — neural networks, using simple associative principles but with many layers — are at the forefront of modern machine learning and artificial intelligence research (Hinton, Osindero, & Teh, 2006; Mnih et al., 2013).

Sloutsky (i.e. Sloutsky & Fisher, 2004b, 2008; Sloutsky et al., 2007, see Sloutsky, 2003, 2010 for reviews) has therefore argued that young children’s reasoning is often based on such perceptual cues, with reliance on conceptual knowledge arising later in development. In contrast, Gelman (i.e. S. A. Gelman & Davidson, 2013; S. A. Gelman & Markman, 1986; S. A. Gelman & Waxman, 2007; Rhodes & Gelman, 2009, see S. A. Gelman, 2004; S. A. Gelman & Meyer, 2011 for reviews) argues that even toddlers rely on conceptual knowledge about the world, and naive theories (Carey, 2009; Gopnik, 2003) to make inferences about the world around them.

S. A. Gelman and Markman’s (1986) triad task, discussed in Chapter 1, has been widely used as a testing ground for these competing accounts. Recall that in this task, children were presented with images of two entities (a flamingo and a bat, for instance), given their labels (“bird” and “bat”), and told a property of each (“This bird’s legs get cold at night”; “This bat’s legs stay warm at night”). They were also shown a third species (a blackbird in this example, labelled “bird”, but which more closely resembled the bat) and asked to indicate which property it would have (“Do this bird’s legs get cold at night, like this bird’s, or stay warm at night, like this bat’s?”). S. A. Gelman and Markman (1986) report that children as young as four predominantly resist the perceptual cue, and reason based on shared category membership – projecting a property from the flamingo to the blackbird, in this case. On the other hand, Sloutsky et al. (2007) present results with artificial categories that seem to show children ignoring category membership in favour of perceptual similarity. S. A. Gelman and Davidson (2013), however, demonstrate that this result only holds under certain circumstances, specifically, when the nature of the learned categories is obscure and not obviously relevant to the properties being projected (i.e. a creature’s ratio of buttons to fingers).

In Chapter 1, I introduced Bright and Feeney’s (2014b) hybrid theory of induc-

tion. This proposes that when reasoning inductively, we can draw on either simple, associative knowledge — such as similarity — or structured knowledge about the relationships between entities and the categories they belong to. The distinction in question in the current chapter, between perceptual cues and simple conceptual knowledge, is perhaps a more fundamental one than that made by Bright and Feeney (2014b) between associative and structured knowledge. Nevertheless, their hybrid account could easily be extended to include it.

A prediction that emerges from this extended version of the hybrid theory is that even adults' inferences may be influenced by perceptual similarity. To date, there has been little evidence suggesting that adults' inferences are driven by perceptual cues. However, as discussed in Chapter 1, this may in part be because methods used in previous work were poorly suited to revealing such conflict, and so in this chapter I explore this question using the mouse tracking paradigm. However, previous studies using this triad task (i.e. S. A. Gelman & Davidson, 2013; S. A. Gelman & Markman, 1986; Sloutsky et al., 2007) both with adult and child participants, have only included trials in which perceptual similarity and conceptual knowledge have conflicted. Typically, adults overwhelmingly respond on the basis of conceptual knowledge on such trials, while the debate has focused on whether children's responses are driven by conceptual knowledge (S. A. Gelman & Davidson, 2013; S. A. Gelman & Markman, 1986) or perceptual similarity (Sloutsky et al., 2007). By not including trials in which both cues agree, however, these experiments do not make it possible to see if adults' reasoning was influenced by perceptual cues in addition to conceptual knowledge: when perceptual cues disagree with conceptual knowledge, participants may be less likely to give the conceptually-cued response, or slower to do so, or be otherwise conflicted, than when the cues agree.

Therefore, in this chapter, I present two experiments, with adult participants, in which perceptual cues either conflict or agree with conceptual knowledge. If adult induction is in part driven by perceptual cues, participants should be less likely to generalise from a base entity to a conceptually-related one when the alternative, foil entity is visually similar to the base. Recording participants' mouse cursor trajectories, it is also possible to investigate how these kinds of informa-

tion interact over time. If participants are driven by *either* perceptual similarity or conceptual knowledge on each trial, they should move the mouse directly to one or other response. Alternatively, if participants are initially driven by quickly-available perceptual cues, but draw on conceptual knowledge later in the reasoning process, we may see reversal trajectories, as participants initially move towards the perceptually-cued option, and change direction mid-trial.

3.2 EXPERIMENT 1

In Experiment 1, participants completed a mouse tracking version of S. A. Gelman and Markman's (1986) triad task, with natural categories. In the original version of this task, children had to choose between generalising a property from a base species to another conceptually-related species, or to an unrelated but perceptually similar one. This experiment is, to my knowledge, the first to include a control condition, in which both perceptual and conceptual information cue the same response. Using this experimental manipulation, it is possible to explore the role these perceptual cues play in reasoning. Additionally, by recording participants' mouse cursor movements, the experiment provides a window into processes during reasoning, rather than just the final inferences participants make.

3.2.1 METHOD

PARTICIPANTS

Fifty nine undergraduate students took part in exchange for course credit.

STIMULI & PROCEDURE

At the start of the experiment, participants were presented with a framing story, based on those used by Sloutsky et al. (2007) and S. A. Gelman and Davidson (2013). This introduced a boy, Mark, who had moved to a country called Elbee, where his teacher was teaching him about the plants and animals found there. They were instructed that they would be told some of the facts that Mark's teacher had told him, and shown pictures of three animals. They were also instructed that

their task was to decide, given a fact about one species, which of the two other species this fact was likely to be true for. This was illustrated using an example.

Participants then completed ten induction trials, in random order. Induction stimuli were similar to those used by S. A. Gelman and Markman (1986), but were created for the current experiment. They consisted of ten sets of species of plants and animals. The full set of stimuli used can be found in Appendix C. Each set was made up of a *base* species, a *correct response* species belonging to the same super-ordinate category as the base, and two *foil* species, belonging to a different category: one that was intended to be perceptually similar to the base, and one that was not. Each species was represented in the experiment by a colour photograph (see Figure 3.1, and Appendix C). For each participant, five stimuli sets were randomly chosen as conflict trials, and presented with the foil that was perceptually similar to the base. The remaining sets were designated control trials and presented with the foil dissimilar to the base.

On each trial, the base was presented in the centre of the screen, with a property of that species shown above the image. Blank genetic properties were used, of the form “This one has gene 4ew. What else do you think has gene 4ew?”. After clicking a button marked “NEXT” in the bottom centre of the screen, images of the two possible response species appeared in the top left and right corners, with their positions randomised on each trial (Figure 3.1). Participants responded by clicking on one or other image, and the position of the mouse cursor was recorded as they did so.

After the reasoning trials, participants completed a post-test check. This was to ensure that participants possessed the appropriate structured knowledge about the relationships between the species used. Each base species was presented twice, once alongside its corresponding correct response, which belonged to the same biological category, and once alongside its perceptually similar foil, which did not. The left-right positioning of these images was randomised for each trial. Participants were instructed to indicate if each pair of species belonged to the same biological group. The order of presentation of the post-test stimuli was totally randomised.

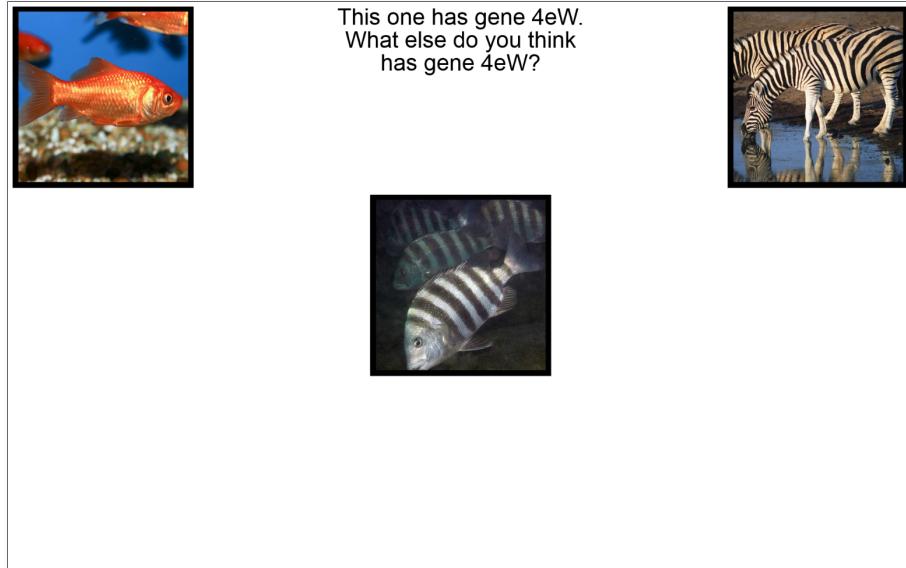


Figure 3.1: Screen shot of conflict trial from Experiment 1. The base image, a striped fish, belongs to the same category as the correct response option, the goldfish shown in the top left, but is perceptually similar to the foil option, the zebra in the top right.

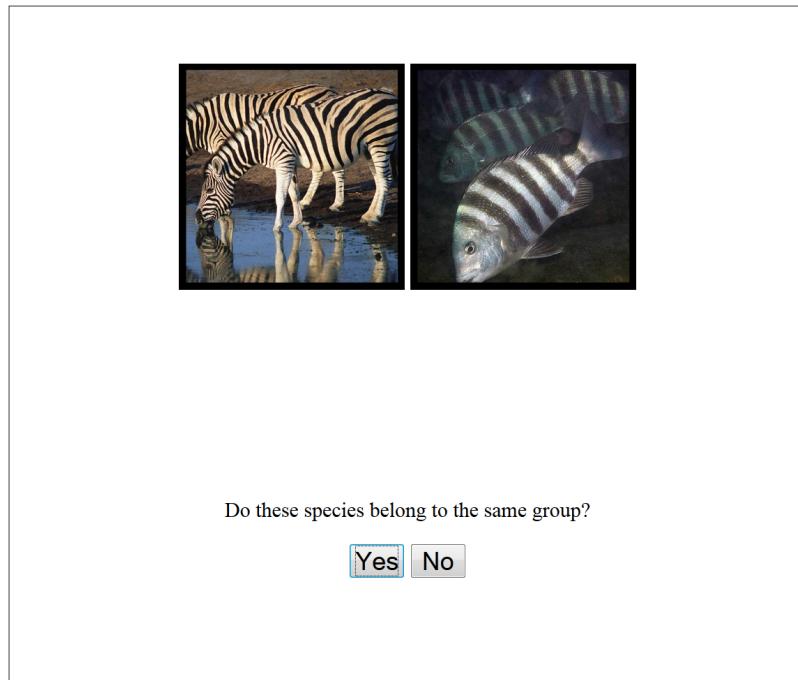


Figure 3.2: Screen shot from the post-test check, Experiment 1.

3.2.2 RESULTS

On the post-test check, for each stimuli set participants achieved at least 91% accuracy in correctly identifying that the base and correct response option belonged to the same biological category, and at least 68% accuracy in correctly identifying that the base and the foil option *did not* belong to the same category (binomial tests by participant, $N = 59$, p 's $< .0075$; minimum accuracy required for $p < .01$ is $40/59$, or 67.8%). Post test scores for each stimulus set are included in Appendix D. Therefore, data from all stimulus sets were included in the analyses.

I analysed the data using linear or logistic mixed models, with random intercepts for each participant and each stimulus set, and random coefficients for the effect of condition for each participant (see Chapter 2). Participants selected the foil species on 4% of trials in the control condition, and 40% of trials in the conflict condition; $e^\beta = 209.1$, CI = [3.6; 12169.2], $z = 2.577$, $p = .0100$). Therefore, perceptual similarity had a robust effect on their responses, leading them to generalise the property to a foil species belonging to a different biological group, rather than to the base species, considerably more often when the base and foil looked alike.

It is worthwhile assessing the pattern of individual differences in this effect. Figure 3.3 shows the number of foil responses each participant gave, by condition. Forty nine participants were more likely to select the foil under conflict, eight never selected it, and two showed the reverse effect, selecting the foil less often under conflict. Therefore, it appears that the effect of visual similarity holds across almost all of the participants.

The time taken to begin moving the mouse cursor (initiation time; IT), and to select a response option (response time; RT) both had positively-skewed distributions, and so were log-transformed before analysis. For correct responses, there were no significant differences in IT between conditions, taking 976 msec under conflict ($SD = 1,000$) and 881 msec in the control condition ($SD = 772$; $t = 0.104$, $p > .9$). However, RT was significantly slower for conflict trials (2,668 msec, $SD = 1,564$) than control trials (2,112 msec, $SD = 1,279$; $e^\beta = 122.9\%$, CI = [114.2%; 132.3%], $t(58.5) = 5.553$, $p < .0001$).

Maximum Deviation (MD) was bimodally distributed (Bimodality Coefficient

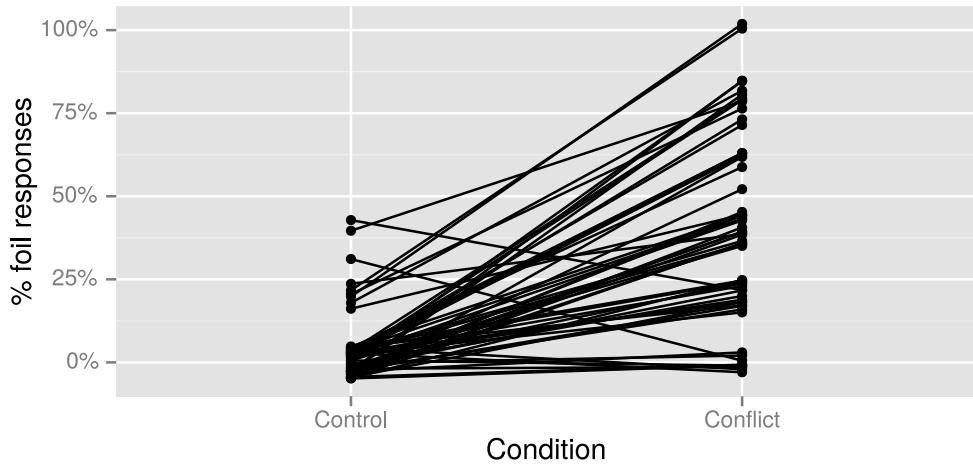


Figure 3.3: Proportion of foil responses given by each participant, per condition, in Experiment 1.

Table 3.1: Summary statistics for correct responses in each condition. Standard deviations are shown in parentheses. Correct responses under conflict took significantly longer, and were more likely to be classed as *reversals*. There was no difference in movement initiation times.

Condition	Response time	Initiation time	Reversals
Control	2,112 (1,279)	881 (772)	19%
Conflict	2,668 (1,564)	976 (1,000)	43%
p	< .0001	> 0.9	< .0001

= .667; Hartigan's D = 0.025, N = 453, p = .0058), with trajectories either going directly to the selected response, or moving towards one response, and changing direction mid-flight (see Appendix B). Trajectories were therefore classed as either *direct* or *reversals*, as described in Chapter 2. When selecting the correct species, participants were significantly more likely to initially move towards the foil option on conflict trials (37.7%) than on control trials (15.9%; $e^\beta = 4.8$, CI = [2.5, 9.3], $z = 4.699$, $p < .0001$).

As discussed in Chapter 2, there were therefore four kinds of mouse trajectories observed in this experiment: direct movements to the correct option (*Direct Correct*), initial movements towards the foil, which changed direction to the correct option (*Reversal Correct*), initial movements towards the correct option, which changed direction to the foil (*Reversal Foil*), and direct movements to the foil (*Direct Foil*). Table 3.2 shows how common each trajectory type was in both conditions, and indicates substantial increases in both incorrect responses and reversal

Table 3.2: The four different possible kinds of cursor trajectory, broken down by condition.

	Direct Correct	Reversal Correct	Reversal Foil	Direct Foil
Control	77.5%	18.7%	1.7%	2.1%
Conflict	34.4%	25.8%	8.2%	31.6%

correct trials, and a decrease in direct correct trials, under conflict.

These different kinds of trial can be better understood in terms of their *transition probabilities* (see Chapter 2). At the beginning of a trial, participants either move towards the correct option, with probability α , or towards the foil, with probability $1 - \alpha$. After initial movement towards the correct option, they can either select this option, with probability β , or the foil, with probability $1 - \beta$. Finally, after moving towards the foil, participants can either change direction and give the correct response, with probability γ , or persevere and select the foil option, with probability $1 - \gamma$. Figure 3.4 shows a schematic of these probabilities. Table 3.3 show each of them across both conditions. Again, these probabilities were compared using logistic mixed models, with random intercepts for each stimulus set, and random intercepts and slopes for each participant.

Participants were significantly more likely to initially move towards the correct response in the control condition (79% of trials) than the conflict condition (43%; regression $e^\beta = 7.3$, CI = [4.7; 11.3], $z = 8.773$, $p < .0001$). After doing so, they were also more likely to then select the correct response in the control condition (98%) than the conflict condition (81%; regression $e^\beta = 25.9$, CI = [6.6, 101.2], $z = 3.078$, $p = .0021$).¹ Finally, on trials where they did initially move towards the foil, participants were more likely to eventually select the correct species in the control condition (90%) than in the conflict condition (45%; regression $e^\beta = 26.6$, CI = [7.0; 101.2], $z = 4.815$, $p < .0001$).

Finally, if the processes drawing participants towards selecting the foil on conflict trials operate early in reasoning, we would expect movements towards this op-

¹ As the value for the control condition was close to 1, regression parameters here would normally approach ∞ , and the model would be unidentifiable, a phenomena known as *perfect separation* (Albert & Anderson, 1984). I therefore used penalised-likelihood estimation to impose an uninformative prior distribution (Zorn, 2005) on the regression parameters, effectively reflecting a belief that when observed probabilities are close to 0 and 1, the true probabilities are likely to be higher, or lower, respectively. In this, and all future cases of perfect separation, I impose a Gaussian prior on the regression weights, with mean 0 and standard deviation 3.

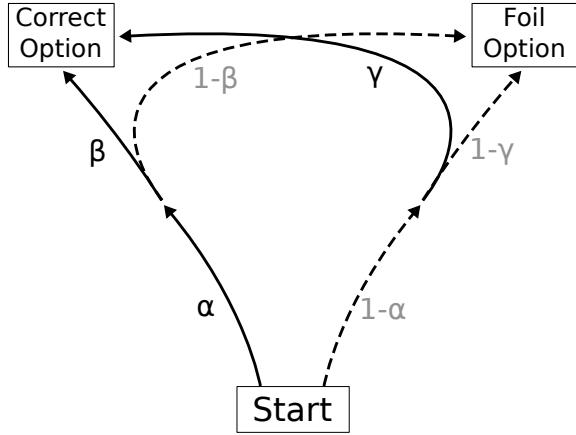


Figure 3.4: The possible cursor transitions that can occur during a trial.

Parameter	Control	Conflict
α	79%	43%**
β	98%	81%**
γ	90%	45%**

Note: ** $p < .01$.

Table 3.3: Transition probabilities for Experiment 1.

tion to be initiated earlier than those towards the correct species. I found this to be the case, with initial movements towards the foil initiated after 687 msec on average ($SD = 599$), and those towards the correct species initiated after 1,182 msec ($SD = 1,119$). Fitting a mixed model with random intercepts for each participant and each stimulus set, this difference was found to be significant; $e^\beta = 130\%$, CI = [120%, 150%], $t(204.6) = 4.042$, $p < .001$).

TIME COURSE

In order to investigate *when* participants' mouse cursor movements were drawn towards each response option, I examined the position of the cursor over the first 4 seconds of each trial. As participants largely moved discretely to one or other response, these positions were not normally distributed, and so I coded the location of the cursor according to whether or not it was on the foil species' side of the screen. By doing so, it is possible to identify at what points in time participants' mouse movements are guided by perceptual similarity, which drives participants towards selecting the foil species on conflict trials (see Chapter 2).

Figure 3.5 shows the proportion of trials on the side of the screen corresponding to each response option. I fitted two series of logistic mixed models, predicting the probability of being on the foil's side of the screen, and one the probability of being on the correct species' side of the screen. Each series consisted of a set of models, one fitted to every 20 msec time window, with condition as a predictor, and

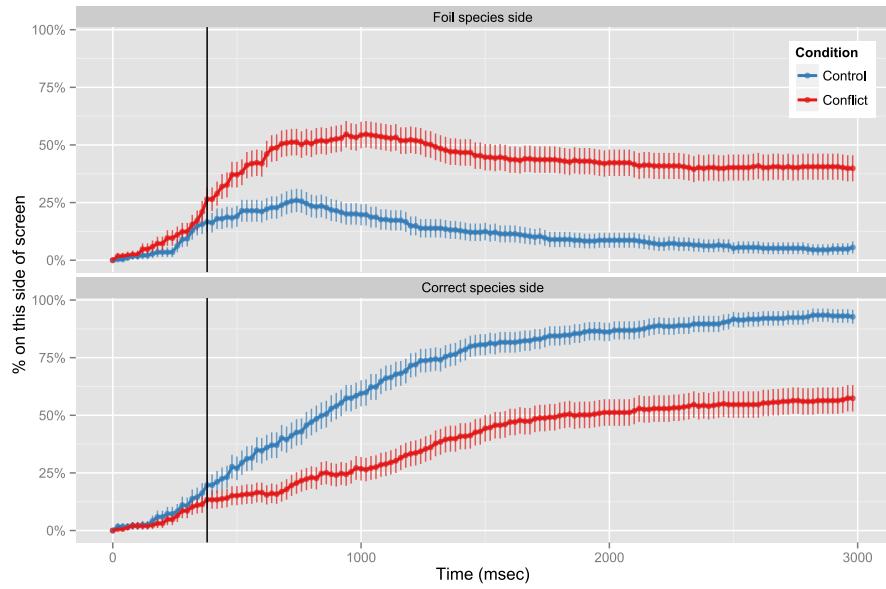


Figure 3.5: Time course of attraction towards each response option. Vertical lines show the points from which the difference between the two conditions is statistically significant ($p < .05$). Error bars show 95% confidence intervals.

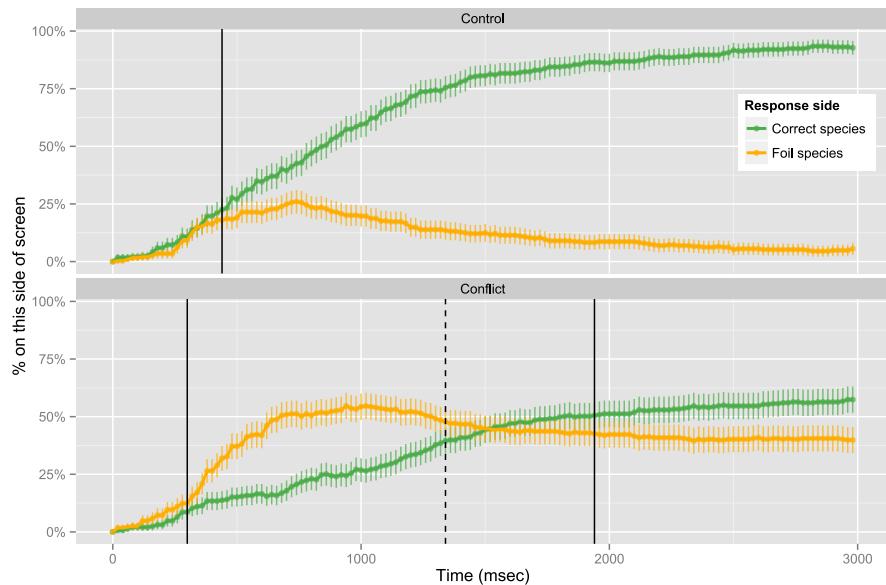


Figure 3.6: Time course of attraction towards each response option. Solid lines show the points from which one response is significantly more likely to be moved towards than the other ($p < .05$), while dashed lines for the points from which these differences are no longer significant. Error bars show 95% confidence intervals.

random intercepts for each participant, and each base species. I used these series of models to find the divergence points: the time from which participants were significantly more likely to move towards the foil species in conflict trials than control trials, and the time from which they were significantly more likely to move towards the correct species in control trials than conflict trials. In both cases, these significant differences between the conditions emerged from 380 msec onwards.

Figure 3.6 shows the same data, but with separate lines for each side of the screen, and separate facets for each condition. In the control condition, participants were significantly more likely to move towards the correct option than the foil from 440 msec onwards. In the conflict condition, however, we see the interaction of the two competing cues. Participants were significantly more likely to move towards the foil option from 300 msec. This preference for the foil persisted until 1,340 msec, at which stage both response options were equally popular. From 1,940 msec, finally, participants were more likely to move towards the correct option than the foil, as the initial attraction towards the foil option has been largely inhibited. In other words, participants were initially drawn towards the foil species on conflict trials, but were later drawn towards the correct species instead.

3.2.3 DISCUSSION

This experiment placed perceptual similarity and conceptual knowledge in conflict in an inductive reasoning task. My first goal in doing so was to test the prediction, based on Bright and Feeney's (2014b) hybrid theory, that adults' inferences would be influenced by perceptual cues, as well as conceptual knowledge. This prediction was confirmed by participants' responses: when conceptual knowledge did not conflict with perceptual similarity, participants selected the conceptually-cued response on almost every trial. When the foil species looked similar to the base, however, participants chose the foil 40% of the time. Additionally, this effect was found for the vast majority of participants, and so is not the result of a few participants who are inappropriately swayed by these visual cues. When participants did give the correct response, it appears that they experienced conflict when perceptual cues supported the foil response, taking significantly longer to respond, and being more likely to initially move towards the perceptually-cued foil option.

The mouse tracking paradigm makes it possible to go beyond this finding, however, and work towards a description of how these two sources of information interacted during reasoning. In Chapter 1, I raised two possibilities: participants may either selectively use one or other cue, or they may use perceptual cues by default, but reject them in favour of conceptual knowledge when they realise those cues to be inappropriate. Consistent with both possibilities, on control trials, where the base and the foil did not look alike, participants initially moved towards the foil 21% of the time. On conflict trials, where they did look alike, participants initially moved towards the foil 57% of the time. Furthermore, these movements were initiated faster than movements towards the correct option in the same condition, suggesting either that conceptual knowledge takes longer to utilise than perceptual cues, or that perceptual cues must be inhibited on these trials.

These two possibilities can be distinguished by looking to what participants do when they have initially moved towards a perceptually-cued foil. In this situation, participants changed direction on 45% of trials to select the correct option instead. By comparison, participants who initially moved towards the correct option under conflict only changed direction to select the foil on 19% of trials. Therefore, it appears that, at least some of the time, reasoning on the basis of conceptual knowledge required the inhibition of the response based perceptual cues. Together, these results suggest that perceptual cues are used by default on this task. By this view, participants can either initiate an early movement, driven by these cues (57% of the time), and subsequently either inhibit this response (45% of the time), or follow through with it and select the foil species. Alternatively they can override these cues before they move the cursor (43% of the time), and move directly towards the correct option.

Analysis of the time course data also appears to support this interpretation. In Figure 3.6 we see that on conflict trials, participants were initially (300 to 1,340 msec) more likely to be on the foil side of the screen than the correct side. By 1,940 msec, however, this trend had reversed, and participants were instead more likely to be on the side of the correct option.

An interesting question, which this experiment does not answer, is what factors dictate whether or not participants moving towards the foil ultimately select

this option, or instead override this initial movement and select the correct species. This issue has come under increased scrutiny lately in the dual process literature (see, for instance, De Neys, 2012; Pennycook, Fugelsang, & Koehler, 2015; Thompson & Johnson, 2014; Thompson, Prowse Turner, & Pennycook, 2011), where the focus has been on whether participants reflect on their intuitive responses. In the induction literature, S. A. Gelman and Davidson (2013) report a series of experiments with children using the triad task. They show that children are more likely to use conceptual knowledge, rather than perceptual similarity, when the conceptual categories used differed at a high, ontological level, (animals versus robots), than when they differed at a lower level (kinds of dogs, or creatures categorised according to their ratio of fingers to chest buttons, as used by Sloutsky et al., 2007). They also demonstrated that children are more reliant on conceptual knowledge when the properties under consideration are meaningfully related to the different categories, for instance animals being warm blooded, or robots containing batteries. Both of these effects make sense from a normative perspective: categories which are more conceptually distinct provide a better basis for induction (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), and certain kinds of category are conducive to the projection of certain kinds of property, for instance biological properties within biological categories (Heit & Rubinstein, 1994; Shafto, Coley, & Baldwin, 2007; Shipley, 1993). We would expect these same factors to influence adults' inferences. However, it is not clear at what point in the reasoning process such variables are important: they may prompt reasoners to attempt to draw on conceptual knowledge in the first place, or they may cause them to be more likely to inhibit their initial inappropriate perceptually-driven responses.

In Experiment 2, I presented participants with a version of this task using artificial categories, specifically, a kind of animal, and a kind of robot. I also manipulated, between participants, the nature of the kind of properties being considered, in order to investigate the effect this has on participants' reasoning.

3.3 EXPERIMENT 2

In Experiment 1, perceptual similarity drove participants' mouse movements early in reasoning, with conceptual knowledge being drawn on later in the process, and

in some cases overriding these perceptual cues. In Experiment 2, I attempted to replicate these finding using artificial categories — a kind of animal, and a kind of robot — that participants learned in the laboratory. I also manipulated, between participants, the nature of the properties to be reasoned about, to be either *specific properties*, saliently related to the two categories — a robot having batteries, or an animal sleeping at night, for example — or fictional *generic properties* — such as being able to make a “zevy sound”. S. A. Gelman and Davidson (2013) showed that children reasoning about specific properties were more likely to draw on conceptual knowledge, presumably as such properties make this knowledge more salient or accessible. Here, this manipulation provides a window into how perceptual cues and structured representations interact during this task.

3.3.1 METHOD

PARTICIPANTS

Forty eight participants completed the experiment for course credit.

STIMULI & PROCEDURE

Stimuli were adapted from drawings by LaRiccia (2005) used by Sussman, Khemlani, and Oppenheimer (2014), and consisted of cartoon pictures of different kinds of creatures. I created two categories: animals called “Flurps”, and robots called “Floobits”. There were four possible body shapes, and four possible colourations, and each body shape and colouration was possible for both Flurps and Floobits (Figure 3.7, see also Appendix E).

Participants were again presented with a framing story about Mark, who had just moved to Elbee, but this time told that Mark was learning about different kinds of things found outdoors there. Two kinds of thing found in Elbee were presented, animals called “Flurps”, and robots called “Floobits”, designed to closely resemble Flurps. Participants were told that both Flurps and Floobits had different body types and came in different colourations, with the same body types and colourations found in both categories. They were also informed that Flurps and Floobits could be differentiated by looking at their heads: Flurps had animal

heads, with tentacles on top, while Floobits had robotic heads (see Figure 3.7).

After these instructions, participants completed eight categorisation trials, with feedback, in which they were presented with four Flurps, and four Floobits, in a random order, and asked to categorise each. These were intended to emphasise the distinction between the categories. After the categorisation trials, participants were told that for the final part of the experiment they would be told a fact about one Flurp or Floabit, and asked to decide which of two others this fact was most likely to also be true for. They were then presented with the sixteen induction trials, in random order. Half of the participants reasoning about *generic* properties that were unrelated to the two categories (i.e. “This one can make a zevy sound”). The remainder reasoned about *informative* properties, so that Flurps were presented with properties specific to animals (“This one has a mummy”), while Floobits were presented with robot-specific properties (“This one has batteries inside”). The full list of properties used can be found in Appendix F.

There were sixteen induction trials, eight with Flurps as bases, and eight with Floobits, and the response options always consisted of one Flurp and one Floabit. Half of these were control trials, in that the response entity belonging to the same category as the base (the correct response) was perceptually identical to the base, having the same body type and colouration, while the other, foil response entity was perceptually different, with a different body type and colouration. The remainder were conflict trials (Figure 3.7), where the correct response was perceptually different from the base, and the foil response was perceptually similar.

3.3.2 RESULTS

Unless otherwise specified, I analysed the data using linear or logistic mixed models, with random intercepts for each participant, and random coefficients for the effect of condition for each participant.

Participants reasoning about generic properties gave the foil response on none of the control trials ($N = 190$), but on 46% of the conflict trials. Those reasoning about specific properties gave the foil response on 4% of control trials, and on 17% of conflict trials (see Table 3.4). A 2 (condition) x 2 (properties) logistic mixed



Figure 3.7: A trial from Experiment 2. Participants were told that the *Flurp* (centre) tries to stay warm, and asked to decide which of the other two things shown, the *Floobit* on the left, or the Flurp on the right, also tries to stay warm. This is a conflict trial: the Floobit has the same colouration and body type as the base, but its robotic head identifies it as belonging to a different category.

Table 3.4: Foil responses in Experiment 2.

Properties	Condition		
	Control	Conflict	Mean
Specific	4%	17%	10%
Generic	0%	46%	23%
Mean	2%	32%	

model² found a main effect of condition ($e^\beta = 95.7$, CI = [28.6; 320.0], $z = 7.403$, $p < .0001$), and a condition x properties interaction ($e^\beta = 74.2$, CI = [7.5; 730.4], $z = 3.691$, $p < .0001$), such that the effect of condition was more pronounced for participants reasoning about generic properties.

Figure 3.8 shows how this effect was distributed across participants. Recall that participants completed eight reasoning trials in each condition. Of the 22 participants reasoning about generic properties, 16 were more likely to give the foil re-

² Once again, as all participants gave the correct response in the control condition with generic properties, I used penalised maximum likelihood to set a normal prior, with mean 0 and SD 3, on the regression coefficients here (Zorn, 2005). This model also omitted the random coefficients for the effect of condition for each participant, due to convergence issues.

Table 3.5: Summary statistics for correct responses, by property type and condition, on the odds of participants selecting the foil option (foil choices), and summary statistics for correct responses: response time (RT), movement initiation time (IT), and number of reversal trajectories (Reversals).

Properties	Condition	Foil choices	RT (msec)	IT (msec)	Reversals
Specific	Control	4%	1,366 (964)	377 (254)	16%
	Conflict	17%	1,611 (1,162)	362 (181)	52%
Generic	Control	0%	1,547 (912)	444 (389)	11%
	Conflict	46%	2,121 (1293)	509 (576)	59%

sponse under conflict, none were more likely to do so on control trials, and 8 never gave the foil response. Of the 22 reasoning about specific properties, 11 were more likely to give the foil response under conflict, 2 more likely to do so on control trials, and 11 never chose the foil option. Therefore, as in Experiment 1, the main effect does not appear to be driven by an effect only a subset of participants.

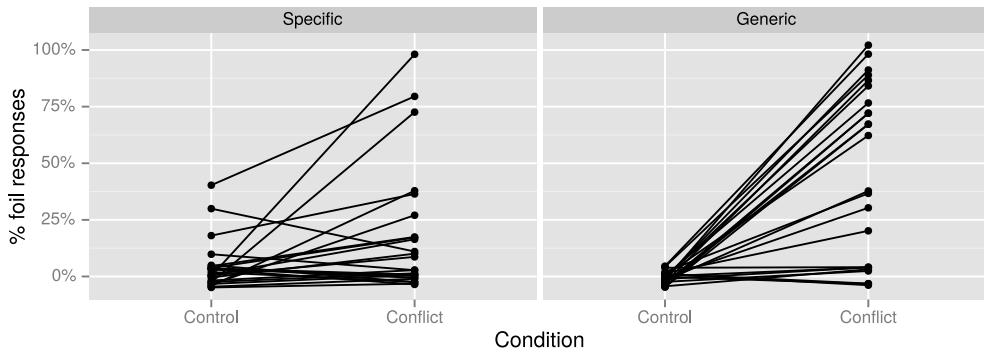


Figure 3.8: Proportion of foil responses given by each participant per condition, in Experiment 2. separately for each property type. A small jitter has been added to vertical positions to show points which would otherwise overlap.

As in Experiment 1, IT did not differ significantly between conditions or property types (p 's $> .2$). Response times were in general slower in the conflict condition (1,815 msec, SD = 1,239) than the control condition (1,458 msec, SD = 941; $e^\beta = 130\%$, CI = [120%; 140%], $t(186.54) = 7.057$, $p < .0001$), and slower for generic properties (1,747 msec, SD = 1,094) than specific properties (1478 msec, SD = 1,064; $e^\beta = 125\%$, CI = [106%, 146%], $t(46.1) = 2.731$, $p = .0089$). There was a significant condition x property interaction, such that the effect of condition was more pronounced for generic properties ($e^\beta = 119\%$, CI = [103%; 138%], $t(186.4) = 2.382$, $p = .0182$).

Table 3.6: The prevalence of each kind of mouse trajectory, by condition and property type.

Properties	Condition	Direct Correct	Reversal Correct	Reversal Foil	Direct Foil
Specific	Control	77.4%	18.9%	2.1%	1.6%
	Conflict	33.0%	49.7%	4.9%	12.4%
Generic	Control	85.8%	14.2%	0%	0%
	Conflict	20.6%	33.3%	12.2%	33.9%

MD was bimodally distributed (Bimodality coefficient = .635, Hartigan's D = 0.02, $p = .0341$), and so trajectories with MD greater than 0.747 were categorised as reversals (see Appendix B). For correct responses, these reversals occurred significantly more often for conflict trials (61% of trials) than for control trials (17% of trials; $e^\beta = 9.2$, CI = [6.0; 14.2], $z = 10.21$, $p < .0001$), with no effect of property type, or condition x property interaction.

There were therefore, once again, four kinds of cursor trajectories: Direct Correct, Reversal Correct, Reversal Foil, and Direct Foil trajectories. Table 3.6 shows the proportion of each trajectory type, broken down by condition and property type. Across both property types, participants were less likely to go straight to the correct option and choose it, and more likely to do so for the foil option, under conflict. In line with the analysis of reversals above, initial movements towards the foil, which redirect to choose the correct option, were more common under conflict, and this difference was more pronounced for specific properties, while this manipulation made participants reasoning about generic properties more likely to instead give the foil response.

We can better understand these effects, once again, by turning to the transition probabilities. Table 3.7 shows these transition probabilities, broken down by condition and by property type. I analysed these probabilities using 2 (condition) x 2 (properties) logistic mixed models, with random intercept terms for each participant. Initial movements to the correct response, with probability α , were significantly more common for control trials (82.6%) than conflict trials (32.3%; $e^\beta = 10.1$, CI = [7.0; 14.6], $z = 2.312$, $p < .0001$), but with no main effect of property type. A marginally significant condition x property interaction ($e^\beta = 2.0$, CI = [0.9; 4.0], $z = 1.904$, $p = .0569$) indicated that the effect of condition on α was

Table 3.7: Transition probabilities for Experiment 2. Regardless of property type, participants are more likely to initially move towards the correct option (α), and to ultimately select the correct option, either after initially moving towards it (β) or after initially moving towards the foil (γ) on control trials, where the correct option and the base look alike. On conflict trials, participants who initially moved towards the foil option were more likely to ultimately select the correct option instead (γ) when reasoning about specific properties than when reasoning about generic properties. Participants were initially less likely to move towards the correct option on conflict trials (blue), and after initially moving towards the foil on conflict trials, they were more likely to change direction and select the correct option if reasoning about specific properties than generic (red).

Condition	Properties	α	β	γ
Control	Specific	80%	97%	92%
	Generic	86%	100%	100%
	Both	83%	99%	96%
Conflict	Specific	38%	87%	80%
	Generic	33%	63%	50%
	Both	35%	76%	64%

slightly more pronounced for generic properties.

For trials where participants initially moved towards the correct option,³ the probability of ultimately selecting the correct option (β) was greater in control trials (98.7%) than conflict trials (75.8%; $e^\beta = 46.1$, CI = [11.6; 182.9], $z = 5.453$, $p < .0001$). There was no main effect of property type, but there was a significant property x condition interaction ($e^\beta = 47.1$, CI = [3.7; 601.5], $z = 3.692$, $p = .0030$), such that the effect of condition for generic properties was greater than that for specific properties. It should be noted, however, that β was largely positive. Even on conflict trials, with generic properties, participants who initially moved towards the correct option ended up selecting in 63% of the time.

For the trajectories that initially moved towards the foil, the probability of changing direction and ultimately selecting the correct option (γ), was lower for conflict trials (64%) than control trials (95%; $e^\beta = 34.4$, CI = [7.1; 166.2], $z = 4.402$, $p < .0001$). There was no main effect of property types, but there was, crucially, a property x condition interaction ($e^\beta = 66.6$, CI = [3.8; 1149.1], $z = 2.890$, $p = .0039$), indicating that the effect of condition was greater for participants reasoning about generic properties. Post-hoc comparisons showed no difference between par-

³ Again, in both the analysis of β and γ , as means of 0 and 1 occurred in a number of cells, penalised maximum likelihood logistic regression was used, with a Gaussian prior (mean 0, SD 3) on each regression coefficient.

ticipants reasoning about specific and generic properties in the control condition ($\gamma > 92\%$; $p_{adjusted} > .4$), but significantly more correct responses after initially moving towards the foil option on conflict trials when reasoning about specific properties (80%) than when reasoning about generic properties (50%; $e^\beta = 19.2$, $CI = [2.1, 174.5]$, $z = 2.622$, $p_{adjusted} = .0088$.

In Experiment 1, I found that initial movements towards the foil option on conflict trials were initiated more quickly than those towards the correct option. Here, I fitted a 2 (initial movement direction) \times 2 (property type) logistic mixed model, with random intercepts for each participant and each base species, and log-transformed initiation times from the conflict trials as the dependent variable. There was a significant effect of initial direction, such that movements towards the foil option (430 msec, $SD = 527$) were again initiated faster than movements towards the correct option (558 msec, $SD = 564$; $e^\beta = 118\%$, $CI = [1.07\%, 131\%]$, $t(346.2) = 3.325$, $p = .0010$). There was no effect of property type, or condition by property interaction ($p's > .1$).

TIME COURSE

I repeated the time course analysis reported for Experiment 1, looking at the proportion of trials on the side of the screen corresponding to each response option over time. Firstly, Figure 3.9 shows the proportion of trials on the side of the screen containing the foil option, over time, broken down by condition, and property type. To infer when each variable began to influence participants' motor output, I fitted a series of logistic mixed models, predicting the proportion of responses on the foil side of the screen at each 20 msec interval. To test the effect of condition, I fitted models with condition as a predictor, and random intercepts for each participant. To investigate the effect of properties, I fitted models with property as a predictor, and random intercepts for each participant, to the data from conflict trials only, as control trials were unaffected by manipulations of the property. Participants were more likely to move toward the foil option in conflict trials than control trials from 280 msec (solid vertical line in Figure 3.9). On conflict trials, participants reasoning about generic properties were more likely to remain on the foil side of the screen than those reasoning about specific properties from 620 msec

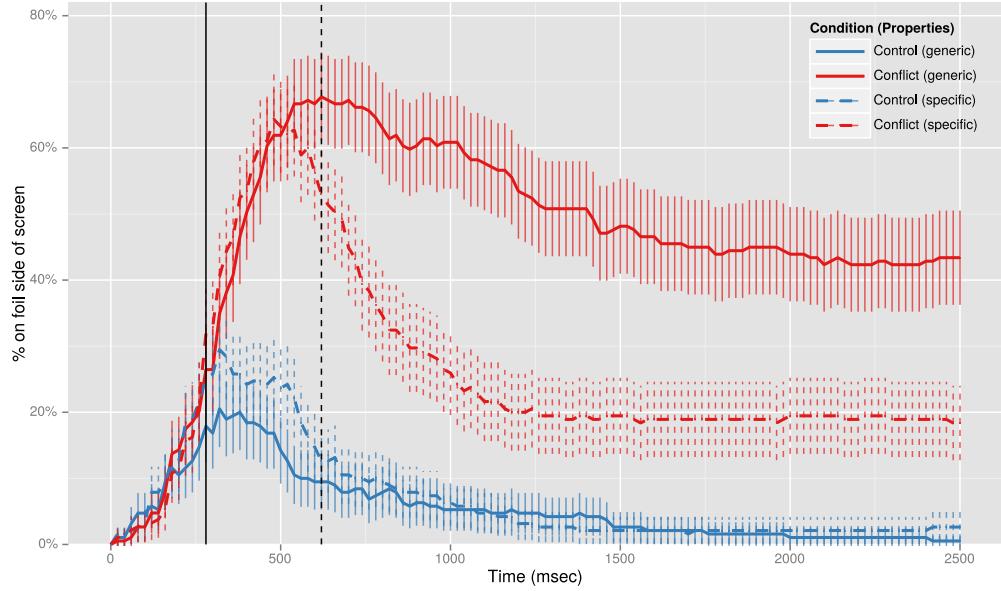


Figure 3.9: Proportion of trials on the side of the screen containing the foil option, over time. Participants were more drawn towards the foil option early in conflict trials (red) than control trials (blue) from 260 msec (solid vertical line). Later on conflict trials, participants were more likely to remain on the foil side when reasoning about generic properties (solid red line) than specific properties (dashed red line), from 620 msec (dashed vertical line).

(dashed vertical lines). In other words, participants reasoning about specific properties were more likely to override their movement towards the foil from this point.

Figure 3.10 shows, for conflict trials, the proportion of trials on the side of the screen corresponding to each response. The top panel shows these trends for participants reasoning about specific properties. Participants were significantly more likely to be on the side of the foil option than the correct option from 300 msec (first solid vertical line) to 620 (vertical dashed line), and subsequently significantly more likely to be on the side of the correct option from 720 msec onwards (second solid vertical line). The bottom panel, for participants reasoning about generic properties, shows a similar trend. Participants were again significantly more likely to be on the side of the foil option from 300 msec (first solid vertical line), but this initial effect persevered for much longer when reasoning about generic properties, to 1,220 msec (vertical dashed line), and participants never showed a significant preference for the correct option's side of the screen.

Therefore, participants reasoning about both kinds of property were initially driven by perceptual cues in the conflict condition, but those reasoning about spe-

cific properties were more likely than those reasoning about generic properties to subsequently draw on conceptual knowledge and move instead to the side of the screen containing the correct response option, and were faster to do so.

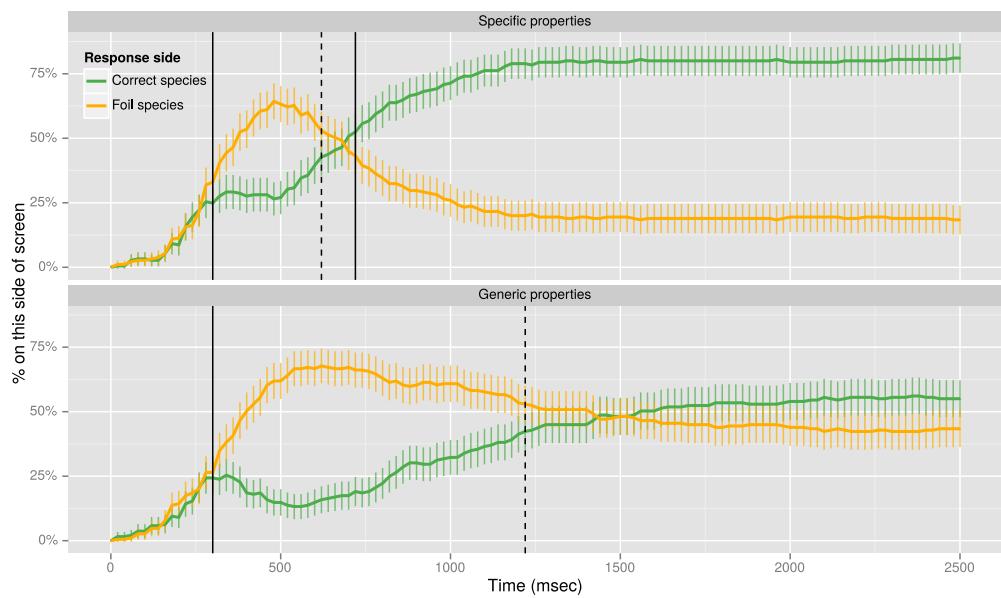


Figure 3.10: Time course of attraction towards foil and correct response options on conflict trials with generic (top) and specific (bottom) properties. Participants reasoning about both kinds of property show an early preference for the foil option (statistically significant from the first solid vertical lines to the dashed vertical lines) which is inhibited quickly for specific properties, and more slowly, and less often, for generic properties. Only participants reasoning about specific properties showed a later significant preference for the correct option (second solid vertical line).

3.3.3 DISCUSSION

In this experiment, participants completed a version of the inductive triad task using artificial categories. Perceptual cues influenced both participants' early cursor movements, and their ultimate responses. When perceptual cues and conceptual knowledge conflicted, participants were more likely to initially move the cursor towards the foil option, particularly if they were quick to initiate their movement, and were also more likely to ultimately select the correct option, regardless of initial movement. Similarly, the time course data showed that participants' cursor movements were driven by these perceptual cues towards the foil option early on conflict trials, but that this initial tendency was generally overridden. In this sense, the current experiment replicates the results of Experiment 1.

This experiment goes further by manipulating the kind of property about which participants reasoned: either specific properties, which were related to the ontological distinction between the categories, or uninformative, generic properties, which were not. This manipulation had no significant effect on participants' early movements, and indeed in the control condition, where participants rarely moved towards the foil option, they appeared to play very little role at any stage. In the conflict condition, however, where participants initially moved towards the foil option on most trials, I found that participants reasoning about specific, distinctive properties were more likely to override their initial, perceptually-driven movement towards the foil than participants reasoning about generic properties. In other words, although reasoning about specific properties made participants no less likely to initially rely on perceptual cues during reasoning, it did make them more likely to reject their initial perceptually cued representation (see Table 3.7, and Figure 3.10).

3.4 GENERAL DISCUSSION

In these two experiments participants completed versions of the inductive triad task where they were asked to generalise a property from a base category to one or other response category. They could rely either on conceptual knowledge, and so generalise the property to the category that belonged to the same category as

the base, or on perceptual cues, and generalise to the category that looked like the base. In both experiments, I manipulated whether the cues agreed or disagreed, so that on conflict trials perceptual cues would lead participants to select the foil response option rather than the conceptually-related one. On these conflict trials, participants were more likely to select the foil category. Furthermore, they were more likely to make fast initial mouse movements towards the foil, which they often overrode to select the correct category instead. These results are consistent with Bright and Feeney's (2014b) hybrid theory of induction.

I also raised the question of how perceptual cues and conceptual knowledge might interact during induction. One option was that participants selectively relied on perceptual cues *or* conceptual knowledge. The alternative was that participants were initially driven by perceptual cues, but sometimes overrode these cues when they realised them to be inappropriate, and relied on conceptual knowledge instead.

The results showed that participants, although usually initially moving towards the perceptually-cued foil, sometimes moved directly towards the conceptually-cued option, with these movements generally taking longer to initiate than those towards the foil. There are two possible explanations for such trials. It may be that participants simply draw on conceptual knowledge here, a process which takes longer, and then act on it. Alternatively, participants on these trials may have first processed the perceptual cues, but inhibited them and replaced them with their conceptual knowledge before initiating their cursor movement. It is not clear at present how these possibilities could be disentangled, and so for the time being this particular issue remains an open question.

In Experiment 2, I manipulated the kinds of properties participants reasoned about: either specific properties, that were saliently related to the distinction between the two categories, or generic properties, that were not. This manipulation mainly affected what participants did on trials where they initially moved towards a perceptually-cued foil option. Participants reasoning about specific properties were significantly more likely to override their initial movement and select the correct option instead than those reasoning about generic properties. The manipulation did not make participants any less likely to initially move towards the foil

option, and its influence on the time course data emerged relatively late in reasoning (see Figure 3.9). It also had no influence on control trials, where participants almost invariably selected the correct option. Therefore, it would appear that this manipulation mainly served to make participants more likely to inhibit their perceptually-driven responses on trials in which they were initially driven towards giving them. This is consistent with much previous work, both using the current paradigm (S. A. Gelman & Davidson, 2013) and other inductive tasks (Heit & Rubinstein, 1994; Ross & Murphy, 1999), indicating that the kind of information people draw on in inductive reasoning is contingent on the nature of the properties to be projected. A future question raised by this result concerns whether this manipulation serves to make participants more likely to inhibit perceptual cues, or if it makes certain structured knowledge easier to retrieve by cuing or priming it.

The purpose of these experiments was to investigate adults' inductive reasoning, and the results are consistent with Bright and Feeney's (2014b) hybrid account. Specifically, they suggest that induction cannot be explained entirely by either accounts based on unstructured associative knowledge such as (perceptual or representational) similarity (i.e. Fisher et al., 2015; Rogers & McClelland, 2004; Sloman, 1993; Sloutsky & Fisher, 2004a) or by purely structured, conceptual knowledge (i.e. S. A. Gelman & Markman, 1986; Griffiths & Tenenbaum, 2009; Kemp & Tenenbaum, 2009; Osherson et al., 1990). Instead, both kinds of information appear to influence adults' reasoning, with simple perceptual similarity drawn on earlier in the process, and perhaps serving as a default. However, these results also have implications for theories of children's reasoning. As discussed in Chapter 1, and earlier in the current chapter, a number of experiments have claimed to show that young children's inferences are either driven by conceptual knowledge (S. A. Gelman & Davidson, 2013; S. A. Gelman & Markman, 1986; S. A. Gelman & Waxman, 2007; Rhodes & Gelman, 2009), or that they are driven by perceptual similarity (Sloutsky & Fisher, 2004b, 2008; Sloutsky et al., 2007). These previous experiments (i.e. S. A. Gelman & Davidson, 2013; S. A. Gelman & Markman, 1986; Sloutsky et al., 2007), however, have focused on a binary question: do children draw on perceptual cues, *or* on conceptual knowledge during reasoning? Therefore, these experiments only presented participants with conflict trials, and

their responses were classed as either consistent with reliance on perceptual similarity, consistent with reliance on conceptual knowledge, or not significantly different from chance in either direction. As discussed in Chapter 1, an experimental control condition, of the type used here, where both cues agree, makes it possible to discover not only which cue dominates when both conflict, but also whether the neglected cue, perceptual similarity in this case, has any influence at all.

Therefore, these results with adults suggest a new interpretation of the developmental data: if both perceptual similarity and conceptual knowledge influence adults' reasoning, they likely both also play a role in children's inferences. This perspective may make sense of apparently contradictory results in the developmental literature, where children seem to draw on conceptual knowledge in some scenarios but not others. It is likely that these studies differ in terms of the factors which make participants more or less likely to inhibit initially influential perceptual cues. Thus, while children may have access to both perceptual cues and information about conceptual knowledge across all of these experiments, they are more likely to inhibit the former in favour of the latter when categories differ at a high, ontological level, when entities are more easily categorised, or when the properties under consideration are conceptually related to the distinction between the categories (S. A. Gelman & Davidson, 2013). In short, the current results suggest that developmental researchers should be less concerned about *whether* children rely on similarity or on conceptual knowledge, and instead ask *when* do children rely on either form of information.

CHAPTER 4

Associative vs. Structured Knowledge in Induction

4.1 INTRODUCTION

In Chapter 3, we saw that induction can be driven by more than one kind of information, as perceptual similarity and conceptual knowledge came into conflict during reasoning. Bright and Feeney (2014b), in their hybrid theory, proposed a more subtle distinction in inductive reasoning, between associative and structured knowledge. While a number of results show that both kinds of knowledge can drive inductive reasoning, however, less is known about how these kinds of knowledge interact. This is the question I seek to answer in this chapter.

As Bright and Feeney (2014b) note, theories of induction can be classed in two ways. Some theories propose that induction is based on structured knowledge: the world is organised into coherent categories, and specific knowledge about these categories, and the often complex relationships between them, is used as the basis for inference. One such theory is Osherson et al.’s (1990) similarity coverage model, that describes inferences about species of animal with reference to the taxonomic relationships between them. More recent accounts have generalised this idea, casting induction as a process of Bayesian inference (Griffiths & Tenenbaum, 2005, 2009; Heit, 1998; Kemp & Tenenbaum, 2009). At the core of these accounts is the notion that in category-based induction, we attempt to use the information given in the premises (i.e. “Carrots have disease X”) to update our beliefs about the dis-

tribution of this property across all categories. To do so, we must be able to express how various categories are related: the probability that rabbits have a given disease, given that carrots have it, depends on the means by which diseases can be transmitted between various species, in this case, through ecological interactions, such as a food chain. Similarly, biological properties, such as genes, are most strongly projected according to the distance between species in the taxonomic tree (Heit, 1998; Osherson et al., 1990), while if we know that certain artefacts are found in one city, geographical distance is used to decide which other cities are likely to house them (Kemp & Tenenbaum, 2009).

On the other hand, a number of theories of inductive inference, and category-based induction in particular, rely on simpler, *associative* forms of knowledge. Similarity, including visual similarity, discussed in Chapter 3, is one such form of knowledge; things that are similar (share many properties, or features, that we know of; see Medin, Goldstone, & Gentner, 1993) are likely to also share novel properties. For instance, on learning about two animals, both of which live underwater, have scales, and breathe through gills, we do not need to know what a “fish” is to predict that if one lays eggs, the other likely does as well. Of course, as discussed in Chapter 3, similarity comes in many forms, and Fisher et al. (2015) make a useful distinction between perceptual and representational (i.e. knowledge-based) similarity. Perceptual similarity is most often proposed as the basis of induction in children (Sloutsky, 2010; Sloutsky & Fisher, 2008; Sloutsky et al., 2007, see also Chapter 3). A number of theories of induction in adults, however, are based on the overlap of features in our mental representations (Rogers & McClelland, 2004, 2008; Sloman, 1993) or perceptual input (Sloutsky & Fisher, 2004a, 2008). Other associative accounts of induction, including of inductive generalisation during learning, are based on what is variously referred to as contiguity, co-occurrence, or thematic relations: things that are often seen together are more likely to share properties than things that are not (Kruschke, 1992; Rescorla & Wagner, 1972; Rumelhart, 1986).

Bright and Feeney (2014b) argue that neither structured nor associative accounts of induction alone are complete. As noted by Murphy and Medin (1985), associative accounts of categorisation and induction fail to capture some of the

flexibility and complexity seen in human reasoning. In particular, participants have been shown to be sensitive to property effects when reasoning inductively: the strength of an argument is dependent on the kind of property projected (Heit & Rubinstein, 1994; Shafto et al., 2007; Shafto, Kemp, Baraff, Coley, & Tenenbaum, 2005). Therefore, transmittable properties such as infectious diseases are thought to be shared by animals that are related ecologically, such as predators and prey in a food chain, whereas biological properties such as genes are shared only by animals that are close together in their taxonomic tree. It is difficult to account for such flexibility in a purely associative account (but see work by Rogers & McClelland, 2004; Sloutsky & Fisher, 2008). Structured models such as the Bayesian accounts discussed above (i.e. Kemp & Tenenbaum, 2009), in contrast, draw on different knowledge structures for different inferences, and so can capture this flexibility in human reasoning.

At the same time, it seems unlikely that structured knowledge alone drives inductive reasoning. Induction is an ubiquitous phenomena, encompassing any inference that goes “beyond the information given” (Bruner, 1973). This ranges from simple perceptual inferences, to the development and postulation of scientific theories (see Kemp & Jern, 2014, for a taxonomy of inductive inferences). Furthermore, inductive reasoning is not exclusive to human adults: similar inferences, although in simpler domains, must be made both by children, and throughout the animal kingdom. Associative theories of induction provide accounts that make more reasonable claims about the cognitive abilities required to reason inductively. Indeed, in the spirit of H. A. Simon (1956), associative processes may *satisfice*, often yielding the same inferences as structured accounts, but with considerably less effort.

Faced with this dichotomy between associative and structured knowledge in induction, Bright and Feeney (2014b) proposed a theory that combines both forms of knowledge. This *hybrid* theory claims that both kinds of knowledge are drawn upon during reasoning, with a key distinction between the two being their processing characteristics. By this account, associative knowledge can be retrieved quickly and easily, while structured knowledge is slower to retrieve, and places greater demands on working memory to utilise.

Evidence for this hybrid theory comes from experiments that manipulate the processing conditions under which participants reason. This research is discussed in more detail in Chapter 1, but to recapitulate, participants' inferences (their choices, or ratings of argument strength) are predicted by measures of associative knowledge under all circumstances. Furthermore, under favourable conditions only (i.e. in the absence of time pressure, or cognitive load), inferences are additionally predicted by appropriate measures of structured knowledge, such as whether participants believed two species belonged to the same taxonomic group when reasoning about biological properties.

Of particular relevance to this thesis, which focuses on *conflict* in reasoning, Bright and Feeney (submitted, see also Crisp-Bright, 2010, Chapter 5) report a version of the inductive triad task (S. A. Gelman & Markman, 1986) that places associative and structured knowledge directly in conflict. In the trial shown in Figure 4.1, for example, participants were told a biological property of carrots (they have "C5s cells"), and chose between generalising this property to rabbits, or to bamboo. Carrots and rabbits are strongly associated, and so a participant relying on associative knowledge would project a property from carrots to rabbits. This is despite the fact that the link between the species is a food chain, and so doesn't provide a means for them to share a biological property. Carrots and bamboo, on the other hand, are both plants, and so are more likely to share such a property. The associative link between the two species, however, is substantially weaker, and so projecting the property to bamboo requires that one inhibits the associative knowledge linking carrots and rabbits before one can draw on the structured link between carrots and bamboo instead. Consistent with the hybrid theory, participants were less likely to select the taxonomically-related response option over a strongly associated foil in this task when under heavy cognitive load, or if they had poor semantic inhibitory control (a measure of their ability to inhibit task-irrelevant semantic information; Burgess & Shallice, 1996; Markovits & Doyon, 2004). Like most studies that analyse only participants' ultimate responses, however, while these results do reveal what information drove participants' final choices, we are limited in what conclusions we can draw about how different processes interact during this task. To address this shortcoming, in this

chapter I present a mouse tracking extension of the Bright and Feeney (submitted) triad task.

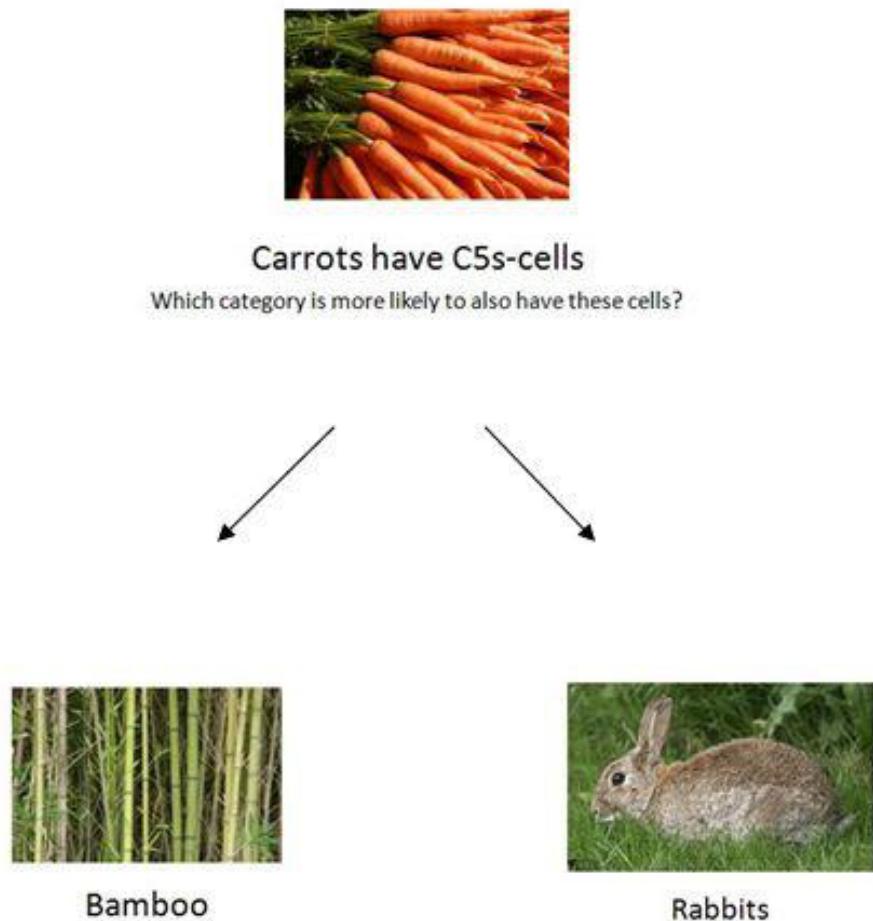


Figure 4.1: A conflict trial from Bright and Feeney (submitted). Participants learn that carrots possess the given biological property, and asked which of the other two species, bamboo or rabbits, are likely to share this property.

How might associative and structured knowledge interact during this task? Again, I propose two possibilities. First, it may be that people selectively draw on associative knowledge *or* structured knowledge for a given inference. In this case, we would expect to find little evidence of actual conflict during reasoning, as both kinds of knowledge would not compete during a single trial. This would lead to cursor trajectories where participants move directly towards one or other option, and then select it, rather than changing direction mid-flight. The second possibility is that associative knowledge may be activated early in reasoning, but be

later overridden, at least some of the time, by slowly-retrieved, more cognitively demanding structured knowledge. In this case, participants would be conflicted when they do override, or at least attempt to override, their association-driven response. Cursor trajectories in this case should largely be initially drawn to the foil species when it is cued by associative knowledge, but also likely to override this initial movement and select the correct species instead at least some of the time.

4.2 EXPERIMENT 3

In this experiment, I adapted Bright and Feeney's (submitted) inductive triad task to record participants' mouse cursor trajectories as they choose to generalise a property to one or other species. Analysing participants' responses, I hoped to replicate that study's basic finding, that participants are less likely to select the taxonomically-related response species when the foil option is strongly associated with the base species. Going beyond this, however, I hoped to use participants' mouse cursor data to draw conclusions about the processes leading up to these responses.

4.2.1 METHOD

PARTICIPANTS

Forty-one undergraduate students at Queen's University Belfast participated in exchange for course credit. Data were lost from one participant due to a malfunction in the experimental software. Participants completed the experiment in a laboratory. The experiment was programmed using the PsychScript package (see Chapter 2) and run in the web browser.

STIMULI

Stimuli were those used by Bright and Feeney (submitted). There were fourteen experimental stimulus sets, in which participants reasoned about genetic properties. Each set consisted of a base species, which participants were told had the property in question, a correct response species, and two foil species. Each species

was represented in the experiment by a labelled picture (see Figure 4.2). Crisp-Bright (2010, Chapter 2) collected ratings of the strength of association between the base and the correct species, and between the base and strongly associated species. She presented 18 undergraduate students with pairs of species names (i.e. “Snails and Octopuses”) and asked them to rate, on a scale from 1 to 9, how strongly associated they believed the species to be, giving their first, intuitive response. These average association ratings, along with a table of the species used, can be found in Appendix G.

In each stimulus set, the base and correct species were moderately associated with each other, according to Crisp-Bright’s (2010, Chapter 2) ratings, and belonged to the same taxonomic category (mammals, reptiles, fish, birds, or plants). The two possible foil species belonged to a different taxonomic category than the base. Each set had a conflict foil, that was strongly associated with the base, and a control foil, assumed to be weakly associated. Each set was presented twice, once with its conflict foil, and once with its control foil, for a total of twenty eight experimental trials. All stimuli used on experimental trials can be found in Appendix H



Figure 4.2: A screen shot from Experiment 3. Participants were told that the base species, Salmon, had a certain biological property, and where asked which of the two response species, Goldfish or Bears, were most likely to also have this property.

There were a further fourteen filler stimulus sets, in which participants reasoned about diseases. These consisted of the same base species as in the experimental sets, but different response species: the correct response species here were the strongly associated foils from the experimental stimulus set, and so related to the bases in a way that would allow the transmission of disease, while one foil response belonged to the same taxonomic group as the base, and one did not. Again, each set was presented twice, once with each foil. These trials were included in order to prevent participants from adopting a strategy of always selecting the taxonomically-related response species, and were not included in the analyses. Participants were asked about genetic properties on experimental trials, and diseases on filler trials. Each property was given a fictional, uninformative name, e.g. “gene 5U3”/“disease 3k0”.

PROCEDURE

The 28 experimental trials and 28 fillers were presented in randomised order. Table 4.1 shows the format of the reasoning trials. Each trial began with the word “GENE” or “DISEASE” displayed in the centre of the screen in large font for 1 second, and then shrinking and moving to the top of the screen where it remained throughout the trials, to remind participants what kind of property they were reasoning about. Participants were next presented with the specific property (text for filler trials in brackets): “There’s a kind of gene [disease] called x3f, which is found in the body of [is known to infect] either…”, for 2.4 seconds. They then saw one species option in the top left corner of the screen, along with the text “...this species…”, followed by the other species in the top right corner and the text “...or this species”, for 1.6 seconds each, with only one species shown at a time. The location of each response option was randomised on each trial. Both response options were then displayed, along with the text “Which species do you think is most likely to have this gene in their bodies [be infected by this disease], given that it is also found in... [that it also affects...]”.

At this point, participants were required to click a START button in the bottom centre of the screen, after which the text was replaced by a 1.5 second fixation, and then by the base species. After the onset of the base species, partici-

Table 4.1: The procedure followed on reasoning trials in Experiments 3 and 4.

Stage	Stimuli	Duration (msec)
Prime property	“GENE” or “DISEASE” in centre of screen	1,000
Property	“There’s a kind of gene [disease] called x3f, which is found in the body of [is known to infect] either ...”	2,400
Response species #1	“...this species...” (show species in top left)	1,600
Response species #2	“...or this species.” (show species in top right)	1,600
Prompt to start trial	“Which species do you think is most likely to have this gene in their bodies [be infected by this disease], given that it is also found in... [that it also affects...]” (show both response species, and “Next” button)	Click “Next” button
Show base species; Record response	Show base species in centre of screen; Record cursor position as participants select a response species.	< 5,000

participants responded by clicking one or other response species, and the position of the mouse cursor was recorded continuously as they did so. Participants were required to respond within 5 seconds, and were prompted to start moving earlier on trials in which the mouse cursor did not move within the first 1.3 seconds, a time which was selected based on pretests (see Chapter 2). On trials in which participants moved the cursor off the start button during the fixation period, they were brought back to the point before they clicked the START button, and asked not to start moving until they saw the final species.

Following Bright and Feeney (submitted), after the induction trials participants completed a post-test check to ensure that they possessed the appropriate structured knowledge. For each experimental stimulus set, the base species was paired with the correct response species and both foil species, to create 42 pairs. Participants were asked to indicate, for each pair, if the two species belonged to the same “biological group”. There were also 42 filler questions, created by pairing each base

species with its possible responses from the filler induction trials, in which participants were asked if the pair belonged to the same “food chain”. Each question was accompanied by the labelled images of each species used in the induction trials, and participants responded by clicking buttons marked “Yes” or “No”. The 84 post-test questions were presented in random order.

4.2.2 RESULTS

POST TEST CHECK

The post-test data make it possible to ensure that participants possessed the appropriate structured knowledge for each trial. For instance, if participants did not realise that dolphins and llamas belong to the same biological group (mammals) then we cannot assess the interaction of associative and structured knowledge when reasoning about these species. Appendix I shows the proportion of participants who correctly identified that the base and correct response species belonged to the same taxonomic group, and that the base and foil response species did not, for the stimuli used in the experimental trials. Participants only performed above chance on the post-test for the correct response species for 7 of the 14 stimuli sets (binomial test, $N=40$, 67.5% accuracy required for $p < .05$), and so only these sets were included in the subsequent analyses.

REASONING ACCURACY

Due to an error in the software used to run the experiment, mouse cursor data were recorded incorrectly on 40 trials (7.1% of total) in which participants took more than 2,200 msec to respond, and so these trials were discarded from the analysis. In the analyses that follow, I fitted linear or logistic mixed models, with random intercepts for each participant, and for each stimulus set.

Participants were significantly more likely to choose the foil species on conflict trials, where the foil and the base were strongly associated (26%), than on control trials, where they were not (5%; $e^\beta = 10.6$, CI = [5.2; 21.8], $z = 6.421$, $p < .0001$). It may be the case that this effect is driven by a subset of participants who are influenced by associative knowledge. Figure 4.3 shows each participant’s number of

foil responses, by condition. Of the 40 participants, 30 gave the foil response more often under conflict, none did so more often on control trials, and 10 never gave the foil response. Therefore, the main effect on participants' responses appears to apply across all participants, or at least all participants who did not achieve perfect accuracy.

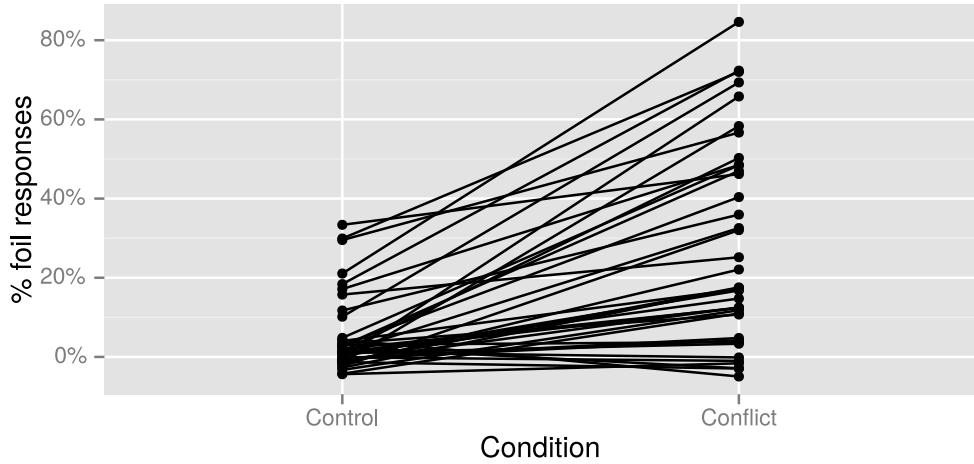


Figure 4.3: Participants' number of foil responses, by condition. Of 40 participants, 30 gave more foil responses under conflict, and 10 never gave the foil response.

CORRECT RESPONSES

On trials in which the correct response was given, there was no difference in response time between the conditions (1,264 msec in conflict condition, SD = 297, 1,273 msec in control condition, SD = 361; $t < .8$, $p > .4$). Participants were significantly faster to initiate the mouse movements for correct responses in the conflict condition (628 msec, SD = 213) than in the control condition (688 msec, SD = 206; $e^\beta = 89\%$, CI = [84%, 95%], $t(405.2) = 3.670$, $p < .0001$).

Maximum Deviation was bimodally distributed (Bimodality Coefficient = .705; Hartigan's D = .0422, N = 520, $p < .0001$), and trajectories were classified as either direct trajectories or reversals, as described in Chapter 2 (MD cut-off = 0.524). A significantly greater proportion of trials where participants selected the correct species were classed as reversals in the conflict condition (16%) than the control condition (8%; $e^\beta = 2.2$, CI = [1.2; 3.9], $z = 2.524$, $p = .0116$). Therefore, although response latencies do not suggest conflict, analysis of participants' mouse

cursor movements shows that responses based on structured knowledge showed a greater attraction towards the foil response option when the foil was strongly associated with the base species.

CURSOR TRAJECTORIES

Once again, cursor trajectories in this experiment can be classed as either Direct Correct, Reversal Correct, Reversal Foil, or Direct Foil trajectories. Table 4.2 shows the proportion of each trajectory type by condition. Under conflict, there was a small increase in the number of Reversal Correct responses, and a considerably larger increase in the number of incorrect responses.

Table 4.2: Proportion of each trajectory type, by condition. Under conflict, participants were markedly more likely to directly move to the foil option, and less likely to directly move to the correct option.

	Direct Correct	Reversal Correct	Reversal Foil	Direct Foil
Control	87%	8%	0.8%	4%
Conflict	62%	12%	3%	23%

Transition probabilities for trajectories from this experiment are shown in Table 4.3. Participants were significantly more likely to initially move towards the correct species on control trials (88%) than conflict trials (65%; $e^\beta = 4.7$, CI = [2.9; 7.6], $z = 6.284$, $p < .0001$). On trials where they have moved towards the correct option, participants were also more likely to select it on control trials (99%) than conflict trials (96%; $e^\beta = 5.2$, CI = [1.04; 26.2], $z = 2.006$, $p = .0448$), although in both cases participants who initially move towards the correct option are extremely unlikely to subsequently change direction. Finally, on the trials where participants did initially move towards the foil option, they are more likely to change direction and select the correct option instead on control trials (68%) than on conflict trials (34%; $e^\beta = 16.9$, CI = [3.6, 80.2], $z = 3.567$, $p = .0004$).

Lastly, on conflict trials, movement initiation times did not differ significantly between initial movements that were towards the correct option (617 msec, SD = 225) and those towards the foil option (652 msec, SD = 201; $t(251.8) = 1.365$, $p = .1733$).

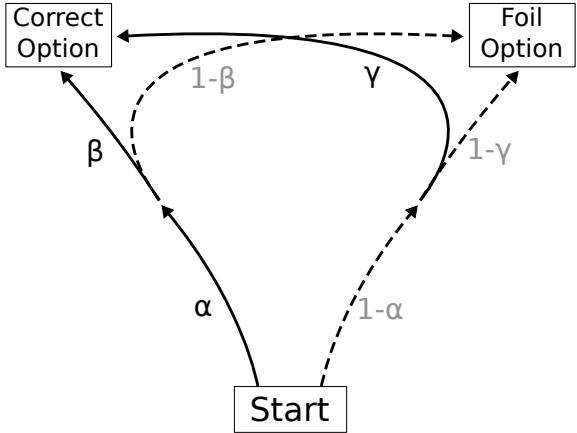


Figure 4.4: The possible transitions that can occur during a trial.

Parameter	Control	Conflict
α	88%	65%***
β	99%	96%*
γ	68%	34%***

Note: * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 4.3: Transition probabilities for Experiment 3.

TIME COURSE

Figure 4.5 plots the proportion of trials in which the cursor is on the foil species' side of the screen (top), and the correct species' side (bottom), broken down by condition. For each plot, a series of logistic mixed models were fit, with the probability of the cursor being on that side of the screen predicted by condition (control or conflict), across each 20 msec window, from 100 to 1,000 msec. Random intercepts were included for each participant and for each base species. By identifying the points from which there were significant effects of condition in each series of models, we can see at what point in time manipulating the association of the foil species affected participants' cursor trajectories to that side of the screen. Trials in the conflict condition were more likely to move towards the foil response than those in the control condition from 480 msec onwards. Those in the control condition were more likely to move towards the correct response than those in the conflict condition from 760 msec onwards.

Figure 4.6 shows the same data, but with separate plots for the control (top) and conflict (bottom) conditions, and separate lines corresponding to the proportion of trials on each side of the screen. Divergence times were calculated in the same way, this time reflecting the point, in each condition, that participants became more likely to move towards the correct species than the foil. In both conditions, this divergence occurred at 480 msec. This means that while participants selected the foil option significantly more often on conflict trials than control tri-

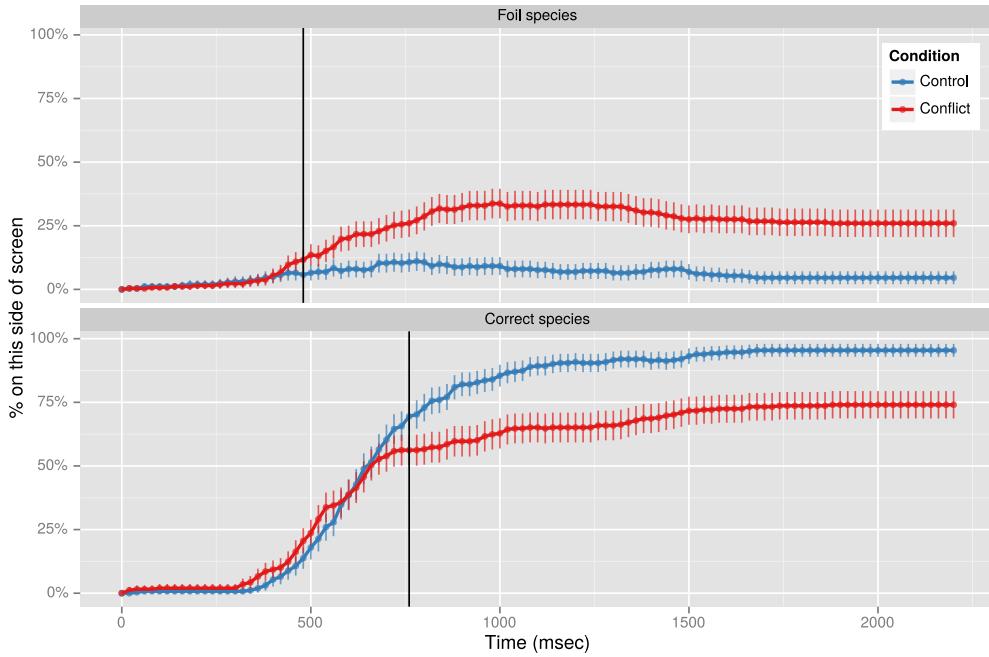


Figure 4.5: Proportion of trials on each half of the screen, over time. Vertical lines show the points from which there were significant differences between the control and conflict conditions.

als, they were no slower to show a preference for the correct species under conflict.

Recall that in Experiments 1 and 2, the plots corresponding to the bottom panel of Figure 4.6 showed a cross-over trend: participants were initially more likely to move towards the foil option, with this initial preference fading, and eventually reversing, as the correct option was chosen on the majority of trials. Here, instead, we see that participants are never more likely to move towards the foil option than the correct one. The implications of these temporal dynamics are discussed below.

4.2.3 DISCUSSION

In this experiment, participants had to generalise a biological property from a base species to one of two other species: one that belonged to the same taxonomic group as the base, and one that did not. On conflict trials, but not control trials, there was a strong association between the base and foil, so participants should have been drawn towards selecting the foil species on these trials if this irrelevant associative knowledge plays a role in inductive reasoning. Replicating the core findings of Bright and Feeney (submitted), I found that participants were more likely to select the foil species on conflict trials, driven by associative knowledge.

Going beyond previous research, this experiment allows us to draw inferences about how these two kinds of knowledge interact. Bright and Feeney (submitted) also showed that participants were more likely to select the foil response under conflict if they had poor semantic inhibitory control, or were placed under heavy cognitive load. This suggests that drawing on the structured knowledge needed to respond correctly on conflict trials requires top-down executive functions. It may be that the actual utilisation of structured knowledge places demands on these executive processes, or alternatively, that associative knowledge must be inhibited before structured knowledge can be retrieved. Perhaps consistent with the latter interpretation, I found that on trials where participants did select the correct species, they were more likely to initially move towards the foil species on conflict trials, where the foil was strongly associated with the base.

Other findings, however, make this less clear. The proportion of correct responses that were reversals, for instance, was low compared to other experiments reported in this thesis: 8% of control trials, and 16% of conflict trials. The current experiment also differed from Experiments 1 and 2 in that participants' response times were no slower for conflict trials than for control trials. Similarly, the participants initially moved towards the foil species on only 35% of conflict trials, but after doing so, they only subsequently changed their minds to select the correct species instead 34% of the time. Finally, the time course analysis showed that even on conflict trials, participants never were more drawn towards the foil species than the correct one. Therefore, a more accurate description of these results is as follows. In some cases participants' mouse movements were initially driven by associative knowledge, and later intervened upon on the basis of structured knowledge. On the majority of conflict trials, however, participants either moved directly to the species cued by structured knowledge, or moved straight to the response cued by associative knowledge.

These results differ from those found in Experiments 1 and 2 in a number of ways. However, before I attempt to interpret these results any further, a limitation of this experiment must be noted. Participants' performance on the post-test check, which was used to ensure that they knew that each correct response species belonged to the same biological group as its corresponding base species, was ex-

tremely poor. Specifically, participants' performance was not significantly above chance when it came to knowing that the following species belonged to the same group: dolphins and llamas, monkeys and seals, snails and octopuses, bananas and tulips, penguins and chickens, mice and goats, and Orca whales and cows. As a result, data from these 7 of the 14 stimuli sets were excluded from the analysis. Given that the purpose of this chapter is to study conflict between associative and structured knowledge, such poor taxonomic knowledge may be problematic. Consider, for instance, a participant who is drawn towards generalising a gene from dolphins to cod, rather than from dolphins to llamas. If this participant knows that dolphins and llamas are mammals, and cod are not, then we can infer that this attraction is driven by unstructured associative knowledge. However, if this participant believes that dolphins and cod belong to the same taxonomic group, then both their associative and structured knowledge would support the same inference. In this case, it is difficult to know what form of knowledge they are drawing upon.

Therefore, In Experiment 4, I attempted to replicate these findings conceptually, using new stimuli for which the appropriate taxonomic relationships are more obvious. Furthermore, Experiment 3 relied on association ratings collected by Crisp-Bright (2010, Chapter 2) from a different pool of university students, and only allowed for the manipulation of the foil species, which was either strongly associated with the base, according to the prior ratings, or (assumed to be) not associated. In Experiment 4, in contrast, I collected association ratings for each species pair from each participant, after they had completed the rest of the experiment. In this way, it was possible both to investigate the relationship between association ratings and participants' choices and cursor movements in a more nuanced way, and to use each participants' actual beliefs about the strength of association between species, rather than aggregate ratings from a separate pool of participants, in the analyses.

4.3 EXPERIMENT 4

In Experiment 4, I attempted to replicate the findings of Experiment 3, but without the need to discard data from trials for which participants lacked appropri-

ate taxonomic knowledge. Beyond this, I wished to explore how each participants' own beliefs about the strength of association between various species interacted with structured knowledge during reasoning. Therefore, in this experiment, I asked participants to rate the association between each base species and every response species it was paired with for the reasoning trials. This improves on the method used in the previous study in two ways. First, while Crisp-Bright (2010, Chapter 2) collected ratings of the association between the base species, the correct response species, and the strongly associated foils used in the previous experiment, she did not collect ratings for the associations between the bases and the foil species assumed to be weakly associated. Second, Crisp-Bright (2010) collected association ratings from one set of participants, and reasoning data from another. Therefore, it was the average association rating between species that she used as a predictor in her analyses. Here, on the other hand, I collected each participants' own idiosyncratic ratings of the associations between the species, allowing a more fine-grained analysis of how such associative knowledge influences reasoning.

4.3.1 METHOD

PARTICIPANTS

Forty four undergraduate students completed the experiment in exchange for course credit, in a laboratory. The experiment was programmed using the OpenSesame experiment builder (see Chapter 2).

STIMULI

In the experimental trials, participants were asked about biological properties, specifically cells. Nine new stimulus sets were generated for the experimental trials, intended to be more familiar to participants than those used in Experiment 3. Each new set had three foil species, one intended to be weakly, one moderately, and one strongly associated with the base. Each set was presented three times, once with each foil species. Stimuli were selected according to a number of partial pretests, in which participants rated the strength of association between species using the procedure from Crisp-Bright (2010, Chapter 2), described above. The full

set of stimuli can be found in Appendix J.

For the filler trials, where participants were asked about diseases, an additional fourteen stimulus sets were generated, each containing a base species, a correct response species likely to share a disease with the base, and three different foil responses, one for each time the set was presented. One possible concern about the design of Experiment 3 is that the species designated as the correct response for each experimental stimulus set was the correct response on every trial it featured in. Therefore, the fourteen correct response species from the experimental trials here were used as foil species (that is, the species that were unlikely to share a disease with the base species) on three different filler trials. This meant that these species were the correct response option in the three experimental trials in which they featured, but also the incorrect response option in three filler trials. The properties to be reasoned about — genes on the experimental trials, and diseases on filler trials — were unchanged from Experiment 3.

To ensure that participants did not complete experimental trials with the same base species in close succession, the order of trials was randomised with constraints for each participant. First, the experimental trials were randomly divided into three blocks of nine trials, with each block containing three weak, three moderate, and three strong foils, and one trial from each stimulus set in each. Nine of the twenty-seven filler trials were then added to each block. Finally, the order of trials within each block was randomised repeatedly, until at least 5 trials separated repetitions of each base species.

PROCEDURE

There were minimal changes to the reasoning trials from Experiment 3. However, this experiment was conducted using the OpenSesame platform, which allowed greater experimental control over the mouse cursor. Therefore, instead of requiring participants not to move the cursor during the fixation period, the cursor's position was automatically reset to the centre of the START button after the fixation.

After the reasoning trials, participants again completed a post-test check. In the first part of the post-test, participants rated the strength of association of the thirty-six base-response species pairs from the experimental trials. These consisted

of the nine base species, each of which was paired with its correct response species and its three foil species. For this section, participants were presented with the following instructions, taken directly from Crisp-Bright (2010, Chapter 2, p. 60):

[...] Please think about all kinds of possible associations, such as causal, functional, categorical, etcetera. Please do not think in detail about the mechanism by which they are related, just give your intuitive response. For example, if you believe that ladybirds and butterflies are strongly associated please give a rating closer to 9. In contrast, if you think cars and ladybirds are unrelated, please give a rating closer to 1. Please give the answer that first comes to mind, as fast as possible.

On each rating trial, the labelled images of each species were shown side by side, with their positions randomised. Participants gave their ratings by clicking on buttons labelled 1 to 9 below the images, with 1 subtitled “Not associated at all”, and 9 subtitled “Very strongly associated”.

The second part of the post-test checked participants’ structured knowledge. Participants were presented with pairs of species, in the same format as the association rating trials, and gave yes or no responses by clicking the marked buttons. There were three blocks in this part of the post-test, and the order of trials within each block was randomised. First, for each of the nine experimental stimulus sets, the base species and the correct response species belonged to the same taxonomic group. These nine pairs of species were presented along with an additional nine pairs that did not belong to the same group. Participants were asked in each case if the pair shown belonged to the same *biological group*, and told “Biological groups are the main branches when you think about the ‘family tree’ of the natural world”, and that the biological groups in the experiment were mammals, fish, reptiles, birds, and plants.

Second, for five of the experimental stimulus sets, the base species was related to the moderately and strongly associated foils via a food chain relationship — one species eats the other. These ten base-foil pairs were presented along with ten other species pairs not related in this way. Participants were asked if each pair shown belonged to the same *food chain*, and told “Species belong to the same food

chain if one is eaten by the other (predators and prey for animals, or a plant which is eaten by an animal)”.

Finally, for the remaining four experimental stimulus sets, the base species was related to the moderate and strong foils in that they shared an ecological habitat.¹ These eight base-foil pairs were presented along with the eight other pairs that did not share a habitat, and participants were asked if the pair shown “live in the same kind of habitat”.

4.3.2 RESULTS

POST TEST CHECK

For each stimulus set used in the experiment, participants correctly said that the base species and the correct response species belonged to the same taxonomic group (accuracy > 70.5%, p’s < .01), and that the moderate and strong foils had the appropriate relationship (food chain or shared habitat) with their base species (accuracy > 75%, p’s < .001). Therefore, no data were excluded from the analyses on these criteria. Full post-test results can be found in Appendix K.

REASONING ACCURACY

In order to analyse data from the reasoning trials, for each trial I divided that participants’ rating of the association between the base and the correct species by their rating of the association between the base and the foil. This produced a ratio reflecting the extent to which that participants’ associative knowledge favoured one or other response option. Values of this ratio ranged from $\frac{1}{9}$ (.111; association of 1 for the correct species, and 9 for the foil), to $\frac{1}{1}$ (equal association for both species), to $\frac{9}{1}$ (association of 9 for the correct species, and 1 for the foil).

For each analysis, I log-transformed this ratio to create a normally-distributed linear predictor, as is standard practice when using ratios as regression predictors (A. Gelman & Hill, 2007). Note that $\log(\frac{1}{1}) = 0$, and correspondingly $\log(\frac{\geq 1}{1}) > 0$ and $\log(\frac{1}{\geq 1}) < 0$ (see Figure 4.7). Therefore, when using the log-transformed ratio

¹ This was not technically true for penguins and arctic wolves, and penguins and polar bears, as these species live in opposite polar regions, but the post-test data showed that participants do not realise this.

as a predictor, a positive regression weight means that the dependent variable is greater/more likely when the ratio is greater than $\frac{1}{1}$, or in other words, when the association rating favours the correct response.

Some aspects of the analyses based on this ratio are slightly unusual, and so I will go through my first analysis, predicting participants' responses, in detail to familiarise readers with the method. Figure 4.8 plots the proportion of correct responses (choosing the species belonging to the same taxonomic group as the base) given, on the y axis, as a function of the association ratio, on the x axis. Note that the x axis is log-scaled. For plotting, I have divided the log-transformed ratio into 13 equal bins, and plotted the mean and standard error of the dependent variable within each bin. For the analyses, I fit log-linear, or logistic mixed models, with random intercepts for each participant, and for each stimulus set. The log-transformed association ratio from each trial was used as the predictor in the models.

The association ratio was a significant predictor of the odds of a correct response ($\beta = 0.89$, CI = [0.68, 1.1], $z = 8.604$, $p < .0001$). Interpretation of the regression β here is not straightforward, but positive β s indicate that the dependent variable (odds of a correct response) was positively related to the size of the association ratio in favour of the correct response. Fortunately, we can simply look to the predicted values from this model (solid red line in Figure 4.8) to see the magnitude of this effect.

An unusual property of these models is that the intercept parameter is also meaningful. As the log-transformed association ratio is the only predictor in the model, the intercept reflects the expected value when this predictor is at 0, or in other words, for a ratio of $\frac{1}{1}$. If participants do not make use of structured knowledge, we would expect participants to select the correct species 50% of the time for such trials, something that would correspond to an intercept of 0. The dashed black line in Figure 4.8 shows the predicted values for such a model, with intercept 0. A significant positive intercept means that participants were more likely to select the correct species than would be expected from the association ratio alone, while a negative intercept means they were less likely. There was a significant positive intercept term in this model ($\text{logit}(\beta) = 72\%$, CI = [56%, 85%], $z =$

2.585, $p = .0010$). I report the *logit* of the regression β weights here as an easily interpretable measure of how much participants were biased towards the correct species. In this case, $\text{logit}(\beta)$ reflects how often participants would be expected to select the correct species on trials where the association ratio = $\frac{1}{1}$, according to the fitted model (i.e. where the red line on Figure 4.8 crosses 1 on the x axis). Together, these results mean that participants were a) more likely to select the correct species when they rated it as being more strongly associated with the base than the foil was, and b) biased towards the correct option rather than the foil beyond the effect of the association ratio.

To analyse individual differences in the number of correct responses (Figure 4.9), I calculated the number of correct responses given by each participant both when the association ratio favoured the foil species, and when it favoured the correct species, excluding trials where species were rated as equally associated. Of the 44 participants, 40 were more likely to select the correct species when the association ratio favoured it, and 4 were less likely to do so. Therefore, it appears that almost all of my participants were influenced by the association between species.

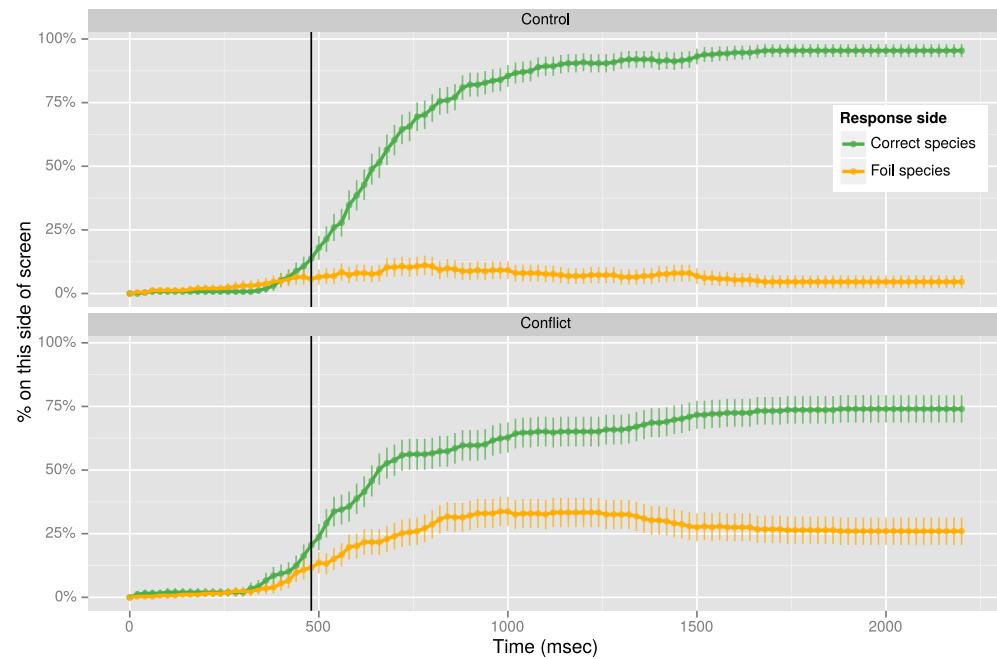


Figure 4.6: Proportion of cursors on each half of the screen, over time. Vertical lines show point from which participants were more likely to move towards the correct option than the foil.

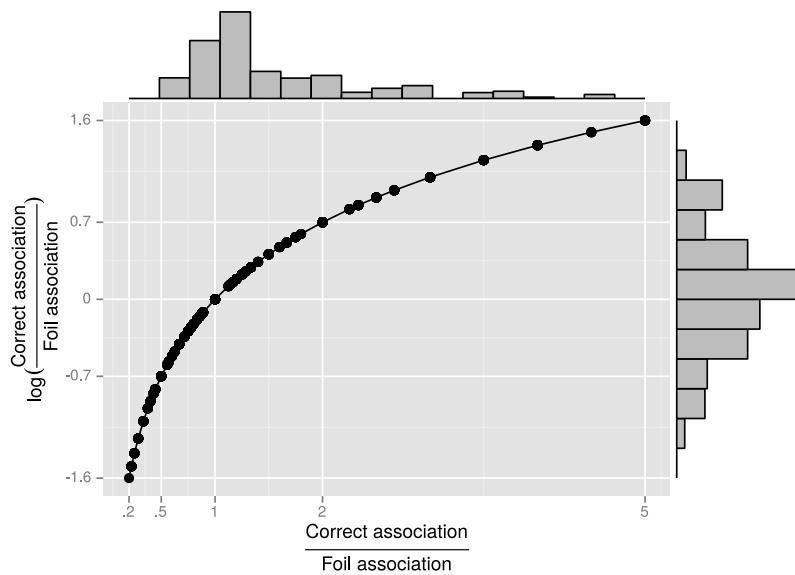


Figure 4.7: To analyse data from Experiment 4, each participants' rating for the association between the base species and the correct response species was divided by their rating for the base and the foil species, to form an association ratio. These ratios were log-transformed to use as predictors in regression analyses: after log-transformation, the difference between 1 and 0.2 (dividing by five) is the same as the difference between 1 and 5 (multiplying by five).

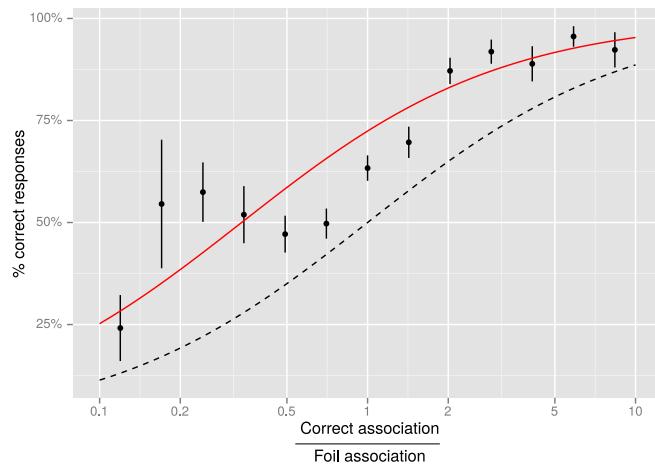


Figure 4.8: Correct responses on Experiment 4, as a function of the association ratio in favour of the correct species. As the ratio increased in favour of the correct species, participants became more likely to select that species. The fitted model (red line) included a significant positive intercept term, meaning that participants were more likely to select the correct species than would be predicted based on the association ratio alone (dashed black line).

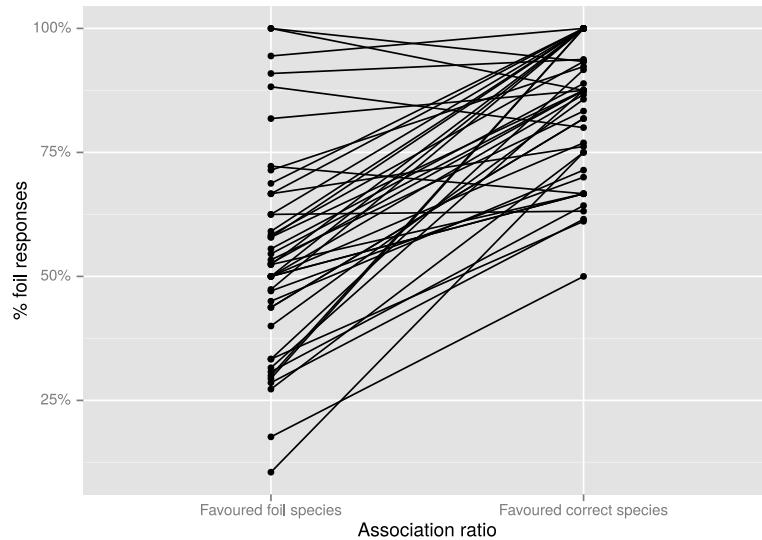


Figure 4.9: Participants' number of correct responses on trials where the association ratio favoured either the correct species or the foil species. 40 of 44 participants were more likely to select the correct species when the association ratio favoured it than when the ratio favoured the foil.

CORRECT RESPONSES

Analysing trials where the correct species was selected, the association ratio had no effect on participants' movement initiation times (mean = 574 msec, SD = 285; $t(529.6) = 0.387$, $p > .6$). There was a marginally significant effect of the association ratio on participants' response latencies (mean RT = 1509, SD = 567; $e^\beta = 98\%$, CI = [96%, 100.4%], $t(787.1) = 1.647$, $p = .100$), meaning that participants were marginally faster to respond correctly as the association ratio in favour of the correct species increased.

Maximum deviation was once again bimodally distributed (Bimodality Coefficient = .635; Hartigan's D = .025, N = 1188, $p < .0001$), and reversal trajectories were classified as described in Chapter 2 (MD cut-off = 0.923, see Appendix B). Figure 4.10 shows the proportion of correct responses that were categorised as reversals — that is, where participants moved initially towards the foil before selecting the correct species — as a function of the association ratio. A logistic mixed model (dashed red line) showed that participants were less likely to follow such a reversal trajectory as the association ratio increased in favour of the correct species ($e^\beta = .8$, CI = [.6, .97], $z = 2.152$, $p = .0314$). However, the data were fit slightly better² (Δ deviance = 0.715) by an alternative model (solid red line), where the odds of a trajectory being classed as a reversal when selecting the correct species were predicted by the *absolute magnitude* of the log-transformed association ratios ($e^\beta = .7$, CI = [.5, .9], $z = 2.289$, $p = .0221$). In other words, when selecting the correct species, participants were most likely to trace a reversal trajectory when the species were equally associated, and less likely to do so as the ratio changed in favour of either species. This trend may appear counter-intuitive, but may make sense when one considers that this analysis does not include trials where the foil species was selected. An explanation for the trend is offered in the discussion, below.

² As these two models were not nested, and contained the same number of parameters, we can identify which model best fits the data by comparing the deviance of each, but we cannot calculate p values for the difference between the models.

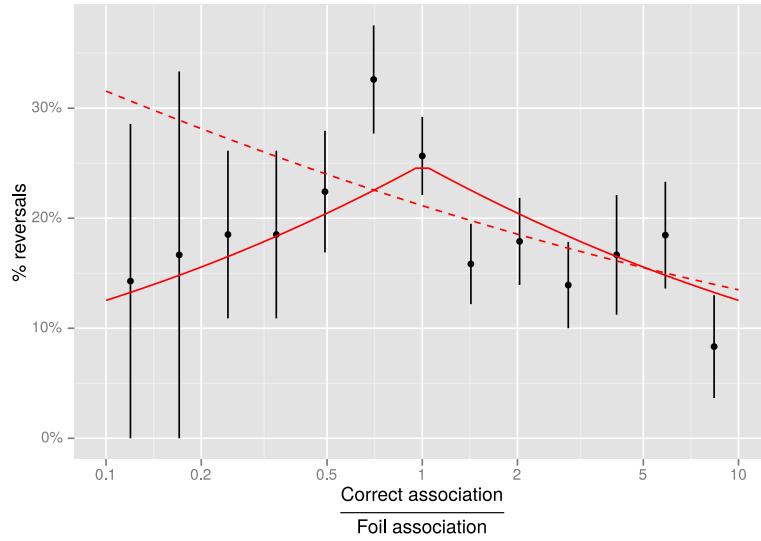


Figure 4.10: The proportion of correct responses where participants initially moved towards the foil species. A model using the log-transformed association ratio as a predictor (dashed red line) showed that participants were more likely to produce these reversal trajectories when the association ratio favoured the foil species. However, the data were slightly better fit by a model (solid red line) which showed that these reversals were most likely to occur when the association ratio favoured neither response.

CURSOR TRAJECTORIES

As before, mouse trajectories in this experiment can be described in terms of whether or not they initially moved towards the correct species (α), whether they selected the correct species after initially moving towards it (β), and whether they changed direction to select the correct species after initially moving towards the foil (γ ; see Figure 4.4). I modelled these three parameters using multilevel logistic regression models, with the log-transformed association ratio and an intercept term as predictors, and random intercepts for each participant and stimulus set.

The model for α (Figure 4.11, top left) had a marginally significant positive intercept ($\text{logit}(\beta) = 62\%$, CI = [50%, 73%], $z = 1.886$, $p = .0593$) indicating that participants initially moved towards the correct species on 62% of trials when the ratio was $\frac{1}{1}$, more than the 50% that would be expected based on the association ratio alone. The association ratio was also a significant positive predictor ($e^\beta = 1.7$, CI = [1.4, 2.0], $z = 6.109$, $p < .0001$) such that participants became more likely to initially move towards the correct species as it was increasingly favoured by the association ratio.

The model for β (Figure 4.11, top right) had a robust positive intercept ($\text{logit}(\beta) = 80\%$, CI = [85%, 92%], $z = 10.567$, $p < .0001$), meaning that after initially moving towards the correct species, participants selected that species 80% of the time when the association ratio was $\frac{1}{1}$. The association ratio was also a significant positive predictor here ($e^\beta = 2.4$, CI = [1.8, 3.2], $z = 5.678$, $p < .0001$), meaning that participants became more likely to select the correct species on trials where they initially moved towards it as the association ratio in favour of the correct species increased. In general, however, participants very rarely changed direction after moving towards the correct species in any case.

Finally, the model for γ (Figure 4.11, bottom left) contained a significant *negative* intercept ($\text{logit}(\beta) = 37\%$, CI = [30%, 43%], $z = 3.963$, $p < .0001$), indicating that on trials where they initially moved towards the foil species, participants only changed direction to select the correct species instead 37% of the time when the association ratio was $\frac{1}{1}$. The association ratio was again a positive predictor here ($e^\beta = 2.6$, CI = [2.0, 3.5], $z = 6.374$, $p < .0001$), meaning that as the strength of the association ratio in favour of the correct species increased, participants were more likely to switch direction and select the correct species when they initially moved towards the foil.

TIME COURSE

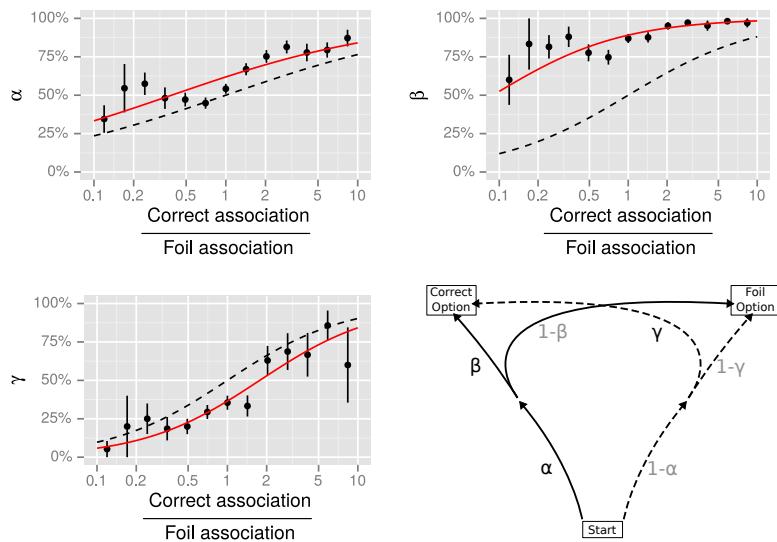


Figure 4.11: Transitions probabilities from Experiment 4, as functions of the association ratios. As the association ratio increased in favour of the correct species, participants became more likely to initially move towards that species (α , top left), to select that species after initially moving towards it (β , top right), and to select it even after initially moving towards the foil (γ , bottom left). Dashed lines show the expected values if participants were influenced by association ratio only. Overview of the three parameters can be seen in the bottom right.

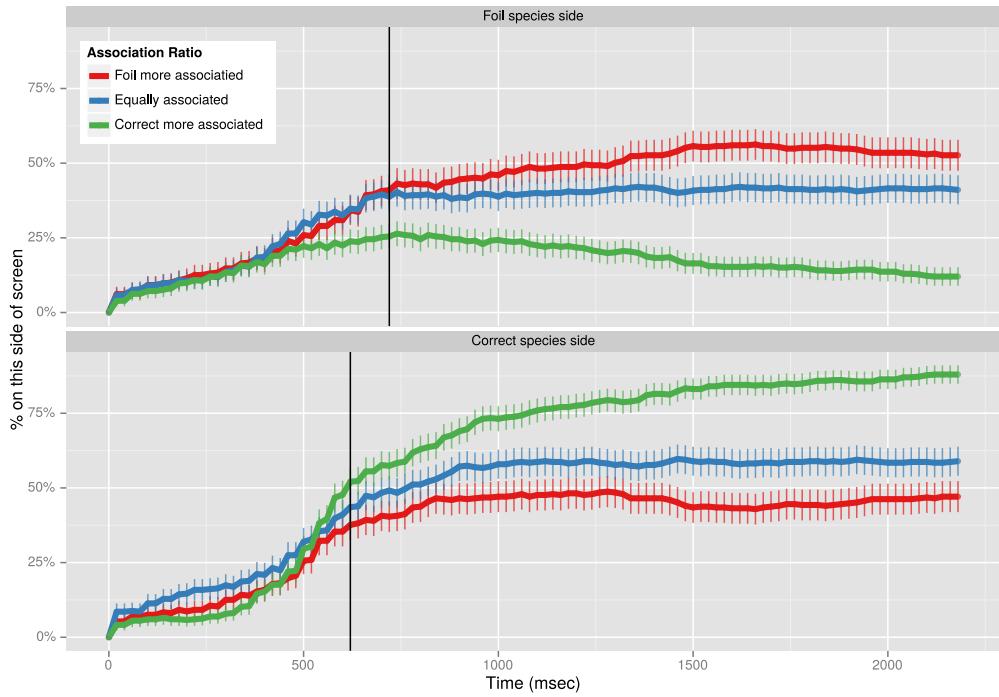


Figure 4.12: Proportion of trials on side of screen corresponding to the correct species (top) and the foil species (bottom) over time. The association ratio is a significant predictor of movements towards the foil from 720 msec, and of movements towards the correct species from 620 msec (solid vertical lines).

As usual, the time course of the mouse movements here reveals more about the points at which participants were drawn towards each response. Figure 4.12 shows the proportion of trials on the side of the screen corresponding to each species over time, for trials where the association ratio favours the foil species ($N = 359$), the two species were rated as equally associated ($N = 397$), or the correct species was more strongly associated ($N = 432$). To estimate divergence times, I fitted two series of logistic mixed models, one series predicting the probability of being on the foil species' side of the screen, and one the probability of being on the correct species' side, separately for every 20 msec time window. Each model included the log-transformed association ratio as a predictor, and included random intercepts for each participant and each base species. The divergence points were the times in each series after which the association ratio is found to be a significant predictor. The association ratio had a significant effect on participants' probability of being on the foil species' side of the screen from 720 msec, and their probability of being on the correct species' side from 620 msec.

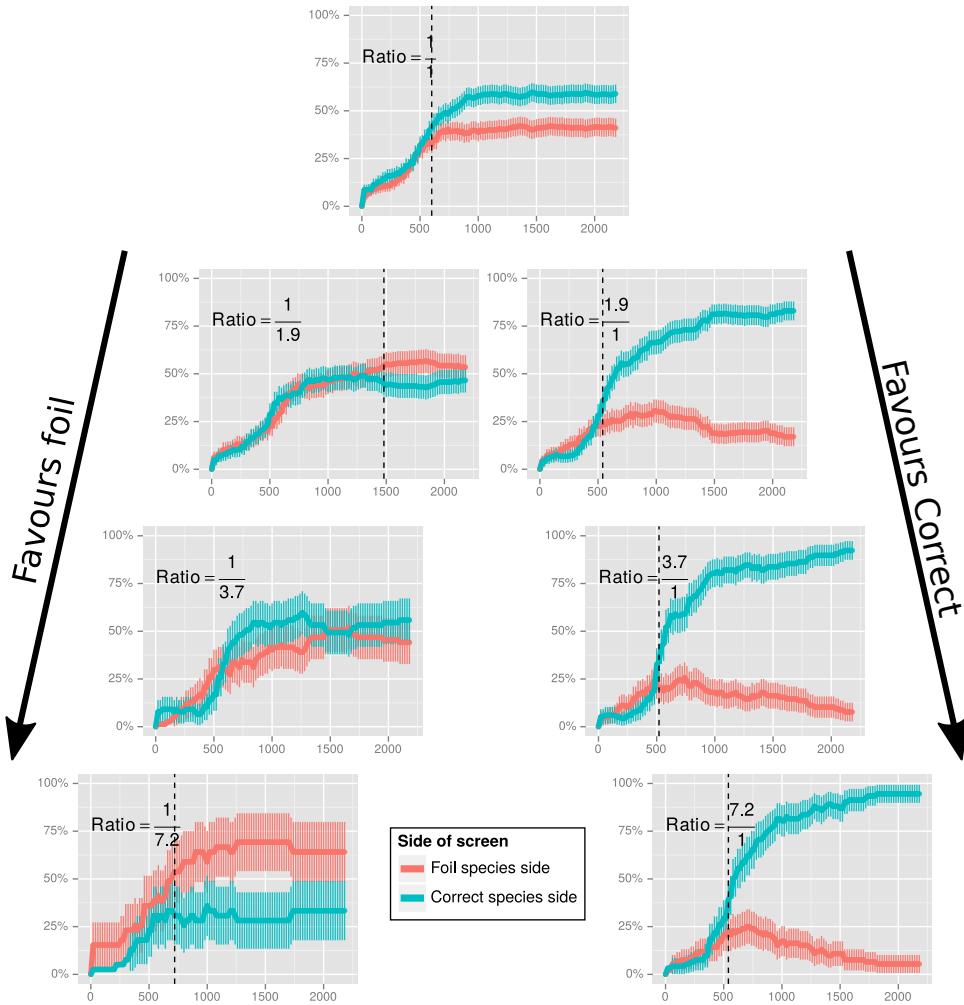


Figure 4.13: Proportion of trials on the side of the screen corresponding to each response, grouped by association ratios. When the association ratio favours the correct option,

Figure 4.13 shows the proportion of trials on either side of the screen, with separate plots representing different levels of the association ratio. For this analysis, I divided the log-transformed association ratio into seven bins of equal width, and labelled each bin according to the mean association ratio within it: $\frac{1}{7.2}$, $\frac{1}{3.7}$, $\frac{1}{1.9}$, $\frac{1}{1}$, $\frac{1.9}{1}$, $\frac{3.7}{1}$, and $\frac{7.2}{1}$. Within each of these bins, I found the divergence point, after which participants were more likely to be on one side of the screen than the other, by fitting a series of logistic mixed models, comparing the probability of being on either side of the screen, with random intercepts for each participant and each base species.

As the association ratio in favour of the correct species increased, from $\frac{1}{1}$ (both

Table 4.4: Divergence times, and number of trials included, for each binned value of the association ratio. For ratios of $\frac{1}{7.2}$ and $\frac{1}{1.9}$, participants were significantly more likely to be on the foil species' side of the screen than the correct species' from the divergence time onwards. For all other ratios, participants were more likely to be on the correct species' side of the screen from the divergence time.

Ratio	Divergence (msec)	N
$\frac{1}{7.2}$	720	39
$\frac{1}{3.7}$	N/A	77
$\frac{1}{1.9}$	1,480	243
$\frac{1}{1}$	600	397
$\frac{1.9}{1}$	540	223
$\frac{3.7}{1}$	520	117
$\frac{7.2}{1}$	540	92

species equally associated) up to $\frac{7.2}{1}$ (correct species 7.2 times more strongly associated than the foil), we see a clear increase in the number of trials where the cursor ends up on the correct species' side, mirroring the earlier analysis of participants' responses. There was little difference, however, in the trends leading up to these final proportions. For all of these bins, the preference for the correct species emerged from 520 – 600 msec, and the shape of the curves were broadly similar, except for the differences in their overall heights. Thus, structured knowledge began to influence participants' motor output from ~600 msec.

When the association ratio was less than $\frac{1}{1}$, however, and so conflicted with structured knowledge, the trends were less clear. When the ratio was strongest in favour of the foil species ($\frac{1}{7.2}$), participants were more likely to move towards the foil than the correct species from 720 msec. Both species were equally attractive with a ratio $\frac{1}{3.7}$, however, and with a ratio of $\frac{1}{1.9}$ a significant preference for the foil only emerged from 1,480 msec. Note, however, that even when the association ratio conflicts with structured knowledge, there is no evidence of a cross-over relationship — in each subplot within Figure 4.13, participants were either drawn to the correct species, or to the foil, but in no case were they primarily initially drawn towards the foil, and later drawn towards the correct species.

4.3.3 DISCUSSION

In general, the results of this experiment are consistent with Experiment 3, but unlike in Experiment 3, I did not exclude any data on the basis of the post-test results. Participants' inferences were influenced by both associative and structured knowledge. Associative knowledge was indexed by the association ratio, reflecting how strongly each participant rated the association between the base and the correct species, compared to the base and the foil. This was a strong predictor of their inferences, with participants more likely to project the property to the correct species as its association with the base increased, relative to the foil. If participants only relied on associative knowledge, however, they should give the correct response only 50% of the time when the association ratio favoured each response equally (ratio of $\frac{1}{1}$). In reality, the fitted model predicted 72% accuracy on such trials, indicating that participants were also influenced by structured taxonomic relationships.

Unlike Experiment 3, the current experiment revealed that response times for correct responses were marginally faster when the association ratio favoured the correct species (i.e. when the associative and structured knowledge cued the same response). There were, however, faster movement initiation times for conflict trials in Experiment 3, a finding that was not replicated here. Of course, these two results are likely related: initiation times are included in total response times, and so faster initiation times under conflict in the previous experiment may cancel out an overall effect on response time. However, my analyses do not rest on these measures, and so I will not attempt to interpret them further.

More interesting are the analyses of the cursor trajectories. First, I found that initial cursor movements were strongly predicted by associative knowledge, and only marginally predicted by structured knowledge: strength of association being equal, participants are expected to move towards the correct species 62% of the time. After moving towards the correct species, participants rarely (20% of the time when the species were equally associated) changed direction to select the foil instead, although they were more likely to do this if the foil was more strongly associated than the correct species. After initially moving towards the foil species, however, participants were somewhat more likely to change direction: they were

predicted to do so 37% of the time when the association strengths were equal, and more often when the correct species was more strongly associated.

From all of this, we can conclude three things. First, associative knowledge predicts participants' movements at every possible juncture, and unsurprisingly, participants are more drawn towards a response option when it is more strongly associated than the alternative. Second, structured knowledge also plays a role at every juncture. Participants were marginally more likely to initially move towards the correct species than would be expected based on associative knowledge alone, and later, participants were in general more likely to change direction to the correct species after initially moving towards the foil (37%) than to change to the foil species after initially moving towards the correct one (20%). Third, once they started moving towards a response option, participants in this experiment were unlikely to change direction.

Analysing the proportion of correct responses where the cursor trajectory was classified as a reversal, an unexpected trend emerged. The data could be modelled using the association ratio as a predictor: correct responses were more likely to involve initial movements towards the foil when the foil was more strongly associated than the correct species. However, the data were slightly better fit by a model that used the magnitude of this ratio (how far it was from a ratio of $\frac{1}{1}$, in either direction), resulting in the inverted U trend seen in Figure 4.10.

However, while this pattern was unexpected, it can be understood in light of the transition probabilities, discussed above. Participants were, first of all, more likely to initially move towards the correct species when the association ratio favoured this response. Additionally, regardless of their initial movement, participants were also more likely to ultimately select the correct species when it was favoured by the association ratio. Combined, these factors produce three kinds of trajectory. When the association ratio strongly favoured the correct species, participants moved straight towards that species, and selected it, rarely moving to the foil at all, and yielding few reversals. When the ratio favoured the foil species, participants who moved towards the foil usually ended up selecting it, leaving few who moved to the foil before selecting the correct species. When the ratio did not favour either species, however, some participants moved towards the foil initially, and some of

these changed their minds, yielding a higher number of reversal trajectories.

Finally, the time course trends here are consistent with those found in Experiment 3. As the association ratio becomes stronger in favour of the correct species, participants became more likely to ultimately select this species, but there were no striking changes in the trends leading up to these responses. Additionally, consistent with the analysis of the transition probabilities, even when associative and structured knowledge conflicted strongly, there was no indication that participants were first driven by associative knowledge, and then by structured knowledge. This is again consistent with the trends seen for Experiment 3.

4.4 GENERAL DISCUSSION

Across two experiments, this chapter investigated how associative and structured knowledge interact during inductive reasoning, using a task that places both forms of knowledge in conflict. Consistent with previous work (Bright & Feeney, 2014b, submitted; Crisp-Bright, 2010, Chapter 5), participants' inferences were influenced by both forms of knowledge. In Experiment 3, participants overwhelmingly generalised genetic properties among species belonging to the same taxonomic group when the foil species was only weakly associated with the base. When the foil was strongly associated, however, participants selected it instead on a number of trials. Similarly, in Experiment 4, while participants were more likely to generalise the property to the taxonomically-related species overall, they were less likely to do so when the foil was more strongly associated with the base than the correct species was.

The current studies go beyond previous work in that I recorded participants' mouse cursor movements as they completed the reasoning tasks. The transition probability analysis showed the proportion of trials where the cursor was moved initially towards each response option, and subsequently the proportion in which participants changed their mind to select the alternative option. In contrast to Experiments 1 and 2, where participants often moved towards one response and then changed their minds, participants here generally moved towards one or other response option, and then stayed there. There were, however, some changes of mind; participants who initially moved towards the foil did sometimes change di-

rection and select the correct species instead. However, again in contrast to Experiments 1 and 2, participants reversed their initial movements towards the foil considerably less than half the time, and additionally, these reversals made up only a small number of the total correct responses. The time course analyses tell a similar story. In Experiments 1 and 2, we saw cross-overs in the time course, as participants were predominantly drawn towards the foil, but then ultimately selected the correct option. Here, on the other hand, whenever the majority of participants were drawn towards the foil, they majority also ended up selecting it.

Together, these patterns would suggest that participants completing these tasks draw selectively on *either* associative knowledge *or* structured knowledge. This stands in contrast to the results of Experiments 1 and 2, on a similar task that pitted perceptual similarity against conceptual knowledge, which suggested that perceptual similarity was usually drawn on first, but later overridden by conceptual knowledge. I compare the results of these two sets of experiments in detail in Chapter 7.

Of course, these results have implications for theories of inductive reasoning. Broadly speaking, participants' inferences were consistent with an account based on structured knowledge: participants recognise that species that belong to the same taxonomic group are more likely to share biological properties than species that do not. In this sense, these results are consistent with a wealth of previous work showing that people use information about category membership, and between-category relationships, when reasoning inductively (S. A. Gelman & Markman, 1986; Murphy, 2004; Murphy & Medin, 1985; Murphy & Ross, 2010; Osherson et al., 1990; Rips, 1975). However, these inferences were also influenced by associative knowledge, a result that is difficult to account for in a theory based on structured knowledge alone. It supports, however, the hybrid theory of induction proposed by Bright and Feeney (2014b; Crisp-Bright, 2010), which allows for both associative and structured forms of knowledge to be used in reasoning.

Finally, the mouse cursor data collected in this experiment are somewhat equivocal. Participants did in some cases move initially towards the foil species cued by associative knowledge before changing direction and selecting the correct species. However, these reversals constituted only a minority of trajectories; on most trials,

participants moved straight to one or other species. As noted above, these results differ from those found in Chapter 3, where reversal trajectories were considerably more common, and these chapters will be compared in detail in Chapter 7.

CHAPTER 5

Descriptions vs. Statistics in Reasoning

5.1 INTRODUCTION

Chapters 3 and 4 explored conflict between different kinds of information in category-based induction. In the reasoning literature, however, there is another form of conflict that has received considerably more attention, between two qualitatively different types of cognitive process. Such conflicts are the domain of dual process theories of reasoning (Evans, 2006; Evans & Stanovich, 2013; Kahneman, 2011; Sloman, 1996), which typically distinguish between fast, effortless, and autonomous *Type 1 processes*, and slower, demanding, and controlled *Type 2 processes*.

These dual process theories have been discussed in detail in Chapter 1. However, it is worth noting again that it is only in recent years that much attention has been given to questions of how Type 1 and Type 2 processes interact during reasoning. Two questions in particular are relevant to the current chapter. First, it is not yet clear how conflict between these processes is resolved, with Evans (2007) outlining three possibilities: a *pre-emptive conflict resolution account* (e.g. Chaiken, 1987; Klaczynski, 2000; Klaczynski & Cottrell, 2004), by which one or other process is selectively activated for a given decision, a *default-interventionist* account (e.g. Evans, 2006; Kahneman & Frederick, 2002) where Type 1 processes provide default responses, which are inspected, and sometimes overridden by Type 2 processes, and a *parallel-competitive* model (e.g. Sloman, 1996, 2014b), where both kinds of process are activated simultaneously, and compete to dictate the response given.

A second question, spurred by the intuitive logic theory (De Neys, 2012, 2014)), concerns the usual assumption in dual process theories that, while Type 1 processes implement the heuristics that lead to systematic biases in human reasoning (Kahneman, 2011; Kahneman & Frederick, 2005), logical reasoning requires the engagement of Type 2 processes. According to this account, Type 1 processes can simultaneously generate both logical and heuristic responses, and as a result that even when we produce heuristic responses, we implicitly and automatically detect the conflict between these multiple Type 1 responses. As discussed in Chapter 1, a number of seemingly implicit measures have supported this *implicit conflict monitoring* prediction: on tasks in which the heuristic and logically-valid response conflict, even when producing heuristic responses, participants are, for instance, slower to respond (De Neys & Glumicic, 2008), less confident in their heuristic responses (De Neys et al., 2013), sweat more during reasoning (De Neys et al., 2010), look more at the logically-relevant information (De Neys & Glumicic, 2008), and even appear to “like” logically valid syllogisms more than invalid ones (Morsanyi & Handley, 2012). Explicit measures such as participants’ verbal reports, on the other hand, rarely show such evidence of conflict (De Neys & Glumicic, 2008), suggesting that conflict is detected by Type 1, rather than Type 2 processes. These questions are relevant to reasoning in every domain. However, the *base rate neglect* paradigm in particular is extensively used as a testing ground both for dual process accounts more generally, and for the intuitive logic theory specifically. It is this paradigm I use here.

5.1.1 BASE RATES IN REASONING

In the psychology of judgement and reasoning, few phenomena have received as much attention as base rate neglect. The phenomena was introduced by Kahneman and Tversky’s (1973) *Tom W.* problem. In its original form, participants read a description of Tom, a randomly chosen student:

Tom W is of high intelligence, although lacking in true creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place. His writing is rather dull and mechanical, occasionally enlivened by somewhat corny puns and by

flashes of imagination of the sci-fi type. He has a strong drive for competence. He seems to have little feel and little sympathy for other people and does not enjoy interacting with others. Self-centered, he nonetheless has a deep moral sense.

Participants were given a list of undergraduate degrees, and variously required to indicate what proportion of American students were enrolled in each degree (the *base rate* of each degree), how similar Tom W. was to a typical student of each subject, or *how likely Tom W. was to be studying for each degree*. Kahneman and Tversky (1973) found that likelihood judgements for each degree correlated almost perfectly ($r = .97$) with similarity judgements (Tom sounds like a typical computer science student) but were unrelated to the base rate judgements (more students study the humanities than anything else). In other words, when asked to make a decision about likelihood, participants responded with a judgement of similarity, or *representativeness* and completely ignored their knowledge about the statistical base rate for each degree. Thus, Kahneman and Tversky's (1973) participants, like many since, displayed *base rate neglect*.

Following this first demonstration, researchers have created a range of base rate problems, where statistical base rate information is available, but typically ignored or underweighted in favour of other evidence. One class of problem pits base rate probabilities against statistics directly about the object at hand. This chapter, however, focuses on a different kind of base rate problem, where base rates are typically ignored in favour of qualitative information, such as a description that allows participants to base their response on how representative something or someone is of a given category (Kahneman & Frederick, 2002; Kahneman & Tversky, 1973; Tversky & Kahneman, 1974). In their original forms, such problems were typically quite complex, such as the example seen at the start of this chapter, from Kahneman and Tversky (1973). In the same paper, they present a simpler test of participants' use of base rates. Participants were told of a group containing 70 engineers, and 30 lawyers (or 30 engineers, and 70 lawyers), and presented with a description of a randomly chosen member of this group, for instance:

Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest

in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.

Participants were asked to indicate the probability that Jack was either an engineer, or a lawyer. Across five descriptions, participants' ratings were minimally influenced by the base rates. When the sample consisted of 30 engineers and 70 lawyers, participants on average said that there was a 50% probability of Jack being an engineer. When it consisted of 70 engineers and 30 lawyers, the average probability was 55%. Therefore, the base rate information was almost completely ignored. Even with an uninformative description (but not with a person for whom no description at all was provided), base rate information was used minimally, or not at all.

More recently, base rate neglect, along with many other phenomena in the heuristics and biases literature, has been reinterpreted in terms of dual process accounts of cognition (see Barbey & Sloman, 2007; Kahneman, 2011; Kahneman & Frederick, 2002, 2005). From this perspective, Type 1 processes are responsible for responses based entirely on representativeness, while Type 2 processes are required to override the description-cued response, and to integrate this information with base rates/prior probabilities. Of course, as noted by Stanovich and West (2000), use of base rate information requires that participants a) are aware of its relevance (see also Bar-Hillel, 1980), b) are motivated to attempt to use it, c) have the cognitive capacity to hold both kinds of information in working memory, and d) can inhibit the intuitively appealing non-base rate response.

More recently still, De Neys and Glumicic (2008) introduced a simplified version of the lawyer-engineer problem that, like the original, placed base-rates and representative descriptions directly in conflict. Participants were told about a different sample on each trial, that contained, for instance, 955 of one social group and 5 of the other. They were again given a description of a "randomly chosen" person from that sample. Their task, in this simplified forced-choice paradigm, was to indicate which group they thought the person described belonged to, rather than give a probability estimate, as in the original version of the task. For each trial, the base rate and description either suggested the same response, or conflicted.

While protocol analysis of participants' "thinking aloud" during reasoning revealed little evidence that they explicitly considered the base rate information, they were nevertheless more likely to correctly remember base rate information from conflict trials, where base rates and descriptions supported different responses, suggesting greater processing of such information those trials, even if not to a degree that participants were explicitly aware of. In a second experiment, De Neys and Glumicic (2008) also showed that even when consistently giving the response consistent with the description, participants took longer to do so, and were more likely to look back at the base rate information, when the cues conflicted. Again, this indicates that participants processed the base rate information, and experienced conflict as a result, on these trials, even when it rarely affected their responses.

The simplified base rate task presented by De Neys and Glumicic (2008) has become a popular tool for testing theories of conflict in reasoning. In an fMRI study, De Neys et al. (2008) presented participants with conflict and no-conflict problems, along with additional control conditions in which either the description was uninformative, or the base rates were equal (500:500). Behaviourally, they found that when the cues conflicted, participants selected the base rate-cued response on 45% of trials. Additionally, they found that both kinds of response were slower under conflict than any other condition, suggesting that even when participants based their decisions under conflict on the description provided, rather than the base rate, they were to some degree slowed by the conflict between the two cues. De Neys et al. (2008) also presented neuroimaging results supporting this interpretation, with greater activation of the anterior cingulate cortex, thought to reflect the detection of conflict (M. M. Botvinick et al., 2004), on conflict trials regardless of the response given than in any other condition. They also found greater activation of the dorsolateral prefrontal cortex, usually an indicator of inhibitory processes (Aron et al., 2004), for base-rate consistent responses under conflict compared to description consistent responses. This suggests that responding based on the base rate requires inhibition of the description-cued response.

A number of other studies have used this forced-choice version of the base rate paradigm. Franssens and De Neys (2009) showed that while placing participants under secondary load led them to be even less likely to select the base rate-cued

response, their recognition memory for the base rate information was unaffected, suggesting that this conflict detection is not cognitively demanding. In a similar vein, De Neys and Franssens (2009) asked participants to complete a lexical decision task after conflict and control base rate trials. They showed that in the lexical decision task, participants were slower to identify words used in descriptions that conflicted with base rates. This effect was found even for those participants who predominantly failed to give the base rate-cued response on such trials, suggesting that even these participants were to some degree attempting to inhibit the description, although unsuccessfully. De Neys et al. (2011) also showed that participants were less confident in their description-cued responses when the base rate disagreed with them, again suggesting some awareness of this conflict.

Collectively, these results provide support for De Neys' (i.e. De Neys & Glumicic, 2008; De Neys et al., 2008) claim that heuristic reasoners are aware that their responses are biased, at least on the base rate neglect paradigm (see Chapters 1 and 6 for evidence for this claim in other paradigms). In the last few years, a number of studies have provided evidence for a stronger claim made by De Neys (2012, 2014), that intuitive Type 1 processes can simultaneously cue both heuristic (description-based, in the base rate paradigm) and logical (base rate-based) responses. Of course, this claim goes against classical dual processes accounts of base rate neglect (Barbey & Sloman, 2007; Kahneman & Frederick, 2005), which claim that Type 2 processes underlie base rate-based responding.

In one study, Pennycook and Thompson (2012) presented participants with problems featuring no base rate information, and problems with base rates that were either consistent or conflicted with the descriptions. Participants were asked to judge the probability of the person described belonging to one or other group. They then analysed the distribution of these probability judgements. Using problems without base rate information as a baseline, they found that when base rates agreed with the descriptions, the probabilities were shifted in the direction consistent with both of these cues. When the base rates disagreed with the descriptions, however, the distribution of the probabilities became bimodal, as participants gave either a rating consistent with the description, or one consistent with the base rate. Crucially, they argue that this pattern of results requires that par-

ticipants always process the base rate, in order that they should integrate it with the description when they agree, or decide to rely on one or other cue when they conflict.

A particularly compelling result is reported by Pennycook, Trippas, Handley, and Thompson (2014), who presented participants with base rate problems where the cues either conflicted or agreed, and instructed them to either base their response on the base rate, or on the descriptions. They analysed the probability judgements participants gave, their confidence in these judgements, and their response times. They found that conflict occurred in both directions. Regardless of what cue they were told to use, participants probability judgements were affected by manipulations of the other cue. These interference effects persisted even when participants responded under time pressure, where they were also slower to respond under conflict, regardless of what cue they were using. Recall that classical dual process accounts (Barbey & Sloman, 2007; Evans, 2006) hold that processing of descriptions is quick, effortless, and obligatory, driven by Type 1 processes, while processing of base rates is slow, effortful, and deliberate, driven by Type 2 processes. These results would suggest that processing both kinds of information can be quick, effortless, and obligatory. Handley and Trippas (2015) further develop this idea, arguing that much of the information processing usually ascribed to Type 1 or Type 2 processes can actually be performed by both sets of processes.

5.1.2 WHAT MOUSE TRACKING CAN ADD TO THE DEBATE

The work cited above goes some way towards revealing the processes that underlie base rate neglect and respect. However, questions remain, which in this chapter I will attempt to address using the mouse tracking paradigm.

To date, work in the base rate problem has been typically based on the analysis of certain form of data, specifically, binary choices, probability judgements, confidence ratings, and response latencies, as well more subtle measures such as recall of base rate information (De Neys & Glumicic, 2008), or subsequent lexical decision times for words from the description (De Neys & Franssens, 2009). The mouse tracking paradigm, however, differs from these methods in that it both allows for the detection of conflict between competing responses on a single trial,

and provides some insight as to the nature of this conflict — for instance, initial movements towards one response option that are later overridden, or changes in the time taken to select a response, in the absence of such cursor reversals.

Naturally, different accounts of the base rate paradigm yield different predictions about what mouse tracking will reveal here. Usually, studies of base rate neglect are couched in terms of dual process theories. However, as discussed in Chapter 1, such theories come in many varieties. First, selective, pre-emptive conflict resolution accounts (e.g. Klaczynski, 2000) would predict that participants engage either Type 1 processes sensitive to the contents of the description, *or* Type 2 processes also sensitive to the base rate information. Such an account would predict that participants take longer to give the base rate-cued response (using Type 2 processes) than to give to the description-cued response (using only Type 1 processes). However, because only one or other kind of process is activated, it would not predict that participants should be conflicted. Therefore participants giving the base rate-cued response should not be slower to do so, or more likely to move towards the alternative option, if the description cues the alternative response. Likewise, participants giving the description-cued response should not be influenced by manipulations of the base rate.

A default-interventionist account (Evans, 2006; Evans & Elqayam, 2007; Kahneman & Frederick, 2002, 2005), on the other hand, would predict that participants are initially driven by Type 1 processes to give description-cued responses. However, this default may be overridden by Type 2 processes that take into account base rate information later in the process. The predictions made by such an account are clear: in addition to the latencies discussed above, early cursor movements should be driven by the contents of the description, but the base rate may become influential later. Therefore, participants giving the base rate-cued response should be slower to respond or initially drawn towards the alternative option when the description cues it instead, and faster when the description agrees with the base rate. However, because these accounts hold that Type 2 processes are only engaged when they override Type 1 responses, participants responding based on the description should be unaffected by manipulations of the base rate.

A number of previous findings, however, discussed above, do suggest that con-

flict occurs in both directions — including many results indicating that participants are sensitive to a conflict with the base rate even when giving the description-cued response (De Neys & Glumicic, 2008; De Neys et al., 2008; Pennycook, Fugelsang, & Koehler, 2012), and Pennycook et al.’s (2014) finding that regardless of which cue participants were told to use, their judgements were influenced by the other cue. One theory that may account for such findings is the parallel-competitive dual processes theory (i.e. Sloman, 1996, 2014b), which claims that both Type 1 and Type 2 processes are activated simultaneously, and compete to produce responses. Alternatively, and more commonly, these findings can be seen as evidence for the intuitive logic theory (De Neys, 2012), by which account Type 1 processes can simultaneously cue both description-based and base rate-based responses, leading to both the implicit detection of conflict, and the interference when the cues conflict, regardless of the response actually given. The account proposed by Handley and Trippas (2015) is in many ways similar. They suggest that both base rates and descriptions can be processed by both Type 1 and Type 2 processes. All of these accounts yield largely the same predictions here: under conflict, participants should be drawn to the competing response, regardless of what response they ultimately give. In other words, beyond the effect of descriptions on base rate-cued responses predicted by the other accounts, they predict that participants giving the description-cued response should be affected by manipulations of the base rate, and show greater evidence of conflict when it disagrees with the description. These accounts do, however, allow for some asymmetries: the description-cued response is thought to be *stronger*, either because it is generated by the faster Type 1 processes (Kahneman & Frederick, 2002, 2005), or because it is the *prepotent* of the two Type 1 generated responses, and so its interference on base rate-cued responses may be greater than the interference in the other direction (see De Neys, 2012). In this experiment, I use the mouse tracking paradigm to test these predictions.

5.2 METHOD

5.2.1 PARTICIPANTS

Fifty undergraduate students at Queen's University Belfast took part in the experiment in return for course credit.

5.2.2 STIMULI

Stimuli were adapted from De Neys et al. (2008), and consisted of forty reasoning problems where participants were presented with the base rates for two social categories within a sample population of one thousand people, followed by a description of a randomly chosen member of that sample, and asked to use this information to decide which social category the chosen individual was most likely to belong to. The base rate information could either indicate a large majority of the sample belonging to one of the groups (995 vs. 5), or an equal distribution of the two groups (500 vs. 500). The description could either match common stereotypes about one or other of the groups, or be totally uninformative.

Questions were presented in four conditions, with 10 questions in each: the base rate and description could both suggest the same response (*congruent trials*), or could disagree (*incongruent trials*), the base rate could be uninformative (500 people from each group; *description only trials*), or the description could be uninformative (equally consistent with both groups; *base rate only trials*). Examples of each kind of trial are shown in Table 5.1.

The same ten questions, with deliberately uninformative descriptions, were used for the base rate only trials for all participants. The remaining thirty questions were randomly allocated base rates (agreeing with the description, disagreeing, or uninformative) for each participant.

5.2.3 PROCEDURE

The experiment was programmed using the OpenSesame program (see Chapter 2). The 40 trials were presented in a random order for each participant. At the beginning of each trial, the base rate information was presented in a boxed region

Table 5.1: Examples of the each kind of problem used in Experiment 5. On congruent trials, the base rate and description cued the same response. On incongruent trials, they cued different responses. On base rate only trials, the description was uninformative, and only the base rate cued a response. On description only trials, the base rate was uninformative (500:500), and so only the description cued a response.

Congruent trials	
Base rates:	995 Engineers; 5 Lawyers
Description:	Jack is 36. He is not married and is somewhat introverted. He likes to spend his free time reading science fiction and writing computer programs.
Question:	What is Jack's occupation?
Options:	Engineer ^{BR,D} ; Lawyer
Incongruent trials	
Base rates:	995 violinists; 5 rappers
Description:	Jason is 20. He grew up in a poor family in a neglected neighbourhood in Birmingham, and didn't finish his A-levels.
Question:	What is Jason?
Options:	Violinist ^{BR} ; Rapper ^D
Base rate only trials	
Base rates:	995 Rolling Stones fans; 5 Beatles fans
Description:	Mark is 43 years old. He weighs 12 stone and is 5 ft 9 inches tall. He has one younger brother and lives in Manchester.
Question:	Which band does Mark prefer?
Options:	The Rolling Stones ^{BR} ; The Beatles
Description only trials	
Base rates:	500 60-year-olds; 500 30-year-olds
Description:	Gladys is a quiet woman. She lives in a little house with her Yorkshire Terrier where she spends most of her time knitting.
Question:	What age is Gladys?
Options:	60 ^D ; 30

Note: ^{BR} Base rate-cued option; ^D Description-cued option

in the top centre of the screen. After 4 seconds, the description text was also displayed, below the base rates. After a further 5 seconds, a button marked “NEXT” appeared in the bottom centre of the screen. On clicking this, a fixation cross was displayed for 600 msec. After this, the probe question was shown in large font in the centre of the screen, and the two possible response options appeared as buttons in the top left and right corners. At the same time, a timer was displayed below the probe question, which filled up from the bottom over the course of the next 6 seconds, in which time participants were required to give their response by clicking on one of the response buttons. With the onset of the probe question, the mouse cursor was reset to the centre of the “NEXT” button. A screen-shot from a trial is shown in Figure 5.1.

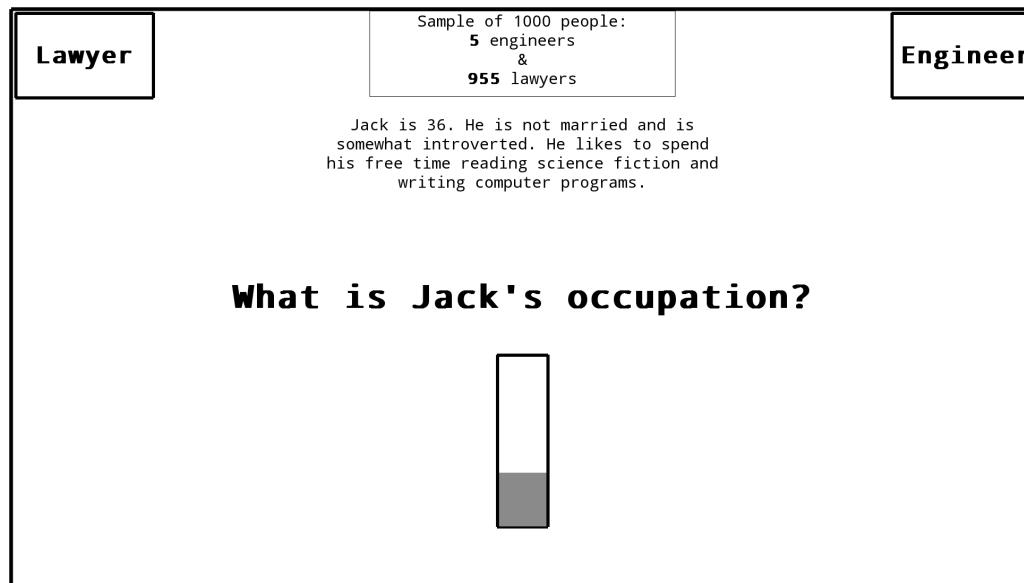


Figure 5.1: A screen shot from Experiment 5. The description suggests that Jack is more likely to be an engineer (the response option on the left), but the base rates suggest he is more likely to be a lawyer (the option on the right). The green timer in the centre of the screen is empty at the start of each trial, and fills up from the bottom over six seconds.

Following trials in which no response was given within 6 seconds ($N = 2$), participants were asked to try to respond more quickly. On trials where participants did not move the mouse cursor from the “NEXT” button within 2 seconds of the onset of the probe question ($N = 4$), a message reminding them that they were under time pressure and asking them to “try to start moving as soon as you see the target”, was shown.

For each trial, I recorded the time taken to finish reading the description and click on the “NEXT” button (*reading time*, measured from the appearance of the “NEXT” button, 5 seconds after the onset of the description itself), the response chosen, the time from the onset of the probe text taken to respond, and the position of the mouse cursor, sampled at 50 Hz (every 20 milliseconds).

5.3 RESULTS

The current results section follows the same format used previously in this thesis; I first present analyses of participants’ responses, followed by measures of conflict on trials where a particular response was given, then the cursor transition probabilities, and finally the time course data. Each of these analyses is broken into three sections. First, I investigated what happens on trials where both cues conflict. Then, I tested the effect of descriptions on base rate-driven reasoning.

There were three kinds of trial in which participants could use base rate information in this task: trials where both the base rate and the description cued the same response (i.e. the description agreed with the base rate), trials where the base rate provided the only useful information (the description was uninformative), and trials where the description cued the opposite response (the description disagreed with the base rate). Most dual process theories would predict that descriptions, being more easily and readily processed, should interfere with base rate-based reasoning.

I also test the effect of base rates on description-cued reasoning. There were also three types of trial where participants could rely on the description: trials where the description and the base rate cued the same response (i.e. the base rate agreed with the description) trials where only the description was useful (the base rate was uninformative; 500:500), and trials where the base rate cued the opposite response (the base rate disagreed with the description).

Analyses were conducted using multilevel models, with random intercepts for each participant, and for each problem, unless otherwise stated. As before, log-linear models were used for the latency data (reading time, movement initiation time, and response time), and logistic models for analysis of participants’ responses, and the transition probability analysis.

5.3.1 RESPONSES

Table 5.2 shows the proportion of responses consistent with each cue in each condition. When the cues disagreed, participants selected the description-cued response on 80.4% of trials, and the base rate-cued response on 19.6%, a statistically significant difference ($z = 5.958$, $p < .0001$).

Table 5.2: Responses consistent with each cue, according to whether the other cue agreed with it, is uninformative, or disagreed with it, in Experiment 5.

Description	Base rate responses	Base rate	Description responses
Agreed	94.4%	Agreed	94.4%
Uninformative	69.4%	Uninformative	93.6%
Disagreed	19.6%	Disagreed	80.4%

EFFECT OF DESCRIPTIONS ON BASE RATE CHOICES

Descriptions had a significant main effect on the probability of participants' giving the base rate-cued response ($\chi^2 = 741.6$, DF = 2¹, $p < .0001$). Participants were significantly more likely²to give the base rate cued response when the base rate and description agreed (94.4% of trials) than when the description was uninformative (69.4%) or when the description disagreed with the base rate (19.6%) They were also significantly more likely to do so when the description was uninformative than when the description disagreed (z 's > 8, p 's < .0001).

Figure 5.2 plots the number of base rate cued responses given by each participant, by condition, and indicates that these effects seem to be operate consistently across all participants, rather than being driven by a subset of participants who are influenced by the description. Of the 50 participants, 49 were less likely to give the base rate cued response when base rates and descriptions disagreed than when they agreed, and 1 always gave the base rate cued response.

¹ Recall from Chapter 2 that Chi-squared tests are used in place of ANOVA F tests for mixed models analysis — a significant chi-squared test denotes a significant main effect of condition here.

²All post-hoc comparisons report Tukey adjusted p values and confidence intervals.

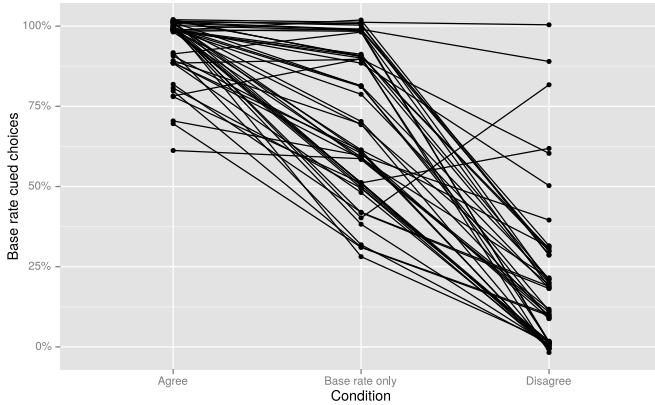


Figure 5.2: Number of base rate responses per condition, by participant, in Experiment 5.

EFFECT OF BASE RATES ON DESCRIPTION CHOICES

Base rates had a significant main effect on the number of description-cued responses given ($\chi^2 = 67.7$, DF = 2, $p < .0001$). Pairwise comparisons showed that participants were less likely to give the description-cued response when the base rate disagreed with the description (80.4%) than when it agreed (94.4%; $e^\beta = 0.20$, CI = [0.12, 0.36], $z = 6.655$, $p < .0001$) or when the base rate was uninformative (93.6%; $e^\beta = 0.24$, CI = [0.14, 0.42], $z = 6.128$, $p < .0001$). There was no difference between trials in which the base rate agreed and those where the base rate was uninformative ($z < .3$, $p > .7$), with both conditions close to ceiling. Therefore, participants were at least in some way sensitive to the base rate information, and in a minority of cases relied on it instead of the description when the two conflicted.

Once again, it is useful to investigate individual differences in this effect. The left panel of Figure 5.3 shows the proportion of description-cued responses given by each participant by condition. This shows three participants who were substantially less likely to give the description-cued response when the base rate disagreed with it. The right panel shows the difference, for each participant, in the number of description-cued responses given from when the base rate agreed to when it disagreed. This shows that, along with the three participants who showed a substantial change, most participants showed a moderate effect in the same direction: they were slightly less likely to select the description-cued response when the base rate disagreed with it, suggesting that base rates had some effect on the majority

of participants. Out of 50 participants, 28 were less likely to select the description-cued response when the base rate disagreed, 13 were equally likely to do so (including 10 who always gave this response), and 9 were actually more likely to select the description-cued response when the description disagreed.

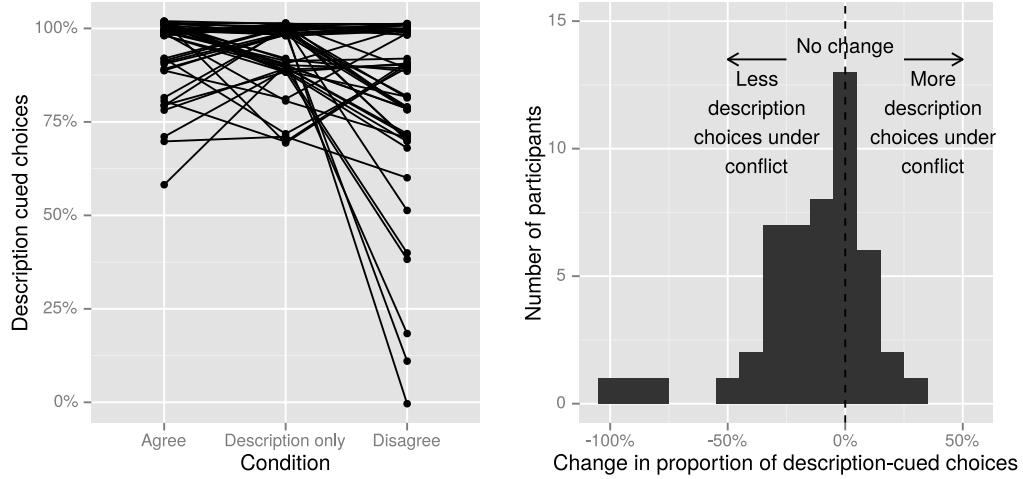


Figure 5.3: Left: Proportion of base rate-cued responses given in each condition, by participant. Right: Difference in the proportion of description-cued responses given by each participant between when the base rate agreed and disagreed with the description. Values below 0 indicate a participant was less likely to select the description-cued response when it conflicted with the base rate than when it agreed.

5.3.2 DESCRIPTIVE STATISTICS

Next, I calculated a number of descriptive statistics for each trial. Maximum deviation of mouse cursor trajectories was bimodally distributed (see Chapter 2; Bimodality Coefficient = .567; Hartigan's D = .0224, N = 1499, $p < .0001$) and so trajectories were classified as either direct or reversals, as described in Appendix B (reversal cut-off: $MD > 0.972$).

When the cues conflicted, participants who gave the base rate-cued response spent more time reading the description (3,229 msec, $SD = 2,943$) than those who gave the description-cued response (2,406 msec, $SD = 2.789$; $e^\beta = 125\%$, CI = [104%, 150%], $t(484.0) = 2.368$, $p = .018$), and were slower to initiate their mouse movements (456 msec, $SD = 266$ vs. 476 msec, $SD = 326$; $e^\beta = 119\%$, CI = [110%, 129%], $t(486.8) = 4.496$, $p < .0001$), but did not differ in their response times ($p > .5$). Participants were also more likely to initially move towards the description-

cued option before selecting the base rate-cued one (38% of base rate responses) than to initially move towards the base rate-cued option and ultimately select the description-cued one (27% of description-cued responses; $e^\beta = 1.7$, CI = [1.03, 2.9], $z = 2.109$, $p = .035$).

EFFECT OF DESCRIPTIONS ON BASE RATE CHOICES

Table 5.3 shows these descriptive statistics for trials where the base rate response was given, broken down for each type of description. Table 5.4 shows effect sizes for the main effect of description type, and pairwise comparisons between descriptions, for each variable shown in Table 5.3.

Table 5.3: Descriptive statistics for trials where the base rate-cued response was given in Experiment 5. All times are in msec. Standard deviations are shown in parentheses. Trials are broken down according to whether the description agreed with the base rate, was uninformative, or disagreed with the base rate.

Note: N = Number of trials in which this response was given.

Description	Reading time	Initiation time	Response time	Reversals	N
Agreed	2,186 (2,153)	444 (247)	1,372 (522)	26%	472
Uninformative	3,046 (2,890)	451 (241)	1,509 (666)	31%	347
Disagreed	3,229 (2,943)	476 (326)	1,776 (739)	35%	98

There was a significant main effect of descriptions on reading times ($\chi^2 = 21.7$, $N = 2$, $p < .0001$). When giving the base rate-cued response, participants were significantly faster to read the description when it agreed with the base rate (2,186 msec, SD = 2,153) than when it was uninformative (3,046 msec, SD = 2,890; $e^\beta = 76.1\%$, CI = [60.9%, 95.2%], $t(33.4) = 2.841$, $p = .0202$), or when it disagreed with it (3,229 msec, SD = 2,943; $e^\beta = 70.8\%$, CI = [60.1%, 83.5%], $t(878.3) = 4.111$, $p < .0001$). Reading times did not differ between uninformative descriptions and those that disagreed with the base rate ($t < .7$, $p > .8$).

Initiation times did not differ according to description type ($\chi^2 < .5$, $N = 2$, $p > .8$). However, there was a main effect of description on response times ($\chi^2 = 34.9$, $N = 2$, $p < .0001$). Participants were faster to give the base rate-cued response when the description agreed with the base rate (1,372 msec, SD = 522) than either when the description was uninformative (1,509 msec, SD = 666) or when descrip-

Table 5.4: Main effects and pairwise comparisons between conditions for the effect of condition on the descriptive statistics shown in Table 5.3, from trials where the base rate-cued response was given. For the main effect, the χ^2 statistic is shown, which was subjected to a chi-squared test with 2 degrees of freedom in each case. For pairwise comparisons, the exponentiated regression weights e^β are shown, reflecting the percentage change in the variables that were log transformed (reading, initiation, and response times), and the odds ratio change for the binary outcome (reversals).

Comparison	Reading time	Initiation time	Response time	Reversals
Main effect (χ^2)	21.68 ****	0.42	34.93 ****	4.78 ·
Agreed/Uninformative	76% *	98%	93% ***	0.7
Agreed/Disagreed	70% ***	97%	82% ***	0.7
Uninformative/Disagreed	93%	99%	89% **	0.95

Note: · $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$; **** $p < .0001$

tion disagreed with the base rate (1,776 msec, SD = 739). They were also faster when the description was uninformative than when it disagreed with the base rate. All comparisons were statistically significant (t 's > 3.2 , p 's $< .01$).

There was a marginally significant effect of descriptions on reversals — the probability of initially moving towards the alternative option before giving the base rate response ($\chi^2 = 4.8$, $N = 2$, $p = .0916$). Despite this, none of the post-hoc comparisons approached significance. However, the trends were in the direction predicted by dual process accounts. On trials where participants gave the base rate-cued response, a greater proportion changed their mind while doing so when the description cued the opposite response than when the description was uninformative, and fewer again did so when the description agreed with the base rate.

In summary, consistent with what would be predicted by most dual process accounts, participants were indeed slower to give the base rate-cued response when it went against the contents of the description, and spent longer reading the description for such problems. For trials where the base rate-cued response was given, descriptions did not have a significant effect on participants' initial cursor movements, in contrast to previous experiments. This pattern, however, can be better understood by considering the transition probability analyses, below.

EFFECT OF BASE RATES ON DESCRIPTION CHOICES

Table 5.5 shows descriptive statistics for trials in which the description-cued response was given, by condition. Contrary to a number of previous studies that found evidence that base rates interfere with participants' description-cued responses, there was no main effect of condition found on any of the descriptive measures here ($\chi^2 < 3.4$, DF = 2, p > .18). Focusing on response time, the measure most commonly analysed in previous studies, reveals a weak trend in the expected direction. Response times when both cues agreed (1,372 msec, SD = 522) were almost identical to when only the description was informative (1,369 msec, SD = 464), but slightly faster than when the base rate cued the opposite response (1,402 msec, SD = 563).

Table 5.5: Descriptive statistics for trials where the description-cued response was given in Experiment 5. Trials are broken according to whether the base rate agreed with the description, was uninformative, or disagreed with the description.

Note: N = Number of trials in which this response was given.

Base rate	Reading time	Initiation time	Response time	Reversals	N
Agreed	2,186 (2,153)	444 (247)	1,372 (522)	29%	472
Uninformative	2,309 (2,260)	447 (258)	1,369 (464)	28%	467
Disagreed	2,406 (2,789)	456 (266)	1,402 (563)	27%	401

In previous tests of the intuitive logic theory (e.g. De Neys et al., 2011; Mevel et al., 2014, see also Pennycook, Fugelsang, & Koehler, 2015), researchers have explored the relationship between individual differences in overt sensitivity to normative principles (i.e. the proportion of normatively correct responses given on conflict trials) and implicit measures of conflict. Therefore, I calculated, for each participant, the proportion of description-cued responses on conflict trials. Of the 50 participants, 26 gave the description-cued response on 9 or 10 of the 10 conflict trials, and were classed as *description-driven reasoners*. The remaining 24 participants gave the description-cued response on average 62% of the time, and were classed as *equivocal reasoners*. Only 4 of these participants gave the base rate-cued response on more than half of the conflict trials.

I repeated the analyses described above — analysing reading time, movement

initiation time, response time, and proportion of reversals, on trials where the base rate-cued response was given — using 3 (description: agreed, uninformative, or disagreed with base rate) by 2 (type of reasoner: description-driven or equivocal) factorial mixed models, with random intercepts for each participant and each problem. None of these models showed significant description by type of reasoner interactions (p 's $> .1$), indicating that there were no substantial individual differences in the null effects of manipulating the description reported above.

However, the analyses reported here are in some ways conservative. Working within the null hypothesis significance testing (NHST) framework, I have calculated p values, comparing the main effect of base rate types on participants' response times to what would be expected under the null hypothesis. Because there were three kinds of base rate — they could agree with the descriptions, be uninformative, or disagree — I then performed pairwise comparisons, using the Tukey HSD method to adjust the p values to maintain a Type 1 error rate of 5%. While I was unable to reject the null hypothesis — that base rates did not influence participants on trials where they gave the description-cued response — this framework does not make it possible to accept this null hypothesis either.

Therefore, I conducted a follow up analysis, comparing responses where the base rate agreed with the description to those where it disagreed, and used Bayesian inference to estimate the magnitude of the difference in response time between these two conditions. Participants on average responded in 1,372 msec (SD 522 msec) when the cues agreed, and 1,402 msec (SD 563 msec) when they disagreed, a slowdown of 30 msec, or 2.1%.

I fit a Bayesian multilevel model using Stan (Stan Development Team, 2015), where log-transformed response times for description-cued responses when the base rate agreed with the description were compared to those where the base rate disagreed. The model included random intercepts for each participant, and for each problem. Uninformative priors were set on all model parameters. The full model specification can be found in Appendix L.

After sampling,³ the posterior distribution for the intercept term (the mean

³ I ran 8 MCMC chains of 1000 iterations each, with 500 warm-up draws discarded from each chain.

response time when the base rate agreed), had a median estimate of 1,357 msec and a 95% credible interval⁴ of [1,285 msec; 1,423 msec]. More importantly, posterior on the regression weight e^{β} is shown in Figure 5.4. The median estimate was 102.8%, indicating a 2.8% slowdown under conflict, and the 95% credible interval was [99.0%; 106.7%]. Therefore, while the most credible effect is a very small slowdown under conflict, there is considerable uncertainty around this effect. We can further understand this estimate by looking to the mass of the posterior above and below certain values. For instance, we can be 91.9% certain that response times under conflict were *slower* than those when the cues agreed — that is, $e^{\beta} > 100\%$ — but also 99.1% certain that the slowdown did not exceed 10% of the response time when the cues agreed. Another way of interpreting a posterior of this sort (Kruschke, 2011) is to define a range of effect sizes that we would consider to be trivial — that is, to define a *region of practical equivalence*; ROPE. For us to be 95% confident that this effect constitutes a null result, we would need to define all changes in response time from 94% to 106% as being essentially null effects. As I believe that effects within this range would be of theoretical importance, the current results do not support the null hypothesis.

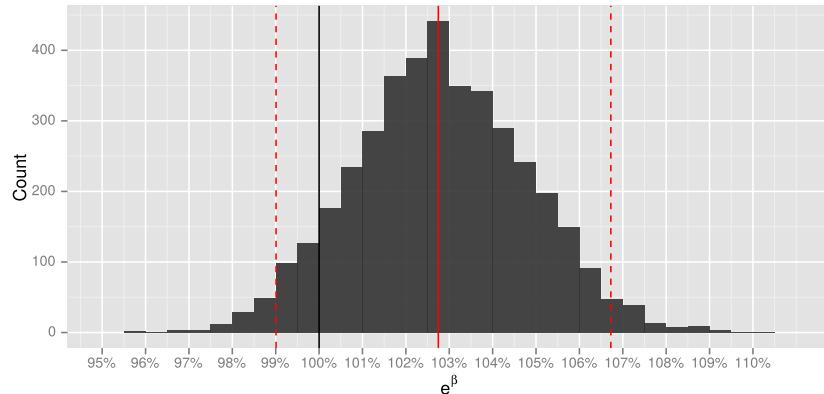


Figure 5.4: Posterior samples for the e^{β} regression weight, reflecting the difference in response times between description-based responses when the base rate agreed with the description and those when the base rate disagreed. The median estimate of 102.8% (solid red line, dashed lines show 95% credible intervals) indicates that participants were 2.8% slower when the base rate disagreed. The solid black line shows the null effect.

⁴ A Bayesian 95% credible interval can be interpreted in the way that frequentist confidence intervals are often misinterpreted: we can be 95% sure that the true value lies between these lower and upper bounds, given our prior beliefs and the current data.

In summary, the evidence here was not sufficient to reject the null hypothesis in a NHST ANOVA design. However, when I directly contrasted trials where the base rate agreed with the description to those where they disagreed, the data supported an extremely small conflict effect with considerable uncertainty, rather than the null hypothesis of no effect whatsoever. Ultimately, the present data are inconclusive, and fail to provide strong evidence either of a slow-down under conflict, or of the absence of such a slow-down.

5.3.3 CURSOR TRAJECTORIES

As described in Chapter 2, cursor transition probabilities were calculated for each condition. When base rates and descriptions disagreed, participants initially moved towards the option cued by the description on 66% of trials, selected this option after initially moving towards it 89% of the time, and selected the base rate-cued option after initially moving towards it 36% of the time.

EFFECT OF DESCRIPTIONS ON BASE RATE CHOICES

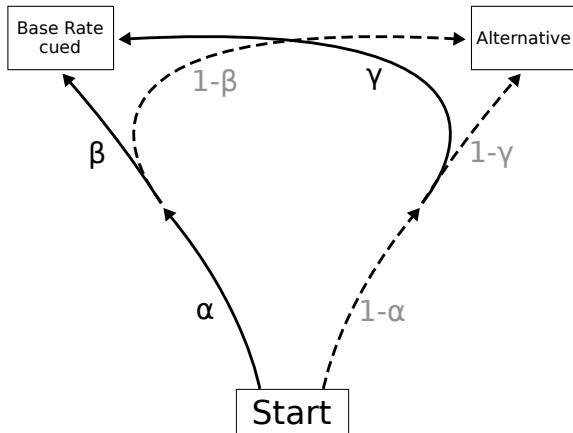


Figure 5.5: The possible transitions that can occur during a trial where the base rate cues a response.

Description	α	β	γ
Agreed	69.4%	96.5%	89.5%
Uninformative	57.4%	79.1%	56.3%
Disagreed	33.8%	36.1%	11.2%

Table 5.6: Transition probabilities for trials where the description either agreed with the base rate, was uninformative, or disagreed with the base rate. There were significant main effects of description type, and significant pairwise differences between each description, on all measures.

There was a significant main effect of the description type on the proportion, α of trials where initially moved towards the base rate-cued option ($\chi^2 = 135.9$, DF = 2, $p < .0001$). Participants were most likely to initially move towards this

option when the description also cued it (69.4%), less likely when the description was uninformative (57.4%) and least likely when the description cued the alternative option (33.9%). Pairwise comparisons between each description type were all statistically significant (z 's > 3.9 , p 's $< .001$).

β indicates the proportion of trials where participants selected the base rate-cued option after initially moving towards it. There was a significant main effect of description type on this measure ($\chi^2 = 262.7$, DF = 2, $p < .0001$). Participants were most likely to select the base rate-cued option after initially moving towards it when the description also cued that option (96.5% of trials), less likely when the description was uninformative (79.1%) and least likely when the description cued the alternative option (36.1%). Pairwise comparisons between each description type were again all statistically significant (z 's > 5 , p 's $< .0001$).

Lastly, γ reflects the proportion of trials where participants selected the base rate-cued option after initially moving towards the alternative option. There was a significant main effect of description type ($\chi^2 = 301.4$, DF = 2, $p < .0001$). Participants were most likely to change direction and select the base rate-cued option after initially moving towards the alternative when the description agreed with the base rate (89.5% of trials), less likely when the description was uninformative (56.3%), and were very unlikely to change direction when that alternative was cued by the description (11.2%). Pairwise comparisons between each description type were once again all statistically significant (z 's > 5 , p 's $< .0001$).

These results show that description interfered with participants' base rate-based reasoning at every juncture, influencing both their initial mouse movements, and their subsequent responses, regardless of their initial movement.

EFFECT OF BASE RATES ON DESCRIPTION CHOICES

The transition probability analysis was repeated to explore the effect of manipulating the base rate on participants' movements towards the description-cued option. Transition probabilities, broken down according to the base rate, are shown in Table 5.7. Here, α represents the proportion of trials where participants initially moved towards the description-cued option. There was no main effect of base rate on this proportion ($\chi^2 = 1.5$, DF = 2, $p > .4$): participants initially moved to-

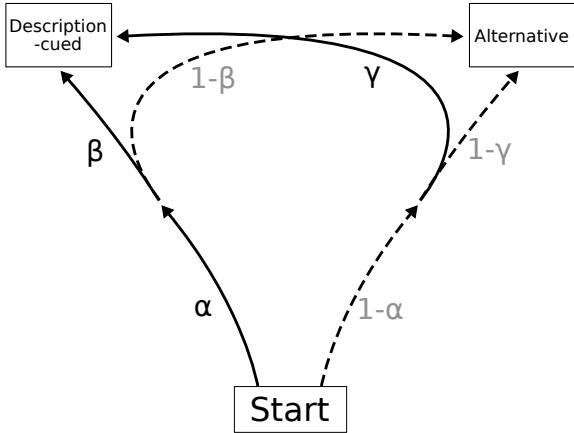


Figure 5.6: The possible transitions that can occur during a trial where the description cues a response.

Base rate	α	β	γ
Agreed	69.4%	96.5%	89.5%
Uninformative	68.5%	98.0%	84.1%
Disagreed	66.1%	88.8%	63.7%

Table 5.7: Transition probabilities for trials where the base rate either agreed with the description, was uninformative, or disagreed with the description. There was no effect of base rate type on α . β and γ were both significantly lower (that is, participants were less likely to select the description-cued option, regardless of initial movement direction) when the base rate disagreed with the description.

wards the description-cued option on average 68% of the time, regardless of the base rate information.

There were, however, main effects of base rate type on β , the proportion of trials where the description-cued option was selected after participants initially moved towards it, and γ , the proportion of trials where it was selected after participants initially moved towards the alternative option (χ^2 's > 30, DF = 2, p's < .0001). In both cases, participants were less likely to ultimately select the description-cued option when the base rate cued the other response option than when the base rate was uninformative or when it agreed with the description (z 's > 3, p's < .001; see Table 5.7).

Therefore, base rates had no early impact on participants' description-based reasoning, and did not influence their initial mouse movements. They showed some effect, however, on ultimate responses, as participants were less likely to select the description-cued option, regardless of what direction they initially moved the cursor.

5.3.4 TIME COURSE

There are many time course analyses that could be presented for this experiment, and so for brevity, I report only the most informative comparisons here. To explore the effect of varying the description on participants' base rate-based reasoning, I examined the proportion of trials on the side of the screen containing the base rate-cued option, on trials containing such an option. Likewise, to explore the effect of varying the base rate on description-based reasoning, I examined the proportion of trials on the side of the screen containing the description-cued option, on trials such an option existed. Both comparisons are plotted in Figure 5.7.

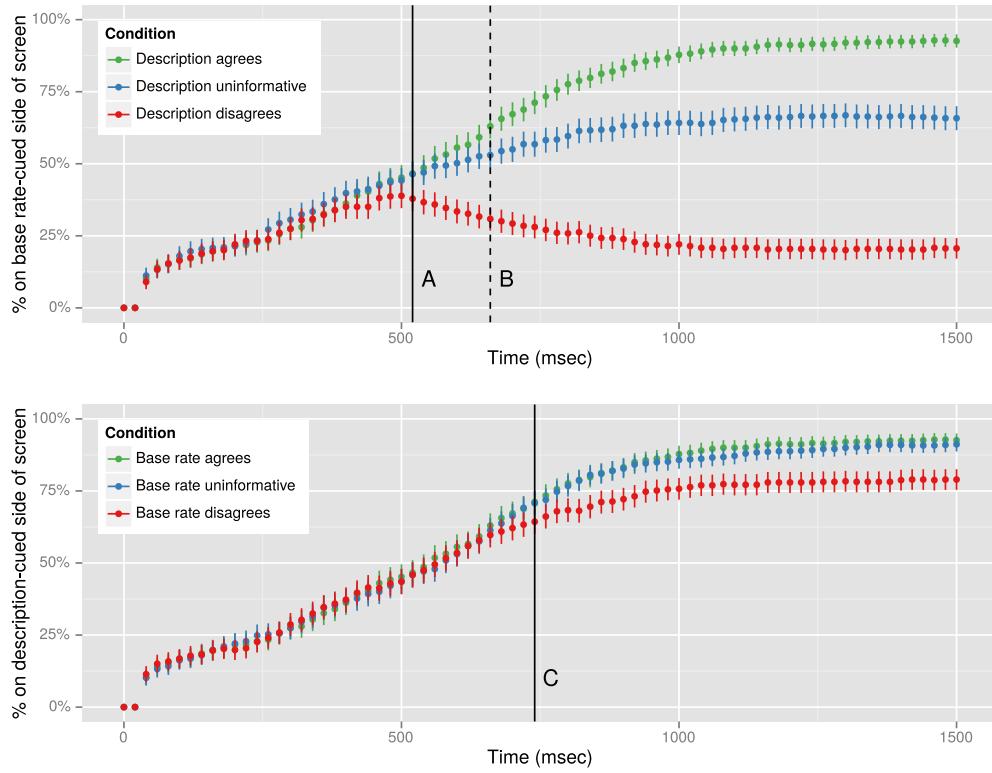


Figure 5.7: Top: Proportion of trials where the cursor is on the side of the screen containing the base rate-cued option, over time, for trials where the description agreed with the description, is uninformative, or disagreed with it. Line A shows the onset of a significant difference between when the description disagreed with the base rate and the other two conditions. Line B shows the onset of a difference between when the description agreed and when it is uninformative. Bottom: Proportion on the side containing the description-cued option, for trials where the base rate agreed with the description, is uninformative, or disagreed with it. Line C shows the onset of a difference between when the base rate disagreed with the description and the other two conditions.

Divergence points were found as before, by fitting two series of logistic mixed models. The first predicted the proportion of trials on the base rate-cued option's

side of the screen, with description type (agreed with base rate, uninformative, or disagreed) as a predictor. The second predicted the proportion of trials on the description cued option's side of the screen, with base rate type (agreed with description, uninformative, or disagreed) as a predictor. All models included random intercepts for each participant. In each series, a model was fit to each 20 msec time window, and divergence points for each pairwise comparison were defined as the times after which these comparisons were found to be statistically significant.

In both plots shown in Figure 5.7, participants were equally likely to move towards each option in any condition over the first 500 msec of each trial, indicating that neither descriptions nor base rates influenced participants' cursor movements before this time.

The effect of descriptions on base rate-driven reasoning was the first to manifest: participants were significantly less likely to place the cursor on the side of the base rate-cued option when the description disagreed with the base rate than when it agreed, or was uninformative, from 520 msec onwards (line A in Figure 5.7). Later, they were more likely to move the cursor to this side of the screen when the description agreed with the base rate than when it was uninformative from 660 msec (line B in Figure 5.7).

The effect of the base rates on description-driven reasoning was slowest to arise. Participants became less likely to be on the side of the screen containing the description-cued option when the base rate disagreed with the description than when it was agreed or was uninformative from 740 msec onwards (line C in Figure 5.7). Mirroring the analysis of responses on these trials, above, no difference emerged between trials where the base rate agreed with the description and trials where the base rate was uninformative.

5.4 DISCUSSION

In this experiment, I combined mouse tracking with a forced choice version of the base rate neglect task, where I manipulated both the descriptions, and the base rates participants reasoned about. In doing so, I build on two lines of research. Much earlier work in the heuristics and biases and dual process literature (i.e. Kahneman, 2011; Kahneman & Tversky, 1973) claims that descriptions are pro-

cessed easily and automatically by Type 1 processes, while processing base rate information requires the optional later engagement of more effortful Type 2 processes. As a result, base rates are more likely to be ignored or underweighted. More recent accounts (i.e. Barbey & Sloman, 2007; De Neys, 2012) mostly agree with this fundamental dual process interpretation, but argue that base rates play a more involved role in reasoning, either because they can be processed by Type 1 processes (De Neys, 2012), or because Type 1 and 2 processes operate in parallel (Barbey & Sloman, 2007; Sloman, 1996).

Analysis of problems in this experiment where base rates and descriptions conflicted provided support for a generic dual process account. Participants gave the response cued by the description on 80% of such trials, and when doing so were faster to read the description, faster to respond, and less likely to initially move the mouse cursor towards the opposite option than on trials where they gave the base rate-cued response. Participants were also much more likely to initially move the mouse cursor towards the description-cued option than the base rate-cued option when the cues disagreed (doing so 66% of the time), and more likely to change direction if their initial movement was towards the base rate-cued option.

Most dual processes accounts — the exception being a selective model (Chaiken, 1987; Klaczynski, 2000) where one or other kind of processing is engaged for a given task — would predict that participants who do give the response cued by the base rate should be conflicted while doing so, as they must inhibit the pull of the description-cued response. This prediction was tested by analysing all trials where participants could rely on the base rate, and looking at effect of manipulating the description.

Consistent with this prediction, the contents of a problem's description had a strong effect on participants' base rate-cued reasoning, as participants were most likely to give the base rate-cued response when this response was also supported by the description (94% of trials), less likely when the description was uninformative (69%), and as mentioned above, only selected the base rate-cued response when it conflicted with the description on 20% of trials.

Similarly, when they did give this base rate-cued response, participants both spent longer reading the description, and spent longer responding, when the de-

scription conflicted with the description than when it agreed. This suggests that participants had to inhibit the description-cued response before they could give the base rate-cued response instead. Unlike previous experiments in this thesis, however, participants' were not significantly more likely to trace reversal trajectories towards the base rate-cued option under conflict, although the trend was in this direction.

The transition probability analysis (Table 5.6) similarly showed that descriptions had a pervasive effect on participants' base rate-driven reasoning, influencing both their initial movements, and their ultimate responses regardless of initial movement direction. The transition probabilities also allow us to make sense of the non-significant effect of descriptions on the probability of participants following a reversal trajectory when giving the base rate-cued response. Regardless of the base rate, participants initially moved towards the description-cued option, where available, on two thirds of trials. Given that they were no more likely to initially move towards the alternative option when the base rate cued it than when it did not, we must assume that base rates play no role at this point. Later in reasoning, the influence of the description is even greater, dictating 80% of final responses when it conflicts with the base rate, and 94% both when it agreed and when the base rate was uninformative. At this point, base rates also clearly play a role, as the response not cued by the description is given on only 6% of trials when it is not cued by the base rate either, but 20% when it is cued by the base rate. To simplify, it appears that only the descriptions, and random variations, have an effect on initial movements, while later movements are even more influenced by descriptions, and also slightly influenced by base rates. The time course data are also in line with this interpretation, as base rates had no influence on early movements. When they did show an effect, after 740 msec, this only slightly reduced the pull of the description-cued option rather than reversing it, as was the case in Experiments 1 and 2, for instance (see the bottom panel of Figure 5.7).

While these results are consistent with most dual process accounts of this task, some accounts, such as the intuitive logic theory (De Neys, 2012; Handley & Trippas, 2015), or a parallel-competitive dual process theory (Barbey & Sloman, 2007; Sloman, 1996), further predict conflict in the opposite direction, with participants

who base their responses on descriptions nevertheless showing sensitivity to the base rate information, and experiencing conflict when base rates and descriptions disagree.

Analysis of participants' responses on trials where the description cued a response showed some evidence of sensitivity to base rates, with participants less likely to give the description-cued response when it conflicted with the base rate than when the base rate agreed with the description, or was uninformative. However, on such conflict trials participants still overwhelmingly gave the response cued by the description. Analysis of individual differences here (Figure 5.3) showed that the majority of participants' responses appeared to be slightly influenced by the base rate, with a small minority strongly influenced by it.

In contrast to previous tests of the intuitive logic theory, however, (De Neys & Glumicic, 2008; De Neys et al., 2008; Franssens & De Neys, 2009; Pennycook, Fugelsang, & Koehler, 2012) there was no significant evidence that participants were influenced by base rates on trials where their responses were driven by the description, as their reading times, movement initiation times, response times, and cursor trajectories did not differ significantly when the base rate agreed with the description, was uninformative, or disagreed with it. However, the trends were in the direction predicted by the intuitive logic account, with participants slower to respond when the base rate conflicted with the description. An exploratory Bayesian analysis suggested that the most plausible effect size was a 2.8% increase in response times under conflict, although the data was largely inconclusive. There was considerable uncertainty around this tiny effect.

Analyses of the transition probabilities and the time course data told the same story. Participants' initial mouse movements, and the position of the cursor before 740 msec, were not influenced by changes in the base rate on trials where participants could rely on the description. Instead, participants were slightly less likely to select the description-cued response option, or to be on its side of the screen from 740 msec onwards, when the base rate cued the alternative response. Taken together, these results suggest that base rates either dictated participants' response to a problem, or were almost totally ignored.

To recapitulate, the results of this experiment provide further support for a dual

process interpretation of base rate neglect (Kahneman & Frederick, 2002, 2005), where, fast, effortless, automatic Type 1 processes underlie description-based reasoning, and slower, effortful Type 2 processes underlie base rate-based reasoning. Results consistent with this interpretation were found in all of the analyses reported. Participants predominantly gave the description-cued response when the base rate also cued it and when the base rate was uninformative, and only did so slightly less when the base rate conflicted with the base rate. Even when giving base rate-cued responses, participants were conflicted when the description disagreed with the base rate. Analysis of the cursor trajectories and the time course data showed that descriptions dictated both early movements (from half a second) and participants' actual responses.

The data were less consistent, however, with some intuitive logic accounts (i.e. De Neys, 2012, 2014; Handley & Trippas, 2015). These would predict that even when participants give the response cued by the description, they should be sensitive to manipulations of the base rate, and previous studies have found such effects (i.e. De Neys & Glumicic, 2008; Pennycook, Fugelsang, & Koehler, 2012; Pennycook et al., 2014). Here, I found no significant effects of manipulating the base rate on participants' cursor movements or response times when they gave the description-cued response. An exploratory Bayesian analysis, however, showed that participants were very slightly slower (around 30 msec) on these trials when the base rate disagreed with the description. Thus, it is not possible to draw strong conclusions either for or against the intuitive logic account from the current data.

The absence of a significant intuitive logic effect on response times in particular was surprising, as response times have been used as a measure of this conflict in a number of previous studies (i.e. De Neys & Glumicic, 2008; Pennycook, Fugelsang, & Koehler, 2012), outlined in Table 5.8. Although it is difficult to say with certainty why this experiment differs from previous work, I can identify a number of possibilities.

First, a number of changes were made to the procedure used by De Neys et al. (2008) in order for it to be compatible with the mouse tracking paradigm. As discussed above, the way in which information was presented here meant that participants were able to process each trials' information at two points: before clicking

on the “NEXT” button to reveal the question (reading time), and after seeing the question but before responding (response time). However, even when I combined these times (not reported), there was little evidence of participants being slower to give the description-cued response when it conflicted with the base rate.

The mouse tracking paradigm also requires that participants respond under time pressure, in this case within 6 seconds. A visual timer in the centre of the screen was used in this experiment to reinforce this idea, filling up over the course of the allowed time. As this time limit was only exceeded on 2 trials out of 2,000, and average response times were below 2 seconds in all conditions, I can be quite confident that participants did reason quickly, although again this does not include time spent reading the description.

These response are considerably faster than those reported in the majority of previous conflict detection studies that used the base rate paradigm and measured response latencies, outlined in Table 5.8. Moreover, of these studies, only two required participants to respond before a deadline. De Neys and Glumicic (2008) presented participants in an fMRI scanner with problems almost identical to those used here, and required them to respond within 8.5 seconds of seeing the questions. They found that participants were slower to give description-cued responses when the cues conflicted (3.5 seconds) than any other condition (~ 2.8 seconds). However, it should be noted that these response times were almost a second longer than those reported here, at a time scale where this constitutes a $\sim 66\%$ increase. Their result, while moderate in size ($\eta_p^2 = .28$), was also not robustly statistically significant, with $F(1, 12) = 4.6, p_{rep} = .87$ (corresponding to approximately $p = .05$).

The other study reporting conflict effects at such a short time scale was Experiment 2 of Pennycook et al. (2014), where participants were asked to respond within 5 seconds. Note, however, that this was an unusual base rate experiment. Participants were instructed before each trial to base their response on either “belief”, or “statistics”, and in Experiment 2 asked to respond within 5 seconds. Analysis of response latencies revealed conflict effects in both directions — when reasoning based on statistics, participants were slower if the description cued the opposite response, and likewise slower when reasoning based on belief if the base rate

cued the opposite response. However, these times remain a second or more slower than those reported in the current experiment. Furthermore, the magnitude of this effect was extremely small — going from 3.70 to 3.79 for statistics-based decisions — a small effect size (η_p^2) of .08. It should also be noted that in Experiment 2 of Pennycook et al. (2014), instructions to rely on belief or statistics were manipulated within-participants, and so the effect within this short time window could be in part due to task-switching effects (Monsell, 2003). Experiment 3 of the same paper demonstrated a similar effect with a between-participants manipulation, but participants were not asked to respond quickly, and the dependent variables were participants' probability judgements and confidence ratings, not their response times. Therefore, it seems possible that in the current experiment a) participants responded too quickly in most cases to detect the conflict between their responses and the base rate; b) having not been explicitly told to use the base rate on 50% of trials, participants may have been less sensitive to the base rate than those in Pennycook et al. (2014).

Another factor may be the scarcity of base rate-cued responses in general in the current experiment. Reasonably, one would expect conflict detection to be related to participants' responses: experiments that yield many base rate-cued responses should also yield greater detection of conflict on problems in which the description-cued response is given. In this experiment, on the other hand, the base rate-cued response was given on only 20% of conflict problems, perhaps unsurprisingly given the fast response times here, and the robust finding that base rate responses are slower than description responses. Consequently, it may not be so surprising that an experiment which yielded relatively few base rate-cued responses should also show little influence of base rates on more subtle measures such as response time.

Lastly, recent work (e.g. De Neys et al., 2010; Mevel et al., 2014) has highlighted that there are individual differences in this conflict detection process — not all participants are slower, or less confident, when their responses conflict with base rates, or with logical principles. At present, we know relatively little about the factors that make some participants, but not others, sensitive to these conflicts. Therefore, given that almost all studies in this area reveal some participants for whom the conflict detection effects do not hold, it should perhaps not be surpris-

Table 5.8: Number of base rate-cued responses under conflict, and response latencies for description-cued responses when cues either agreed or conflicted, in previous base rate studies.

Study	Procedure	BR responses	No-conflict RT	Conflict RT
De Neys et al. (2008)	fMRI	45%	2.8	3.75
De Neys and Glumicic (2008)	Standard	22%	~14	~18
Franssens and De Neys (2009)	Secondary load	47% (no load) 35% (load)	~14	~17
De Neys and Franssens (2009; Exps. 2–4)	Standard	~33%	~14	~17
Pennycook, Fugelsang, and Koehler (2012; Exp. 4; extreme base rates)	Standard	24–26%	16	20
Pennycook and Thompson (2012; Initial responses)	Two-response paradigm (probability ratings)	N/A	12.8	13.4
Pennycook et al. (2014; Exp. 1)	Belief/statistics instructions	N/A	~13	~14
Pennycook et al. (2014; Exp. 2)	Belief/statistics instructions, 5 second deadline	N/A	~3.8	~3.7

ing that these effects are not found in every study.

CONCLUSIONS

To conclude, this chapter reports one of the first uses of the mouse tracking paradigm to investigate the interaction of dual processes during reasoning. Results are consistent with a default-interventionist dual process accounts (Evans, 2006; Evans & Stanovich, 2013; Kahneman & Frederick, 2002), by which descriptions are processed easily and automatically by Type 1 processes, but base rates, thought to require Type 2 processes to process, play less of a role in reasoning on most trials. In fact, when the base rates were attended to, they tended to dictate participants' responses outright. The rich, temporal dynamics of the data collected using this paradigm reveal much more about the underlying processes, for instance that only descriptions influence the initial direction of participants' cursor movements, and that the descriptions have a discernible effect on cursor movements from around 500 msec, while the weaker effect of base rates is not apparent until around 750 msec.

A number of recent studies have also shown evidence of conflict in the opposite direction, as base rates have some effect on participants even when their responses appear to be dictated by descriptions alone. The current data, however, did not support this idea, although this may in part be due to the ways in which the experimental paradigm had to be adapted to suit the mouse tracking paradigm, in particular, the extremely faster response times, and unusual presentation of information. Despite this, however, the current chapter demonstrates another point at which conflict can be found in reasoning, when the right kind of data is collected.

CHAPTER 6

Intuition vs. Reflection in Reasoning

6.1 INTRODUCTION

In Experiment 5 (Chapter 5), I used the mouse tracking paradigm to study the interaction of two kinds of processes in cognition: fast, intuitive *Type 1 processes*, and slower, deliberative *Type 2 processes*. In that experiment, participants were asked to respond within 5 seconds of seeing the probe question, and my analyses focused on the responses they gave, their response latencies, and whether they moved the mouse straight to their response option or went the wrong way initially. However, reasoning is one of the most complex, high-level cognitive processes possible, and naturally not all reasoning happens in less than 5 seconds. In this chapter, I use a slightly different kind of mouse tracking paradigm to investigate conflict in reasoning over a considerably longer time scale. I do this for a task that has become perhaps the classic means of pitting intuition against logic: the Cognitive Reflection Test (Frederick, 2005). Like Chapter 5, this chapter addresses the question of just how Type 1 and Type 2 processes interact in dual process accounts of cognition.

6.1.1 COGNITIVE REFLECTION AND REASONING

The Cognitive Reflection Test (CRT; Frederick, 2005) is a brief test designed to measure individuals' ability to inhibit intuitive responses in favour of reflective and deliberative reasoning. In the bat-and-ball problem, one of the best-known CRT items, participants are asked:

“A bat and a ball together cost £1.10.

A bat costs £1 more than a ball.

How much does a ball cost?”

The appealing but incorrect response, to say “10p”, is believed to be generated effortlessly and automatically. Arriving at the correct response of “5p” may require that this intuitive response is inhibited in favour of the result of sustained, effortful deliberation.

The CRT has become a widely-used measure of individual differences in cognition. Higher CRT scores predict better performance on various cognitive tasks, including reduced framing effects, less discounting of delayed rewards (Cokely & Kelley, 2009; Frederick, 2005) and probability matching (Koehler & James, 2009), resistance to the illusion of explanatory depth (Fernbach, Rogers, Fox, & Sloman, 2013) and conjunction fallacies (Oechssler, Roider, & Schmitz, 2009), greater metacognitive awareness (Mata, Fiedler, Ferreira, & Almeida, 2013), and less endorsement of supernatural beliefs (Pennycook, Cheyne, Seli, Koehler, & Fugelsang, 2012; Shenhav, Rand, & Greene, 2012), less endorsement of vacuous statements as profound (Pennycook, Cheyne, Barr, Koehler, & Fugelsang, 2015), as well as performance on various tasks that pit normative responding against intuition (Toplak, West, & Stanovich, 2011). Scores on the CRT correlate with measures of IQ and personality characteristics, and usually predict performance even when these are controlled for (Toplak et al., 2011).

6.1.2 DUAL PROCESS ACCOUNTS

The CRT is widely seen as a an archetypal application of dual process theories of cognition (De Neys et al., 2013; Frederick, 2005; Kahneman & Frederick, 2005). Consistent with this, performance on the CRT is related to performance on a number of other traditional dual process tasks (Toplak et al., 2011), and to dispositional factors related to willingness to engage in analytic thinking (Böckenholt, 2012; Campitelli & Gerrans, 2013; Campitelli & Labollita, 2010).

However, dual process theories differ in their account of CRT performance. As discussed in Chapter 1, there are a number of ways in which Type 1 and Type 2 processes can interact during reasoning. As has been my focus throughout this

thesis, I will concentrate on the predictions made by each account with regard to when conflict occurs during reasoning. Although I am interested in reasoning in general, I discuss these accounts in terms of the specific points at which they predict conflict should occur during the CRT. According to a selective dual process account (e.g. Klaczynski & Cottrell, 2004), the CRT, like all such tasks, should not evoke any conflict. Instead, participants should selectively draw on heuristic Type 1 processes, which usually produce a response of “10p” on the bat-and-ball problem, or draw on Type 2 processes to apply the simple mathematical rules needed to reach a correct response of “5p”. On the other hand, Type 1 processes are always automatically activated in default-interventionist models (Evans, 2006; Kahneman, 2011; Kahneman & Frederick, 2005). Therefore, such accounts would predict that heuristic responses (“10p”) come to mind automatically for most participants, and that giving the correct response requires that this heuristic response is inhibited, and that Type 2 processes are engaged to derive the correct response. Such a default-interventionist account is assumed, for instance, in Frederick’s (2005) paper introducing the CRT, and Toplak et al.’s (2011) work outlining the relationship between the task and other dual process problems.

A third option (Sloman, 1996, 2014b) is that both Type 1 and Type 2 processes are activated simultaneously, and that they compete for control of behaviour. In common with default-interventionist models, these accounts predict that Type 1 intuitive responses must be inhibited in order to reason correctly. Uniquely though, parallel models would also predict Type 2 processes should attempt to signal the correct response, even when failing to overrule the output of Type 1 processes.

The intuitive logic model (De Neys, 2012, 2014) has also been applied to performance on the CRT (De Neys et al., 2013). This model modifies the traditional default-interventionist model to account for many findings which indicate that when participants provide biased, heuristic responses, they are often implicitly aware of some conflict between their responses and the normative standard. According to this model, Type 1 processes are sensitive to normative principles, such as logical principles in syllogistic reasoning tasks, or mathematical rules on the bat-and-ball problem. As a result, they implicitly signal a conflict when the incorrect heuristic response is given. However, because the heuristic response is usu-

ally prepotent, participants often fail to inhibit it, even when they do detect that it conflicts with normative principles. It is unclear at present, however, how this conflict is actually detected. One possibility is that Type 1 processes simultaneously produce both heuristic and correct responses, and it is the conflict between these two partially active beliefs which is detected directly. In Chapter 1, it is this proposition that I refer to as the intuitive logic theory. Alternatively, the process may be more subtle, with Type 1 processes not generating a fully-formed correct response, but rather detecting, through some other means, that the heuristic response is questionable. I refer to this proposition in Chapter 1 as the dual process conflict monitoring theory. Clearly, these two possibilities make different predictions about conflict between competing response options during reasoning. In the former case, the intuitive logic model would, like a parallel-competitive account, predict that because both responses are partially cued, participants should be drawn towards giving the correct response, even when they ultimately give the heuristic one. In the latter case, if Type 1 processes can signal conflict without actually generating the correct response, participants may experience conflict and uncertainty, but not be actually drawn towards the correct response when giving the heuristic one. For the purposes of this experiment, the same predictions are made by the slightly different intuitive logic account offered by Handley and Trippas (2015). For simplicity, I will refer to De Neys' (2012) account throughout, with the understanding that the same points generally apply to other intuitive logic accounts.

As discussed in previous chapters, evidence of the implicit conflict detection predicted by the intuitive logic model comes from a range of experimental paradigms (see also De Neys, 2012, for a review). Typically, these studies compare conflict problems, where the intuitive, heuristic response is incorrect, to analogous no-conflict versions, where both heuristics and normative principles cue the same response. Type 1 processes cue both the heuristic response on conflict problems and the correct response on no-conflict problems. If participants detect the conflict between normative principles and their heuristic responses, they should show greater evidence of conflict on these problems, compared to the no-conflict problems. Such conflict has been measured using confidence ratings (De Neys et al.,

2011), response times (De Neys & Glumicic, 2008), neuroimaging (De Neys et al., 2008), and galvanic skin response (De Neys et al., 2010), among other measures.

Two studies, however, directly test the intuitive logic model against the CRT. De Neys et al. (2013) showed that heuristic responses on conflict problems were given with less confidence than correct responses on no-conflict versions of the same problems. Gangemi, Bourgeois-Gironde, and Mancini (2015) report similar effects, asking participants to fill out a brief questionnaire measuring their “feeling of error” after answering either the original bat-and-ball problem or a no-conflict control version, both when participants were asked to generate their responses, and when asked to choose between the heuristic and correct responses. These findings all suggest that participants are to some extent aware of the inadequacy of their heuristic responses.

In this chapter I use a novel version of the mouse tracking paradigm to explore this topic. Participants completed a computer-based multiple-choice version of the CRT, including both conflict problems, where the intuitively appealing heuristic response was incorrect, and no-conflict versions of the same problems, where the appealing response was the correct one. There were four response options for each problem, located in each corner of the display. While participants decided on their response, their mouse cursor movements were recorded as they moved the cursor around the screen. Unlike in previous mouse tracking studies, participants were not placed under time pressure, and their mouse movements over the first 60 seconds of each trial were analysed. Rather than moving the cursor directly from its starting point to a response option and clicking on it, participants typically moved the cursor around the screen a number of times on each trial, passing close to multiple response options. As mentioned in Chapter 1, mouse movements over such long time scales have been used in the past as part of research on human-computer interaction, typically involving users interacting with web pages, such as web search results pages (e.g. M. C. Chen et al., 2001; Huang et al., 2011; Rodden et al., 2008). This allowed me to use the patterns in these cursor movements to test predictions derived from the various forms of dual process theory.

According to a selective dual process account, participants should move to the response option they eventually chose without showing any particular attraction

towards any of the other options along the way. According to a traditional default-interventionist account, on conflict trials participants should initially be drawn towards the heuristic response option, but in some cases inhibit this response and instead select the correct option. A parallel-competitive account would likewise predict that participants are drawn towards the heuristic option on trials where they eventually give the correct response, but would also predict the reverse — because Type 2 processes are activated in parallel with Type 1, participants should be in some cases drawn towards the correct option even on trials where they end up giving the heuristic response. The predictions of the intuitive logic model depend on the nature of the conflict detection process. If participants detect conflict because both responses are simultaneously generated by Type 1 processes, then the intuitive logic model, like the parallel-competitive model, would predict conflict in both directions. Alternatively, if the conflict detection process is more subtle (c.f. the dual process conflict monitoring theory, Chapter 1) then like the classic default-interventionist account it would predict that participants should be drawn to the heuristic option when selecting the correct one, but not the other way around.

6.2 METHOD

6.2.1 PARTICIPANTS

One hundred and thirty one students at Queen’s University Belfast participated in exchange for course credit.

6.2.2 MATERIALS

Eight problems were adapted from Primi, Morsanyi, Chiesi, Donati, and Hamilton’s (2015) extended version of the CRT. Each of these problems was modified to create a set of eight corresponding non-conflict problems, in which the intuitively appealing responses were also the correct ones (see the Appendix M). Participants were randomly allocated to complete either conflict versions of items 1, 3, 5, and 7 and no-conflict versions of the rest, or vice versa. Each problem was presented in a 4-option multiple choice format. For the conflict items, the possible responses

were the correct option, the incorrect heuristic option, and two incorrect foil options. For the non-conflict items the correct intuitive option was presented with three incorrect foils.

6.2.3 PROCEDURE

The experiment was administered on personal computers, programmed using PsychScript (see Chapter 2), and run in the web browser. Participants were instructed to respond in their own time to each CRT problem by clicking on one of the four response options presented in the four corners of the display (Figure 6.1). Participants were not made aware of the mouse tracking aspect of the experiment in advance.

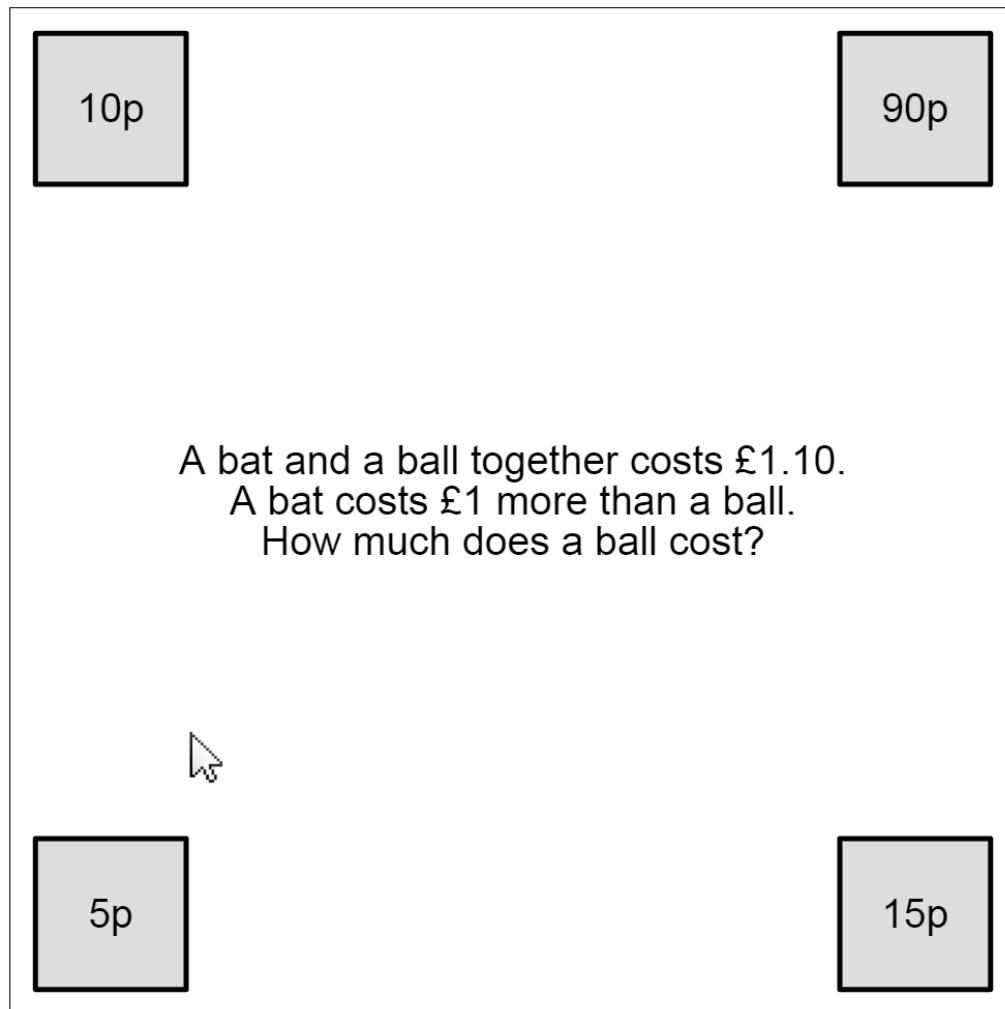


Figure 6.1: A screen shot from the CRT.

Each problem was preceded by onscreen instructions to click on a button marked “Go”, presented in the centre of the monitor. This was done to ensure the mouse cursor was located in the same central position at the beginning of each trial. The button was then replaced by the problem text and the four response options appeared simultaneously in the corners (Figure 6.1). The response options were randomly assigned to the four locations on each trial, with the constraint that the correct and heuristic response options were always adjacent for conflict problems. The mouse cursor was no longer visible at the onset of each trial to prevent it from obscuring the question text. The cursor reappeared once it had been moved more than 5% of the width of the display. Mouse cursor location was recorded every 20 msec.

6.3 RESULTS

6.3.1 BY-TRIAL ANALYSES

After excluding data from 3 participants who did not complete the experiment within the 15 minutes allocated, and 7 trials with response times greater than 100 seconds (.6% of the total), participants selected the correct option on 79.5% of non-conflict problems. On the conflict problems, the correct option was chosen 36% of the time, the heuristic option 58%, and one of the foils 6% of the time.

In the first stage of the analysis, I calculate a number of summary statistics for each trial, and compare these between problem types, and between responses. The measures were response time, the distance travelled by the mouse cursor (scaled so that a straight line from the start point to the response corresponds to 1 unit), the number of times the cursor was moved during a trial (with movements defined as windows of 100 msec or more in motion, separated by 100 msec or more not moving), the closest proximity achieved between the cursor and the non-chosen option (closest proximity to the heuristic response option on trials where the correct option was chosen, and vice versa). These measures were compared using linear mixed models, with crossed random intercepts for each participant, and each problem (Baayen et al., 2008). Response latencies, and the distance travelled by the mouse cursor were log-transformed to normalise their distributions, and a gener-

alised mixed model with a Poisson link was used to model the number of movements.

Consistent with a dual process interpretation, whereby heuristic responses are generated by Type 1 processes, and correct responses under conflict by Type 2 processes, for conflict problems there was greater evidence of conflict across all measures when participants gave the correct response ($N = 181$) than the heuristic one ($N = 297$). The average time to respond was 27.3 seconds ($SD = 16.3$) for correct responses, and 21.0 seconds ($SD = 13.4$) for heuristic responses ($e^\beta = 114\%$, $CI = [102\%, 129\%]$, $t(470.8) = 2.349$, $p = .0192$). The mouse cursor travelled a greater distance before selecting a correct option (6.11 times the minimum needed distance, $SD = 5.6$) than an heuristic option (5.66 times, $SD = 4.74$; $e^\beta = 116\%$, $CI = [102\%, 133\%]$, $t(298.4) = 2.267$, $p = .0241$). There were also more cursor movements on trials in which the correct response was given (5.4, $SD = 4.8$) than when the heuristic response was given (4.9, $SD = 4.5$; $e^\beta = 1.15$, $CI = [1.02, 1.29]$, $z = 2.337$, $p = .0195$). Finally, the minimum distance between the cursor and the heuristic option on trials in which the correct option was chosen was on average 49% of the display width ($SD = 24\%$), significantly less than the minimum distance between the cursor and the correct option on trials in which the intuitive option was chosen (55.5%, $SD = 18\%$, $e^\beta = 0.92$, $CI = [0.89, 0.96]$, $t(72.1) = 4.119$, $p < .0001$).

Most tests of the intuitive logic model compare correct responses on no-conflict problems with heuristic responses on conflict problems, on the basis that heuristic, Type 1 processes should cue both kinds of response, but the chosen response conflicts with normative principles on conflict problems only. Evidence for the intuitive logic model therefore comes from results which indicate greater conflict for heuristic responses to conflict problems ($N = 404$). However, there was no such reduction in conflict for no-conflict problems on any of the applicable measures: response time (23.1 seconds, $SD = 15.3$; $t(14.3) = 0.222$, $p > .8$), distance travelled (5.6, $SD = 5.0$; $t(15.0) = 0.359$, $p > .7$) and number of movements per trial (5.2, $SD = 4.6$; $z = 0.064$, $p > .95$).

Following previous intuitive logic studies (e.g. De Neys et al., 2011; Mevel et al., 2014, see also Pennycook, Fugelsang, & Koehler, 2015), I also calculated the num-

ber of heuristic responses given by each participant on conflict problems, and categorised each participant as either “majority heuristic” (3 or 4 heuristic responses out of four, 53 participants) or “minority heuristic” (0 to 2 heuristic responses, 75 participants). I entered this measure as a participant-level predictor in the models, but found that it was not involved with any interactions in the analyses above (t 's $< .9$, p 's $> .4$). I also repeated this analyses for the most (4 heuristic responses) and least (one heuristic response) biased reasoners only, again finding so significant interactions (t 's < 1.1 , p 's $> .25$). Therefore, these analyses revealed no evidence for logical intuitions in either biased or unbiased participants.

6.3.2 TIME COURSE ANALYSES

In most previous mouse tracking research, both in this thesis and in general (e.g. Freeman, Dale, & Farmer, 2011; Spivey et al., 2005), the location of the cursor is recorded over a few seconds, as participants move it from a starting position to a response option located in either top corner of the screen. Almost invariably, this path follows either a single movement (curved or otherwise) to a response, or as has been more the case in this thesis, a movement towards one response, that changes direction mid-flight. In the current data, unfolding over up to 60 seconds, participants move and rest the cursor many times throughout a trial, an average of 5.1 times, and a maximum of 30. Thus, this data is in ways more similar to eye movement data. A typical mouse cursor trajectory is shown in Figure 6.2, showing a number of movements which pass near to several response options. In order to analyse participants' attraction to each response option over time, the display was divided into quadrants corresponding to each response option. For the first 60 seconds of each trial, the mouse cursor positions at each 200 millisecond time slice were coded according to which section of the screen they occupied, similar to fixation analyses of eye-tracking data.

Figure 6.3 shows, for each response region, the proportion of trials in which the cursor is in that region, over time, for both conflict and no-conflict problems. While the proportions at 60 seconds here largely reflect participants' ultimate responses, earlier proportions show how these preferences developed over time. Both correct responses to no-conflict problems and heuristic responses to conflict prob-

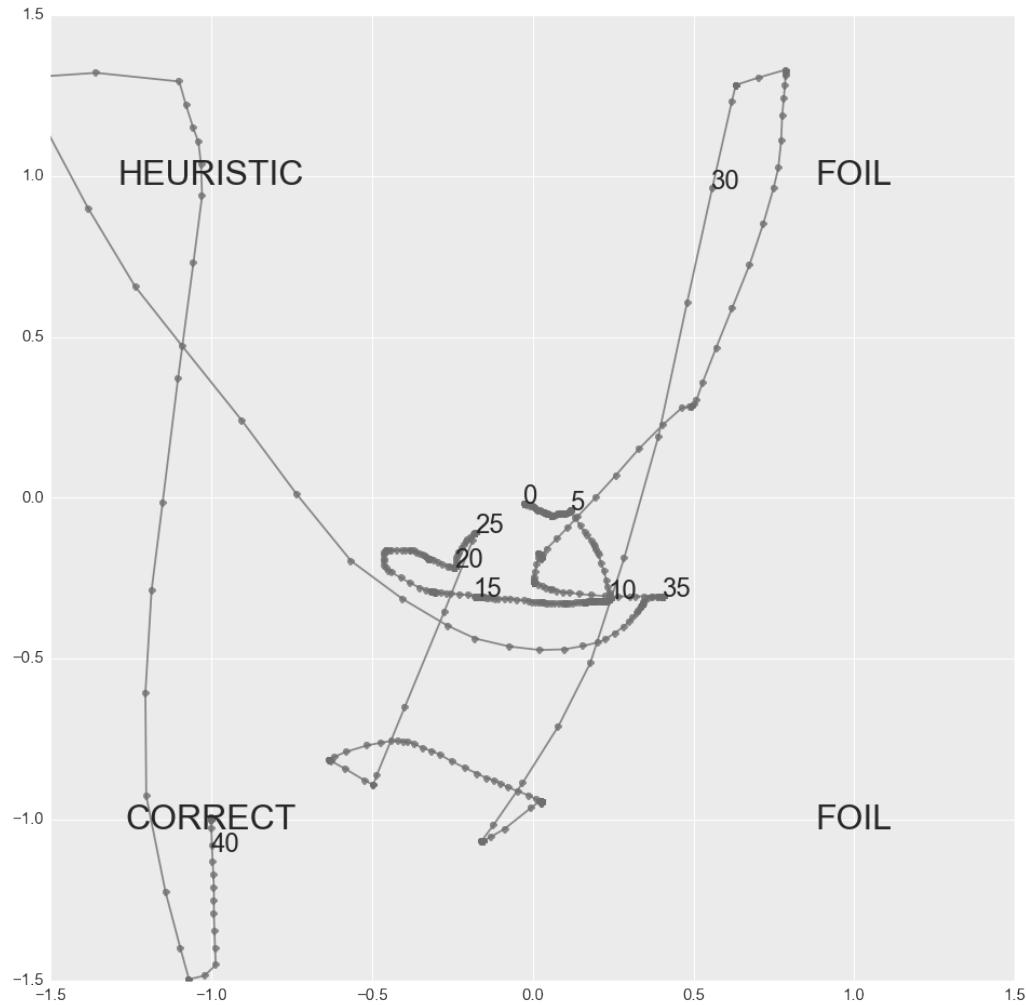


Figure 6.2: A typical mouse cursor trajectory from the conflict condition. Numerical values indicate the time elapsed in seconds. Cursors meandered as participants generated their responses, passing near the response options located in the corners of the display

lems were intuitively appealing, and participants began to move towards both options from before 5 seconds. At approximately 10 seconds, participants also began to move towards the correct response option on conflict problems, and the accumulation of cursors in the region of the heuristic option under conflict slowed accordingly. The proportion of cursors in the region of these foil response options declined steadily in both conditions. Note that the proportions for the foil response options are averaged across the two foil options on conflict problems, and three options on no-conflict problems.

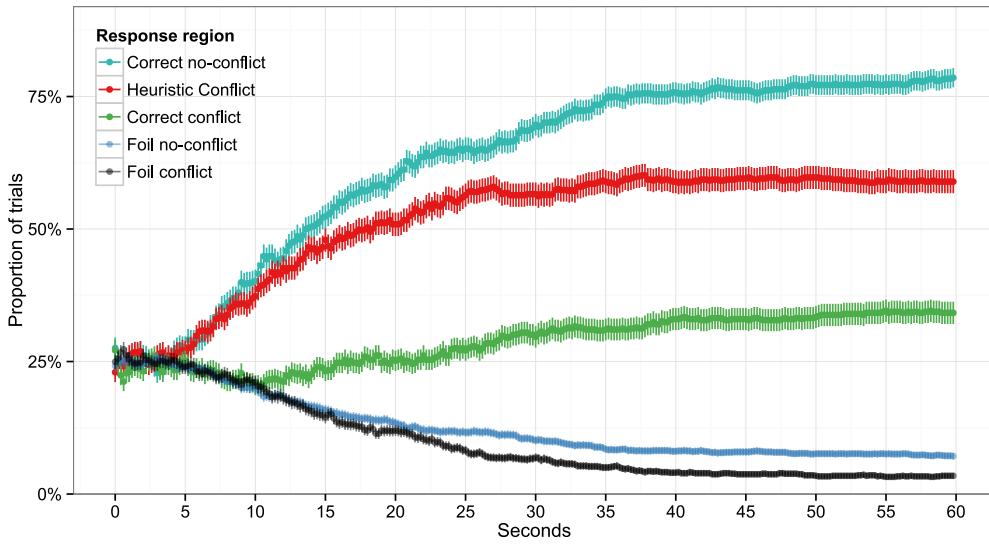


Figure 6.3: Proportion of mouse cursors in the region of the screen corresponding to each response options, over time, for conflict and no-conflict problems.

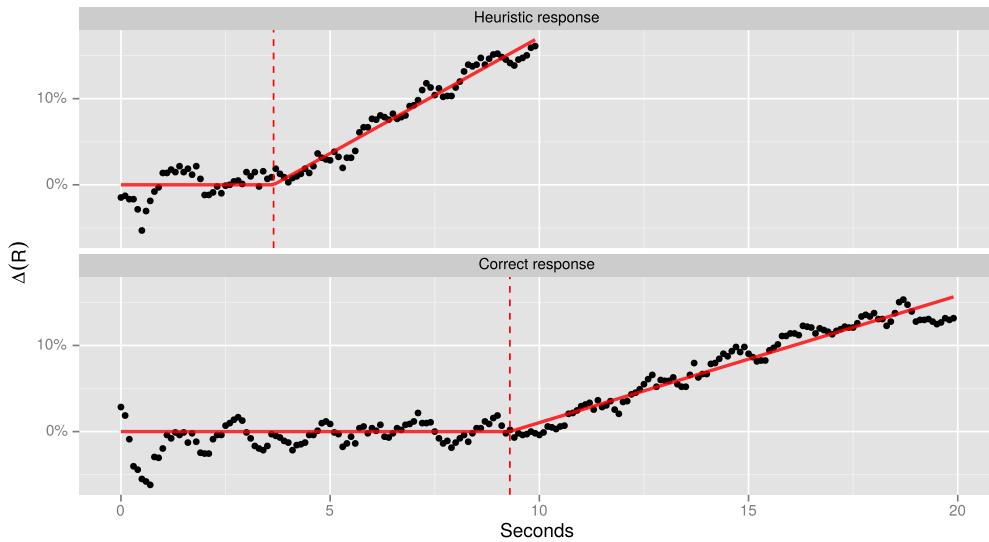


Figure 6.4: Top: $\Delta(\text{Heuristic})$, the difference between the probability of the cursor being in the region of the heuristic option and the probability of being in the region of a foil option. Bottom: $\Delta(\text{Correct})$, the difference between the probability of being in the region of the correct option, and of being in the region of a foil option. Solid red lines show non-linear regression fits. Dashed vertical lines show change points, after which participants began to be drawn towards the option in question.

BAYESIAN CHANGE POINT ANALYSIS

Inspecting Figure 6.3, participants are initially equally likely to move toward each of the four response options on conflict trials, doing so 25% of the time. This is the case until some time before 5 seconds, at which point participants become more likely to be in the region of the heuristic option than the correct option, or either of the foils, reflecting the point at which processes driving participants towards the heuristic response exert their influence. Similarly participants were equally likely to be in the region of the correct option as the foils until around 10 seconds, and after this point more likely to be in the region of the correct option, indicating that participants begin to be drawn towards the correct option from this time.

Of course, visual inspection of these curves is not a particularly accurate means of revealing *when* participants begin to be drawn towards each response option. To formally estimate the times at which participants began to move towards each response, I calculated, across each 40 msec, $\Delta(\text{Heuristic})$: the difference between the average probability of the cursor being in the region of the heuristic option and the probability of being in the region of either foil option, as well as $\Delta(\text{Correct})$: the difference between the average probability of being in the region of the correct option and of being in the region of the foil option. This yielded two series of values (Figure 6.4) that were close to 0 until participants began to be drawn towards the response in question, and increased over time after that point.

I modelled these series using a non-linear regression model of the form

$$\Delta(R) = \begin{cases} 0 & \text{if } t < \tau_R \\ \beta * (t - \tau_R) & \text{otherwise} \end{cases}$$

where t is the time in seconds, τ_R is the point at which participants begin to be drawn towards response R , and β is the slope of the regression line after time τ_R .

Modelling $\Delta(\text{Heuristic})$, I analysed the first 10 seconds of each trial, and set a uniform prior on the value of $\tau_{\text{Heuristic}}$ between 0 and 10 seconds (i.e. that participants were equally likely to start being drawn towards the heuristic response any time between 0 and 10 seconds into a trial). Modelling $\Delta(\text{Correct})$, I analysed the first 20 second, and again set a uniform prior on τ_{Correct} between 0 and 20 seconds.

Table 6.1: Posterior estimates from the change point analysis. Participants began to be drawn towards the heuristic option from 3.65 seconds, and the correct option from 9.30 seconds.

Parameter	Median	2.5%	97.5%
$\tau_{Heuristic}$	3.65	3.36	3.93
$\tau_{Correct}$	9.30	8.86	9.64
$\beta_{Heuristic}$	0.027	0.025	0.029
$\beta_{Correct}$	0.015	0.014	0.016

In both cases, I set an uninformative normal prior with mean 0 and SD 1 on the slope, β .

Figure 6.4 shows the fitted regression models, and Table 6.1 shows the posterior estimates for the parameters. The posteriors for the τ parameters represent estimates of the point at which participants began to be drawn to each response. The median posterior estimate for $\tau_{Heuristic}$ was 3.65 seconds (95% credible interval [3.36, 3.93 seconds]), and the estimate for $\tau_{Correct}$ was 9.30 seconds (95% credible interval [8.86, 9.64 seconds]). The β parameters reflect how quickly the proportion of participants in the region of each option increased after time τ . The estimate for $\beta_{Heuristic}$ (median 0.027, or a 2.7% increase per second, 95% credible interval [2.5%, 2.9%]) was almost twice as large as that for $\beta_{Correct}$ (median 0.015, or a 1.5% increase per second, 95% credible interval [1.4%, 1.6%]). To summarise, participants began to be drawn towards the heuristic option from 3.6 seconds, and towards the correct option from 9.3 seconds. After these onsets of attraction, there was a greater increase in the proportion of trials where the cursor was in the region of the heuristic option (2.7% per second) than in the proportion of trials where it was in the region of the correct option (1.5% per second).

GROWTH CURVE MODELLING

This time course data also allows us to supplement the response time analyses reported above by looking at the speed at which participants moved the mouse cursor to the region of the response option they eventually did select. Figure 4 shows this measure for correct responses to no-conflict problems, and for both heuristic and correct responses to conflict problems. The curve for each response region over time was modelled using third-order polynomial logistic regression models (or

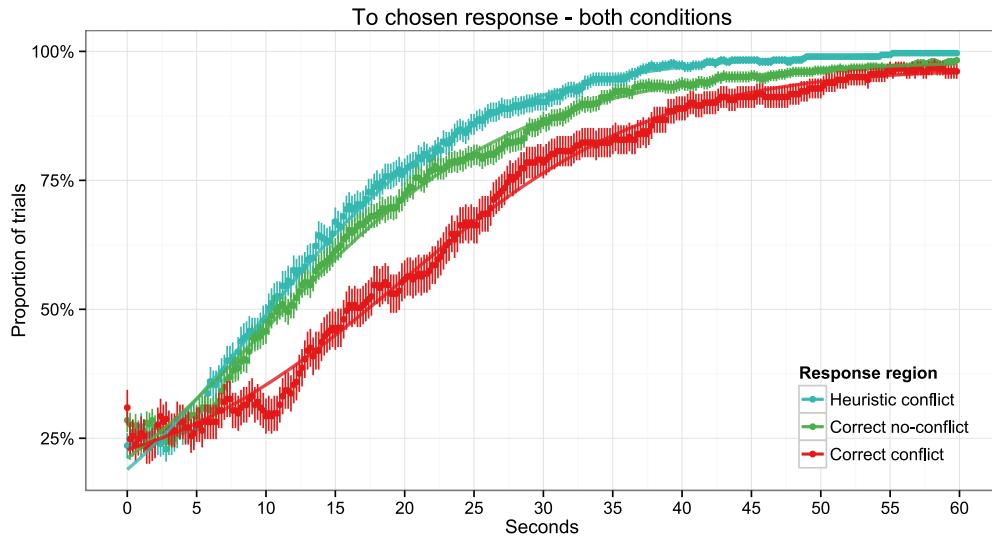


Figure 6.5: Proportion of mouse cursors in the region of the response option which was ultimately selected on that trial.

growth curves; see Mirman, 2014b), such that the log odds of the cursor being in that region were given as $\alpha + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$. Natural polynomials were used, meaning that the intercept corresponded to the log odds at 0 seconds, the linear term to the simple change over time, and the quadratic and cubic terms to higher-order 'wiggles' later in the time course.¹

To test for a significant difference between two curves, a null model, in which the weights were the same for each curve, was compared with a full model, in which there were different weights for each curve. Chi-squared tests were used to compare the deviance of each model, with degrees of freedom corresponding to the number of parameters added in the full model. Note that α , the intercept, was not allowed to vary between curves. Finally, a random effect for the linear time term was included for each participant, to allow for individual variability in how quickly each participant moved towards a response in general. Random effects on other terms, by participant, or by problem, were considered, but led to convergence issues, and so only this term, which was found to account for the most variance, was included.

¹ One disadvantage of using these natural polynomial terms is that they are by definition correlated, and so the model suffers from mild multicollinearity, which leads to some loss of statistical power. However, as the alternative, orthogonal polynomial terms would be difficult to interpret individually, I believe this approach lends itself to a clearer description of the data.

Mirroring the response time analyses, and as predicted by all dual process accounts, participants were faster to move towards the heuristic response option when selecting it than the correct option for conflict problems ($\chi^2 = 4515.7$, DF = 3, p < .0001), with the curves differing significantly on the linear, quadratic, and cubic terms (z 's > 5, p's < .0001; see Figure 6.5). Again consistent with the response time analyses, and contrary to previous findings supportive of the intuitive logic model, participants were faster to move towards the heuristic response on conflict problems than to move towards the correct response on no-conflict problems (χ^2 , DF = 3, p < .0001). This effect was mainly driven by a significant difference on the linear term between the curves ($z = 2.352$, p = .0187).

Most dual process theories, including default-interventionist, parallel-competitive, and intuitive logic accounts, would predict that participants should be drawn towards the heuristic option on trials where they ultimately give the correct response. In order to test for this attraction, I compared the probability over time of the cursor being in the region of the heuristic option with the average probability of it being in the region of either foil option on those trials (Figure 6.6). A higher probability of being in the region of the heuristic option than the foils constitutes evidence of an attraction towards that heuristic response. Visual inspection of Figure 6.6 shows that this is the case from approximately 10 seconds onwards. Again, third order polynomial regression models were fit to this data, which showed that the difference between the curves was statistically significant ($\chi^2 = 428.2$, DF = 3, p < .0001), with significant differences on the linear, quadratic, and cubic terms (z 's > 2.1, p's < .05). Therefore, when selecting the correct response, participants were more drawn to the heuristic option than to the foils, as predicted both by default-interventionist and parallel-competitive or intuitive logic accounts.

A more interesting comparison is between the attraction towards the correct response option, and that towards the foil options, on conflict trials where the heuristic response is given. According to the default-interventionist account, Type 2 processes have not become engaged at this point, and so the correct response option should not be any more attractive than the foil options. According to the parallel-competitive account, on the other hand, both Type 1 and Type 2 processes should be engaged on such trials, and so participants should be drawn to-

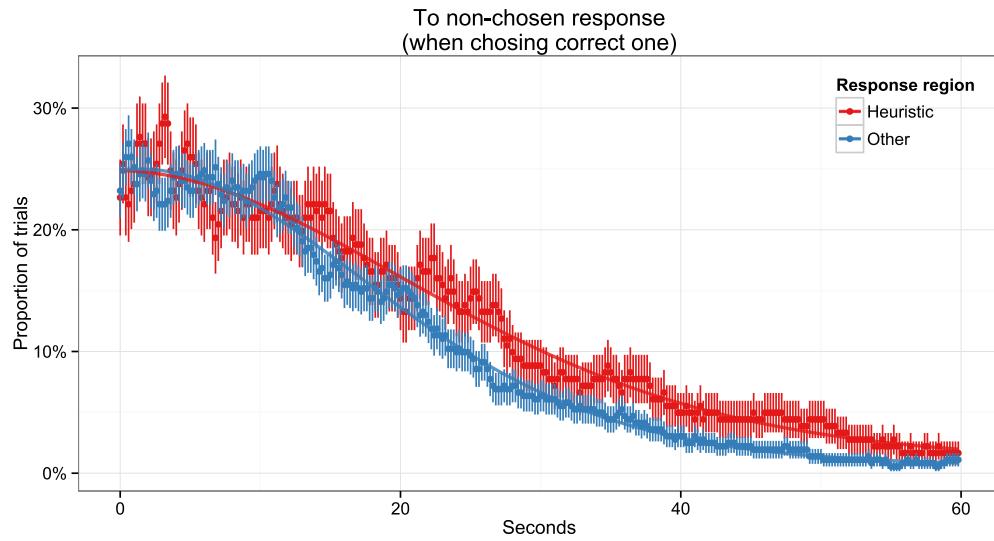


Figure 6.6: Proportion of trials in the region of each option, over time, for trials in which the correct option was eventually chosen, for conflict problems. Error bars show standard error of measurement. Lines show fitted polynomial regression curves. Participants were more likely to be in the region of the heuristic response from around 10 seconds onwards.

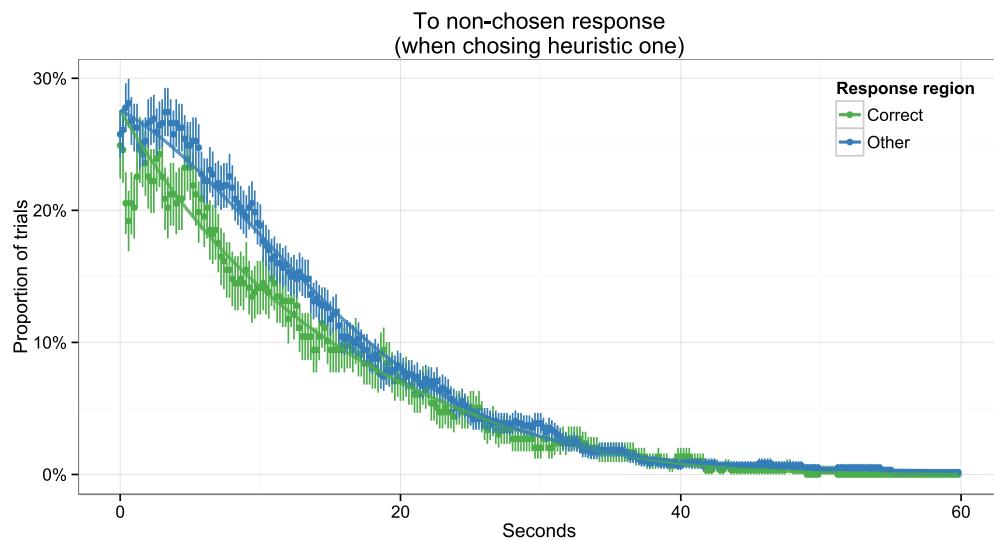


Figure 6.7: Proportion of trials in the region of each option, over time, for trials in which the intuitive option was eventually chosen, for conflict problems. Participants were less or equally likely to be in the region of the correct option than a foil throughout.

wards giving the response cued by Type 2 processes (that is, the correct response). Either result could be consistent with the intuitive logic theory, depending on the mechanism by which conflict is actually detected. If conflict detection occurs because Type 1 processes simultaneously cue both the correct and heuristic responses, then attraction towards the correct response option should be seen here. Conversely, if conflict is detected without Type 1 processes actually producing the correct response, then the intuitive logic theory, like the classic default-interventionist account, would predict no attraction towards the correct response option here. Of course, no conflict would be predicted by a selective theory.

Figure 6.7 shows that, contrary to the prediction of the parallel-competitive and intuitive logic accounts, participants are not more likely to move towards the correct response option than either of the foils before giving the heuristic response. Participants were in fact less likely to be in the region of the correct option than the foils. The polynomial regression model showed that the difference between the curves shown is again significant ($\chi^2 = 208.0$, DF = 3, $p < .0001$), with significant differences between the curves on the linear, quadratic, and cubic terms (z 's > 9 , p 's $< .0001$). This result indicates that the correct responses were on average actually less attractive than the foils. This is perhaps unsurprising, given that part of the difficulty of the CRT lies in the failure of intuition to support the correct response.

6.4 DISCUSSION

These results are broadly consistent with a default-interventionist dual process theory (Evans, 2006; Kahneman & Frederick, 2005). On problems with an incorrect but intuitively appealing heuristic response, this response was given more quickly, and with less evidence of conflict, than the correct response. Participants began to systematically move the mouse cursor to the region of the heuristic response option within approximately 5 seconds, compared to 10 seconds for movements to the correct response option, and this trend was evident both when analysing all trials, and trials in which the response in question was given. This also appears to be true of both biased and unbiased participants.

When participants did give the correct response on these conflict problems, they

spent more time in the region of the heuristic response option than either of the foil options before doing so — a finding consistent with default-interventionist, parallel-competitive, and intuitive logic accounts, suggesting that these participants considered the heuristic response before they reached the correct one. This finding is also consistent with modelling work (Böckenholt, 2012; Campitelli & Gerrans, 2013), and individual differences studies (Liberali, Reyna, Furlan, Stein, & Pardo, 2012) which have shown that inhibition of the heuristic response is an important predictor of accuracy on the CRT. However, contrary to the prediction made from a parallel-competitive dual process theory (Sloman, 1996, 2014b) or by the intuitive logic account (De Neys, 2012, 2014; Handley & Trippas, 2015) on trials where the heuristic response was given participants were no more likely to place the cursor in the region of the correct response option than either foil option.

These results also have implications for the logical intuitions theory (De Neys, 2012, 2014). First, a number of previous studies using simpler reasoning tasks have found that heuristic responses to conflict problems take longer than correct responses to no-conflict problems, despite both being cued by Type 1 processes (De Neys & Glumicic, 2008; Stupple & Ball, 2008). To my knowledge, the current study is the first to report response times for conflict and no-conflict versions of the CRT, and although this analysis was not the main focus of the this experiment, I found no such effect. In fact, when analysing participants speed of movement to the response option they ultimately selected, a more sensitive measure, I found the opposite effect, with participants faster to move to the heuristic option under conflict than the correct option for no-conflict problems on trials where these responses were given. All of these findings were true for both participants who gave the heuristic response to most conflict problems, and those who did not. Thus, unlike a number of studies using simpler reasoning problems, I found no evidence that participants were slower to give intuitively-cued responses which were wrong than intuitively-cued responses which were right.

Secondly, as discussed above, I found no evidence of an attraction towards the correct response option on conflict problems where the heuristic response was given. This suggests that Type 1 processes did not simultaneously cue both responses on such trials. This result is not, however, totally inconsistent with the logical intu-

itions theory. Previously, I differentiated between a dual process conflict monitoring account (De Neys & Glumicic, 2008; De Neys et al., 2008), that proposes that we detect when our reasoning is biased, and a fully-fledged intuitive logic theory (De Neys, 2012, 2014; Handley & Trippas, 2015), where this conflict detection is the result of Type 1 processes simultaneously cuing both the correct and the heuristic response. These results would appear to contradict the latter account, whereby Type 1 processes cue both the heuristic response (“10p”) and the correct response (“5p”) at the same time for the bat-and-ball problem and other CRT items. They do not, however, rule out the possibility that participants experience uncertainty (De Neys et al., 2013), or a feeling of wrongness (Gangemi et al., 2015) while solving these conflict problems. If this is the case, further work is needed to reveal how this feeling comes about.

At this point, I would like to note again that, De Neys et al. (2013) and (Gangemi et al., 2015) notwithstanding, previous evidence for the intuitive logic theory has come from simpler experiments, such as simple syllogistic reasoning (Morsanyi & Handley, 2012) and forced-choice base rate neglect (De Neys & Glumicic, 2008) paradigms. The operations required to reach the correct answer to these CRT problems are considerably more complex than those needed to evaluate a simple syllogism, or apply basic statistical principles. Therefore, while I do not find evidence that Type 1 processes automatically generate correct responses on the CRT, this does not rule out the possibility that they can generate correct responses on these simpler tasks. For instance, it has been demonstrated that participants report “liking” syllogisms which are logically valid more than those which are invalid, even when not asked to evaluate their logical status (Morsanyi & Handley, 2012), but also that this effect only holds for simpler logical forms (Klauer & Singmann, 2013, see also Handley & Trippas, 2015). Indeed, De Neys (2012), when proposing the intuitive logic account raised the possibility that it may not apply to all problems.

One might argue that the absence of evidence for either the parallel-competitive or intuitive logical theories here do not reflect evidence against these accounts, but rather the inability of this paradigm to reveal the effects predicted by these accounts. It may be the case, according to this line of reasoning, that participants

are drawn towards the correct option on trials where they give the heuristic response, or that participants are more conflicted when their heuristic responses are wrong than when they are right, but that I was unable to detect these mental states using this new paradigm. While I cannot completely rule out this possibility, I believe two factors go against such an interpretation. First, this paradigm does reveal effects, such as attraction towards the heuristic option before giving the correct response, consistent with the default-interventionist model. Second, for the two comparisons above, rather than finding no effect, I found significant effects in the opposite direction to those predicted by parallel-competitive and intuitive logic accounts.

Additionally, there is extensive evidence that even subtle, implicit cognitive processes influence motor output in detectable ways (Bargh, 2006; Miles, Betka, Pendry, & Macrae, 2010; Tucker & Ellis, 2004; Xiao & Yamauchi, 2014). Therefore, if participants do experience conflict, but this conflict does not influence their motor output, then this raises the question of what mechanism produces this conflict while not influencing motor output. I return to this issue in Chapter 7. Finally, I would note again that these results should be interpreted as constraining the intuitive logic account, rather than falsifying it.

Of course, all of the above assumes a dual process interpretation of the CRT, as most treatments of the task do. Even in accounts which focus instead on dispositional factors (Campitelli & Gerrans, 2013; Campitelli & Labollita, 2010), it is acknowledged that responding correctly typically requires the inhibition of the heuristic response. While I am unaware of any accounts of the CRT which do not rely on such an inhibition, I cannot rule out the possibility of such explanations being offered in future. The current results, however, provide an additional constraint to such accounts, in that they should predict not only observed choices, but also the patterns at the process level reported here.

As a side note, it may be noted that it is unusual to present the CRT as a multiple-choice test, and that this may affect the processes engaged during this experiment. However, multiple-choice versions for the test have been previously reported by Morsanyi, Busdraghi, and Primi (2014), Primi et al. (2015, Experiment 3), and Gangemi et al. (2015, Experiment 2), without any clear effect on participants' re-

sponses.

Finally, since its introduction in 2005, the CRT has been hugely popular as a measure of individual differences in thinking, despite only limited evidence as to what underlies performance on the task. These results go some way towards filling this gap, and suggest that responding correctly does require the activation of otherwise dormant Type 2 processes to override incorrect intuitions. Future work might address the relationship between conflict on this task and individual differences. Stanovich and West (2008) proposed that normative decision making requires (1) awareness of the limitations of intuition; (2) desire to overcome those limitations; (3) inhibition of the intuitive response and (4) ability to generate the correct response. Each of these requirements is a distinct reason for failure to produce the correct response on the CRT, and each should produce a distinctive pattern in mouse cursor movement data.

To conclude, I recorded participants' mouse cursor movements over a considerable period of time while they reasoned about CRT problems. Trends in these movements were consistent with a default-interventionist dual process theory of reasoning, where participants are initially drawn towards heuristic responses only, but in some cases engage further effortful processing to find correct solutions. I did not find evidence that participants were drawn to correct responses on trials where these responses were not actually given, inconsistent with a parallel-competitive dual process account. Finally, contrary to previous work using simpler reasoning tasks, and confidence ratings collected on the CRT, I found no evidence that participants were conflicted when giving incorrect heuristic responses.

CHAPTER 7

Summary, Discussion and Final Conclusions

7.1 RECAPITULATION

In this thesis, I have used the mouse tracking paradigm across a series of experiments to investigate conflict in reasoning. In doing so, I hoped to further understand under what circumstances conflict arises, what form this conflict takes, and at what points in time it occurs.

In Experiments 1 and 2 (Chapter 3), I pitted perceptual cues in the form of visual similarity, against conceptual knowledge in the form of category membership, in a forced-choice induction task, using both natural (Experiment 1) and artificial categories (Experiment 2). In both experiments, I found that perceptual cues are an early driver of participants' motor output, and that conceptual knowledge is brought online later in reasoning, sometimes causing participants to change direction if they had begun to move towards the perceptually-cued option. Analysis of the time course data showed that participants began to be drawn towards the perceptually-cued foil from 300 msec (in both experiments), and did not begin to inhibit this attraction until \sim 750–1,500 msec. In Experiment 2, I further showed that participants were more likely to override their perceptually-driven movements when reasoning about properties that were conceptually related to the distinction between the categories, likely because these properties make conceptual knowledge more accessible, or help participants realise that the perceptual cues were inappropriate. Again, this effect, being dependent on conceptual knowledge, was not visible until \sim 620 msec.

Experiments 3 and 4 (Chapter 4) were similar, but pitted associative knowledge against structured knowledge. In Experiment 3, I contrasted conflict trials, where the foil response was strongly associated with the base according to separate association ratings, to control trials, where it was not. In Experiment 4, I collected association ratings from each participant for each pair of species. This allowed me to conduct regression analyses with both each participants' associative knowledge (the ratio of the association ratings in favour of each response) and structured knowledge as predictors.

In both experiments, I again found that participants were influenced by both kinds of information: in Experiment 3 participants gave the correct response on ~95% of control trials but only ~75% of conflict trials, while in Experiment 4 both structured and associative knowledge were significant predictors of participants' responses. However, compared to the previous experiments, there was less evidence here that participants were initially drawn towards the associatively-cued option, and then subsequently towards the structured response. While this did happen on some trials, on most trials participants either moved straight to one response option, or straight to the other.

In Experiments 5 and 6, I explored conflict from the perspective of dual process theories of reasoning. In Experiment 5 (Chapter 5), participants completed a mouse tracking version of the base rate neglect task (De Neys & Glumicic, 2008; Kahneman & Tversky, 1973) where they could draw on either stereotypical descriptions or statistical base rates to decide someone's social category. I manipulated both the descriptions and the base rates, so that they could either agree, disagree, or only one cue was informative.

Consistent with a default-interventionist dual process model (i.e. Evans, 2006), I found that participants' responses were predominantly determined by the contents of the descriptions, and that participants who opted to ignore the description and rely on the base rate instead experienced conflict. In line with previous work (e.g. Kahneman & Tversky, 1973; Tversky & Kahneman, 1982), but counter to intuitive logic accounts (De Neys & Glumicic, 2008; Pennycook et al., 2014) or a parallel-competitive dual-process account (Sloman, 1996), the influence of the base rates was less pronounced. Specifically, participants did sometimes give the base rate-

cued response even when it conflicted with the description (~20% of trials). Aside from this, however, participants responding on the basis of the descriptions did not show signs of conflict when the base rate was manipulated to disagree with the description. A Bayesian follow-up analysis, finally, showed that despite the non-significant effect, there was not considerable evidence for the existence of a null effect either. Rather, it seems that participants were very slightly (~3%, or around 30 msec) slower giving the description-cued response when it disagreed with the base rate.

In Experiment 6 (Chapter 6), participants completed the more complicated Cognitive Reflection Test (CRT; Frederick, 2005): a series of questions for which the first response that comes to mind is often incorrect. Unlike Experiments 1 to 5, there were four response options here, located in each corner of the screen. Participants were allowed to respond in their own time, and I analysed movements of the mouse cursor over the first sixty seconds of each trial as participants moved the mouse around the screen deciding on a response. This novel form of mouse tracking made it possible to record participants' ongoing decisions over a longer period of time. It also allowed me to infer participants' attraction towards four response options, rather than the two used in standard mouse tracking. This was useful in differentiating between effects where participants are actually drawn towards a particular alternative response option, and those where participants are merely slow to move towards the option they do choose. I also included, following De Neys et al. (2013), no-conflict versions of the CRT problems to serve as a control condition where participants' first, heuristic responses are the correct ones. Finally, this experiment was, to my knowledge, the first to record response latencies on any version of the CRT.

Consistent with a dual process perspective, heuristic responses on the CRT were given more quickly, and participants were faster to approach these response options before selecting them, than correct options. Contrary to response time analyses of other tasks (e.g. De Neys & Glumicic, 2008) and confidence rating data from the CRT (De Neys et al., 2013; Gangemi et al., 2015) however, participants were no faster to give the heuristic (and correct) response on no-conflict problems than to give the heuristic (and incorrect) response on conflict problems.

The nature of the mouse tracking data here allowed for a number of novel analyses. First, I demonstrated that on conflict problems participants began to be drawn towards the heuristic option from 3.7 seconds, and towards the correct option from 9.3 seconds, including time spent reading the problems. More relevantly, to test for the presence of conflict I calculated, for trials where participants gave one response, the degree to which their mouse cursor was in the region of the alternative, competitor response beforehand, compared to the regions of the other two non-competing responses. Thus, for instance, if a participant giving the correct response is drawn towards the heuristic option while doing so, their cursor will spend more time in the region of the heuristic option, on average, than either of the other two foil options. Again in line with a default-interventionist model, I found exactly this effect, as participants were drawn towards the heuristic option before selecting the correct one. However, contrary to some other accounts, the reverse was not true: participants did not hover over the correct option before ultimately selecting the heuristic one.

In this final chapter, I try to make sense of these results and discuss the implications they have for what we know about conflict in reasoning. I do this both within the specific sub-domains that these experiments addressed — information selection in induction, and dual process theories of reasoning — and in terms of conflict in reasoning and cognition more broadly.

7.2 CONFLICT IN REASONING

It is clear, both from this thesis and from a wealth of previous work (M. M. Botvinick et al., 2004; Crisp-Bright, 2010; De Neys, 2012; E. K. Miller & Cohen, 2001), that conflict in cognition arises in many places. In Chapter 1, I introduced two particular junctures: conflict between competing representations in induction, and conflict between Type 1 and Type 2 processes. At this point, it is worthwhile considering how these two forms of conflict fit into the broader scope of the interacting processes that underlie reasoning.

I propose that both kinds of conflict can be understood in terms of the role of working memory in reasoning, and the interaction of working memory-based processes with other cognitive processes.

Some cognitive functions are almost certainly achieved by fast, associative, processes. Typical, but not defining properties of these processes are that they are autonomous (they operate automatically when presented with their triggering cues), they are associative (rather than rule-based), they generally operate quickly, and are not cognitively demanding. In dual process accounts of cognition, these are known as Type 1 processes (or, in the past, as *System 1*; e.g. Sloman, 1996). Stanovich (2005, 2009a), highlighting the diversity of these processes, labels them the *Autonomous Set of Systems*. In his definition of this set of processes, Stanovich (2009a) includes both domain specific evolved *modules*, and general learned associations, which have become autonomous through practice or repeated exposure. While some of these Type 1 processes serve to provide information to other cognitive process, including for use in more effortful, deliberate cognition, others can affect our actions directly without being mediated by other processes.

Clearly, under this definition, many cognitive processes are autonomous. To fully catalogue these processes would take a lifetime, and so for the purpose of this discussion it suffices to say that there exist a constellation of processes that are generally fast, associative, and autonomic. Well-known such processes include face recognition (largely an evolved processes, localised to the fusiform gyrus; Kanwisher, McDermott, & Chun, 1997), reading text (a learned skill that maps onto the left fusiform gyrus; McCandliss, Cohen, & Dehaene, 2003), and word recognition (e.g. Spivey et al., 2005).

Other processes, however, require the sustained representation, maintenance, and manipulation of information in working memory, *decoupled* from interference from competing Type 1 processes (Evans & Stanovich, 2013; Gilbert, 1991; Stanovich & West, 2008). More recent dual process accounts (see Evans & Over, 2004; Evans & Stanovich, 2013; Stanovich & Toplak, 2012) propose that this decoupling in working memory is the defining feature of Type 2 processes. In this account, Type 1 processes, in contrast, are simply those that are *autonomous*, or do not require controlled attention or working memory resources (Evans & Stanovich, 2013, p. 236).

These Type 2 processes, making use of working memory, have the advantage of being extremely flexible. Many Type 1 processes are thought to be *domain specific*

(see Cosmides & Tooby, 1994), in that they are restricted to processing only one kind of information. An archetypal example of this is the process responsible for face perception, a small system localised to the bilateral fusiform gyri (or *fusiform face areas*; see Kanwisher et al., 1997). This system is exquisitely well adapted to recognise human faces in visual input, but when presented with other stimuli will either not activate, or mistakenly indicate that it has seen a face.¹ Type 2 processes, however, are *domain general*, and can equally well process information about faces, words, geographical locations, and abstract concepts. These processes are also thought to be unique in their ability to sequentially apply rule-based operations (Anderson, 1996; Anderson & Fincham, 2014; Cooper & Shepard, 1973), in contrast to Type 1 processes, which are thought to work on associative principles (see Sloman, 1996). This flexibility comes at the cost, however, of limited capacity. Working memory capacity — and thus the capacity of Type 2 processes — is limited first of all in that, compared to Type 1 processes, through which an enormous amount of information streams in parallel, only a very small amount of information can be held in working memory at one time (famously 7 chunks of information, ± 2 , according to G. A. Miller, 1956). Working memory is limited secondly in that many theories hold that only one state of the world, or mental representation, can be held in working memory at any one time. This means that we can consider one possible mental representation, followed by another, but that we cannot simultaneously consider two contradictory representations. This limitation has been noted across a number of research traditions, including in dual process theories as the *singularity principle* (Evans, 1984, 2006), in mental models accounts of reasoning (Johnson-Laird, 1983) as *focusing* (Legrenzi, Girotto, & Johnson-Laird, 1993), by those interested in diagnosis as an inability to consider more than one hypothesis at a time (Mynatt, Doherty, & Dragan, 1993), as well as in the literature on working memory itself (Baddeley, 2007).

My proposition is as follows. First, the inductive triad tasks used in Chapters 3 and 4 require that participants hold the three categories presented in working memory, to compare the inductive potential from the base to each of the candidate responses. To do this, they must draw on information about the relationships

¹ The common illusion of perceiving faces in non-facial stimuli is known as pareidolia.

between the categories shown. This information, naturally, must come from elsewhere, and in these tasks it can come from a number of Type 1 processes: from the visual system, for instance, or from different aspects of long term memory (Jackson et al., 2015), including both associative and structured knowledge. As only one representation of the world can be held in working memory at a time, conflict arises when multiple Type 1 processes provide multiple contradictory representations. I contend that the induction tasks reported here involve conflict in this sense, as representations based on perceptual cues or conceptual knowledge (Experiments 1 and 2), or on associative and structured knowledge (Experiments 3 and 4) vie to be realised.

Second, there are some problems for which responses can be cued either by Type 1 processes, operating on associative principles and placing few demands on working memory, and by Type 2 processes that involve maintaining information in working memory while applying sequential, rule-based operations. These are, of course, the problems addressed by dual process theories of reasoning. (e.g. Evans, 2008; Evans & Stanovich, 2013; Kahneman, 2011). I investigated two such problems in this thesis: reasoning about base rates and stereotypes (Experiments 5), and the Cognitive Reflection Test (Experiment 6). My account for performance on these tasks is no different from the generic dual process account. On the base rate task, Type 1 processes are responsible for processing the descriptions, and automatically cue stereotype-consistent responses. To process and respond on the basis of the base rates, however, one must engage Type 2 processes to relate the statistical information to the task at hand and choose a response (and, under most accounts, to inhibit the description-cued response). Similarly, on the CRT, Type 1 processes automatically cue incorrect heuristic responses, without the need for sustained Type 2 processing. To reach the correct response, however, participants must represent the problem in working memory and apply the rules of arithmetic, an archetypal Type 2 operation.

Figure 7.1 shows a simple sketch of this framework for thinking about conflict in reasoning. In short, conflict can arise both as multiple Type 1 processes attempt to project information to working memory, or because both Type 1 and Type 2 processes attempt to produce responses to the same problem. With this framework

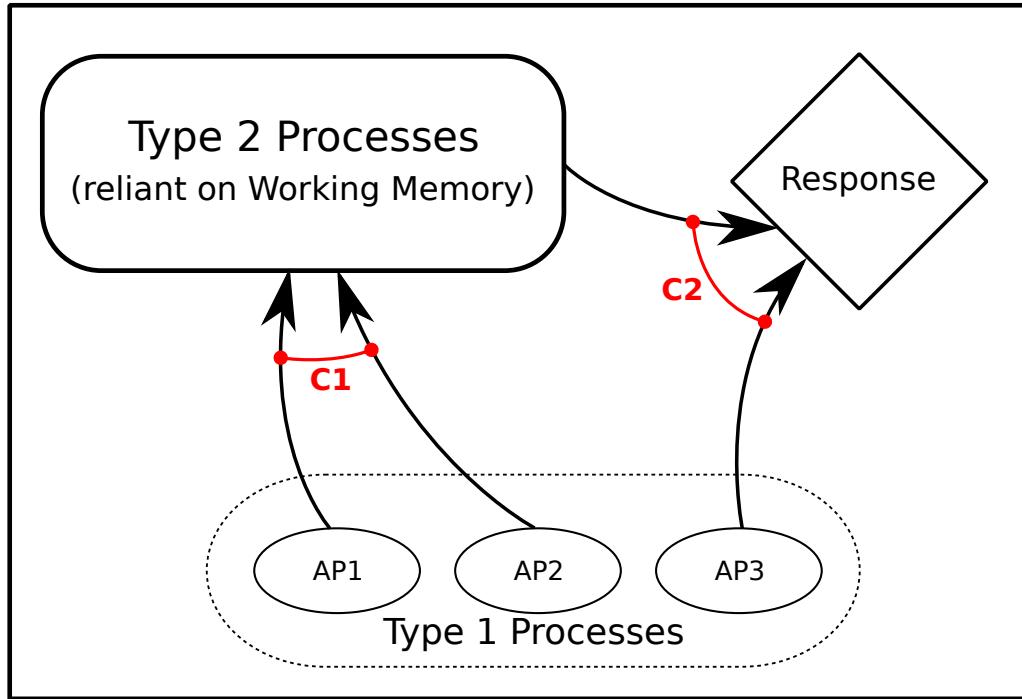


Figure 7.1: A simple framework for conflict in reasoning. Reasoning requires the interaction of associative, Type 1 processes (AP1, AP2, AP3, etc.), and Type 2 processes, reliant on working memory. Conflict in reasoning arises a) when multiple Type 1 processes attempt to project information to working memory (**C1**), and b) when both Type 1 and Type 2 processes attempt to produce responses (**C2**). I propose that conflict of type **C1** occurs when multiple sources of information, such as perceptual cues, associative knowledge, or structured knowledge, are available during inductive reasoning. Conflict of type **C2**, conversely, occurs in reasoning when both Type 1 and Type 2 processes can generate responses, for instance description and base rate-cued responses in Experiment 5, or heuristic and correct responses in Experiment 6.

in mind, I now return to the interpretation of my results.

7.3 KNOWLEDGE AND REASONING

In Experiments 1 and 2 (Chapter 3) and 3 and 4 (Chapter 4), I looked at what information people drew on during inductive reasoning, using triad tasks. In these, participants were asked to project a property from a base category, known to have the property in question, to one of the two other categories. Above, I argued that this task requires that participants represent the three categories in working memory, and then draw on information about each category to decide to which of the two response categories the property likely generalises. Thus, the framework above provides a mechanistic account of the operations proposed by Bright and Feeney's (2014b) hybrid theory of induction.

In Experiments 1 and 2, I pitted perceptual cues against conceptual knowledge, and in Experiments 3 and 4, I pitted associative knowledge against structured knowledge. An interesting question that arises at this point, therefore, is if these two kinds of conflict really unfold in the same way. In Experiments 1 and 2, the interaction of perceptual cues and conceptual knowledge was straightforward: perceptual cues drove participants' motor output early in reasoning, and conceptual knowledge was retrieved later, causing participants driven towards the foil response to sometimes change direction to select the right one instead. In Experiments 3 and 4 the picture was less clear. On the one hand, participants did sometimes initially move towards the foil option before giving the correct response, and a greater proportion of correct responses were reversals under conflict. On the other hand, these reversals were considerably rarer than in Experiments 1 and 2, as participants for the most part moved towards one other response option, and then selected it.

These differences can be captured by two of the parameters in my analysis of the transition probabilities of cursor trajectories: $1 - \alpha$, the probability of initially moving towards the foil option, which was higher for perceptual cues ($\sim 50\%$) than for associative knowledge ($\sim 35\%$), and γ , the probability of overriding an initial movement towards the foil, which was again higher for perceptual cues ($\sim 50\text{--}80\%$) than associative knowledge ($\sim 25\text{--}35\%$).

One possibility here is that the differences between the two sets of experiments are only quantitative. It may be that participants are very prone to draw on perceptual cues during reasoning, but that they are also relatively good at overriding the influence of these cues, doing so, at minimum, on 50% of trials. Conversely, it may be that associative knowledge may only be retrieved some of the time, but participants who do retrieve this knowledge are considerably less likely to subsequently inhibit or override it. From this point of view, the phenomena studied in both Experiments 1 and 2 and in Experiments 3 and 4 are examples of conflict occurring as different Type 1 processes attempt to project information to working memory.

An alternative possibility, however, is that the difference between the two sets of experiments is qualitative. An extensive body of research (see Goodale, Westwood,

& Milner, 2004, for a review) shows that information passes through the human visual system in two streams. A dorsal stream leads from the visual cortices to the prefrontal cortex and allows us to consciously perceive stimuli and to hold them in working memory. A ventral stream, however, leads directly to the motor cortex and allows our actions to be informed by perceptual cues without conscious awareness.² It could be the case that rather than participants seeing the visual cues in Experiments 1 and 2, representing this information in working memory, and reasoning from there, this information may have been passed directly from the perceptual system to the motor cortex. In this case, these results could also be interpreted as a dual process phenomenon, if we allow that this perceptual-action coupling constitutes a Type 1 process.

On the basis of the current data, it is not possible to differentiate between these two possibilities. A possible future approach, however, may be to investigate the relationship between performance on both of these versions of the triad task and measures of different kinds of inhibitory control. According to Diamond (2013), there are broadly speaking two kinds of inhibitory processes. One is *response inhibition*, which allows us to override automatically executed motor commands. The purest (i.e. least confounded by other factors) measures of this kind of control are the Go/No-Go (Donders, 1868/1969) and Stop-signal (Lappin & Eriksen, 1966) tasks. Both of these require participants to quickly give a response associated with a given stimulus, except on trials where a no-go or stop signal is presented. The other inhibitory process is *cognitive inhibition*, which inhibits and filters out irrelevant or unwanted mental representations and information. Of particular relevance here is *semantic* inhibitory control, a subset of cognitive inhibition measured using the Hayling task (Burgess & Shallice, 1997). In the non-clinical version of this task (Markovits & Doyon, 2004), participants are primed with incomplete sentences (e.g. “The captain wanted to stay with the sinking ____”), and then shown a single word and asked to indicate if it is an appropriate word with which to finish the sentence. Crucially, on lure trials the probe was a pseudo-word similar to a real word primed by the sentence — i.e. “shifp” when “ship” was expected. This re-

² In the extreme, damage to ventral system can leave individuals able to respond to visual stimuli, without being consciously aware of seeing them, a condition known as *blindsight* (Sali et al., 2015; Weiskrantz, 1986)

quires that participants inhibit their automatically primed semantic knowledge to correctly reject this word.

Bright and Feeney (submitted) presented participants with the triad task from Experiment 3, which placed associative and structured knowledge in conflict. They also collected measures of both response inhibition (the Stop-signal task) and semantic inhibition (the Hayling task). Their results showed that the effect of associative knowledge — fewer correct responses when associative knowledge cued the foil — was more pronounced for participants who performed poorly on the Hayling task (responded that the lure pseudo-words were appropriate), but was not related to individual differences in performance on a Stop-signal task. Therefore, they concluded that foil responses on this task are the result of a failure to inhibit inappropriate semantic knowledge in favour of structured knowledge, rather than a failure to inhibit the inappropriate response itself.

Applying this logic to Experiments 1 and 2, I would predict that if foil responses on these tasks are the result of conflict between representations in working memory like foil responses in Experiment 3 and 4, and in Bright and Feeney's (submitted) experiment, participants lacking in semantic inhibitory control should be more likely to err in this way. Conversely, if these foil responses are instead the result of a failure to inhibit the response automatically cued by the visual cues, it should be participants with poorer response inhibition that err in this way instead. Clearly, further research is needed to resolve this question.

Beyond this question, however, the experiments reported in both Chapters 3 and 4 have implications for theories of induction more broadly. First, these results are consistent with Bright and Feeney's (2014b) argument that theories of induction that draw on just one kind of knowledge are incomplete. Instead, it seems that induction can be driven by multiple forms of information. Bright and Feeney (2014b) showed that different sources of information are used depending both on individual differences (semantic inhibitory control) and on situational factors (the presence of secondary load). The current results showed that multiple kinds of information can drive reasoning even within a single trial, leading to conflict.

Formal modelling plays an important role in research on induction (see, e.g., Hawkins et al., 2015; Kemp & Tenenbaum, 2009; Osherson et al., 1990; Sloman,

1993). These results, however, are challenging to model using existing frameworks. A natural next step, following on from Bright and Feeney's (2014b) hybrid account, would be to attempt to model these time course data using a hybrid of existing models based on associative and structured knowledge.

This could potentially be achieved, for instance, by a change-point model where reasoning is driven by one kind of information before time t , and by another model afterwards. Another potentially interesting avenue here is the CLARION cognitive architecture (e.g. Sun, 1995; Sun & Zhang, 2006), which combines implicit similarity-driven and explicit rule-based operations. Sun and Zhang (2006), for instance, use this architecture to model performance on a number of inductive reasoning problems. However, at present this model does not account for the time course data presented here, and so further work would be needed to model the full temporal dynamics of reasoning.

7.4 DUAL PROCESSES AND REASONING

In Experiments 5 and 6 (Chapters 5 and 6), I tested the predictions of dual process theories of reasoning. At the outset, I introduced a number of types of dual process architecture outlined by Evans (2007): selective/pre-emptive conflict resolution models, where reasoners selectively draw on Type 1 *or* Type 2 processes each time; corrective/default-interventionist models, where Type 1 processes are activated automatically, and optionally overridden by Type 2 processes; and parallel-competitive models, where both types of process operate in parallel, and compete for control of behaviour. I also discussed what I referred to as the dual process conflict monitoring account (e.g. De Neys & Glumicic, 2008; De Neys et al., 2008), which proposes that participants implicitly detect that their heuristic reasoning conflicts with normative standards, and the more recent *intuitive logic* accounts of reasoning (De Neys, 2012, 2014; Handley & Trippas, 2015), which propose that this conflict detection happens because Type 1 processes can simultaneously cue both biased heuristic responses and normatively correct ones.

These accounts differ in when they predict conflict should occur, and mouse tracking provides a means of detecting this conflict when it does take place. Pre-emptive conflict resolution models predict that Type 1 and 2 processes never con-

flict, as we selectively activate only one or other kind of process on each task. Default-interventionist models predict conflict only when we activate Type 2 processes to override Type 1 responses. Parallel-competitive models, however, predict conflict to be more common, as both types of process usually compete to control behaviour. Like parallel-competitive models, intuitive logic accounts would predict that we are often conflicted during reasoning, but while the former claim this is because Type 1 and Type 2 process cue conflicting responses, the latter claim this occurs because Type 1 processes simultaneously cue multiple responses, often including both biased heuristic responses and normative correct ones.

Experiments 5 and 6 both tested these various predictions: Experiment 5 using the base rate neglect paradigm, where participants responded within five seconds, and Experiment 6 using the four-alternative Cognitive Reflection Test (CRT; Frederick, 2005), where participants were not placed under time pressure. In both cases, I found results strongly in favour of a default-interventionist model. First, in the base rate neglect task, Type 1 processes are traditionally thought to underlie description-driven responses, and Type 2 processes to underlie base rate-driven responses (Barbey & Sloman, 2007; Kahneman & Frederick, 2002, 2005). Participants' early, initial movements were sensitive to manipulations of the description, but not to manipulations of the base rate, as they moved towards the description-cued option ~66% of the time, regardless of the base rate. It was only later in the reasoning process (after 750 msec or so) that base rates began to influence participants' movements, and ultimately, some of their responses. Therefore, it appears that descriptions were processed automatically, by default, and only in some cases did participants attend to the base rates and override their default responses. Also consistent with a dual process interpretation, when the two cues conflicted participants overwhelmingly gave the description-cued response, and were less conflicted while doing so.

In the CRT, similarly, Type 1 processes are traditionally held (Frederick, 2005; Kahneman & Frederick, 2005) to drive heuristic responses (e.g. “10p” on the bat-and-ball problem) whereas Type 2 processes drive correct responses (e.g. “5p”). Consistent with this, participants predominantly gave the heuristic response on conflict problems, and were faster to do so, showed less signs of conflict, and were

also faster to approach this heuristic response before clicking it. Analysis of participants' cursor movements also showed that participants giving the correct response were more likely to hover in the region of the heuristic option than any of the other foil options before doing so. This would indicate that these heuristic responses must be inhibited before Type 2 processes can produce the correct response.

The dual process conflict monitoring theory (e.g. De Neys & Glumicic, 2008; De Neys et al., 2008), intuitive logic accounts (De Neys, 2012, 2014; Handley & Tripapas, 2015), and the parallel-competitive dual process theory (Sloman, 1996, 2014b) make additional predictions about when conflict should occur. According to the conflict monitoring theory, participants responding heuristically should show evidence of conflict on problems where their response is not also the normatively correct one, compared to problems where it is (in other words, they show signs of conflict when their heuristic response is wrong, but not when it is right). The intuitive logic and parallel-competitive accounts go further, and propose that this conflict occurs because both heuristic and correct responses are being cued simultaneously. The former proposes that Type 1 processes cue both responses but that the heuristic response, being prepotent, is the one normally given. The latter proposes that Type 2 processes cue the correct response at the same time as Type 1 processes cue the heuristic one.

In the two-alternative base rate neglect task, I tested these predictions by looking to problems where participants give the description-cued response, while the base rate either agreed with the description, was uninformative, or disagreed. Previous studies (i.e. De Neys & Glumicic, 2008; De Neys et al., 2008; Pennycook, Fugelsang, & Koehler, 2012; Pennycook et al., 2014) have found that participants are slower to respond when the description disagrees with the base rate. Here, I did not find such an effect, in response times or any other measure — although there was not strong evidence in favour of a null effect either. It appears that participants either gave the base rate-cued response, or (largely) ignored the base rates altogether. Therefore, while these results do not support the intuitive logic theory, they do not falsify it either.

It should be noted, however, that most studies showing an intuitive logic effect

in this paradigm involved participants responding much more slowly than was the case here. Therefore, it may be that participants rarely process base rates when required to respond in less than six seconds (see the Discussion of Chapter 5). I asked participants to respond quickly in this experiment, as in Experiments 3 and 4, in order to ensure that they were still making their decisions while moving the mouse cursor so that the cursor trajectories would reflect the reasoning process. In Experiment 6 however, due to the fortuitous fact that people will spontaneously move the cursor while inspecting and thinking about options located around the screen, I was able to let participants complete the CRT in their own time.

The design of this experiment allowed me to disentangle the predictions of the dual process conflict monitoring theory — heuristic responses to be slower when they are incorrect than when they are correct — from those of the stronger intuitive logic theory, and of a parallel-competitive dual process model — attraction towards the correct option before selecting the incorrect heuristic option on conflict problems. However, I not only failed to confirm these predictions, but in both cases in fact found significant effects in the opposite directions. Participants were slower to approach the heuristically-cued correct option on no-conflict problems where it was selected than to approach the heuristically-cued incorrect option on conflict problems where it was selected. Likewise, participants giving the correct response on conflict problems actually spent *less* time in the region of the correct option than the other foil options. However, these results were not predicted by any of the theories considered, and so were unexpected. Despite this, it seems that neither the imposition of time pressure nor a lack of statistical power can explain the absence of effects consistent with the intuitive logic theory here.

7.5 CONFLICT MORE BROADLY

Aside from the implications of my results for theories of reasoning, there is much in this thesis that may be of interest to those interested in conflict in cognition more broadly. In this section, I discuss a number of these broader issues.

7.5.1 CHANGES OF MIND IN COGNITION

In Chapters 1 and 2 (see also Appendix A), I discussed previous applications of the mouse tracking paradigm. Most of these studies to date, studying simple cognitive and perceptual tasks, have revealed continuous attraction effects, as participants are partially and simultaneously drawn towards two responses. However, a number of studies, mostly of more complex, high-level cognition (e.g. Barca & Pezzulo, 2015; Dale & Duran, 2011; McKinstry, Dale, & Spivey, 2008; Tomlinson et al., 2013) instead show *reversals*, or *changes of mind* as participants move first towards one response option, and then towards the other. In Experiments 1 to 5 of this thesis, similarly, I found that my high-level reasoning tasks yielded these changes of mind, and little evidence of continuous attraction (see Appendix B). At present, however, it is not yet clear why conflict works one way (continuous attraction) for simple tasks, and another (changes of mind) for more complex ones.

Above, I proposed a simple framework for thinking about conflict in reasoning. An important idea here was that, according to many accounts (Baddeley, 2003; Evans, 1984, 2006; Johnson-Laird, 1983; Legrenzi et al., 1993; Mynatt et al., 1993), only one state of the world can be represented in working memory at a time: it can believe that the correct response is A, or that it is B, or consider each in turn, or assign explicit probabilities to each possibility, but it cannot simultaneously represent a world where the correct response is A *and* a world where the correct response is B. Such fuzzy, graded, partially-overlapping representations, however, are fundamental to other theories of cognitive function, including those based on neural networks (Hinton & Anderson, 2014; Kruschke, 1992; McClelland, Mirman, Bolger, & Khaitan, 2014; McClelland, Rumelhart, & PDP Research Group, 1986; Rogers & McClelland, 2004), sequential sampling models (Leite & Ratcliff, 2010; Ratcliff, 1978; Ratcliff & Smith, 2010; Townsend & Busemeyer, 1995; Usher & McClelland, 2001), and dynamical systems (Freeman & Ambady, 2011a; Port & Gelder, 1995; Spivey, 2007; Van Gelder, 1998). More broadly, there has been philosophical debate as to whether mental representations in general are discrete and symbolic (Dietrich & Markman, 2003; Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991; A. Newell & Simon, 1972), or continuous, graded, and fuzzy (Beer, 1995; Huette & Spivey, 2012; Port & Gelder, 1995; Spivey, 2007).

I propose that continuous graded representations, and continuous attraction effects, are the hallmark of processes that do not require sustained representations in working memory, or, in other words, of Type 1 processes. Due to the unique nature of the symbolic, domain-general processing done in working memory, however, processes that rely on this (that is, Type 2 processes) operate in an all-or-nothing fashion: we ultimately decide that either A or B is the correct response. Of course, it is possible to explicitly consider uncertainty in the world, but doing so requires a great deal of cognitive effort (Malt et al., 1995; Murphy et al., 2012; Tversky & Kahneman, 1982). Therefore, there is a *symbolic bottleneck* in working memory — representations that were continuous and graded are forced into discrete symbolic states when they are passed to working memory. Dale and Spivey (2005) provide a formal model for how continuous and symbolic representations could map onto each other, using the mathematical framework of *symbolic dynamics*.

This distinction corresponds well with the claim by Oaksford and Chater (2010, 2012) that the representations used by Type 1 processes are probabilistic in nature, while those used by Type 2 processes are approximately symbolic. It also dovetails with recent experiments showing that probabilistic tasks that participants normally struggle with (e.g. Kahneman & Tversky, 1979; Murphy et al., 2012) can be solved easily when presented in a way that allows the inferences to be performed by perceptual, motor, and low-level cognitive systems (S. Y. Chen, Ross, & Murphy, 2013; Glaser, Trommershauser, Mamassian, & Maloney, 2012; Kording & Wolpert, 2004; Trommershäuser, Landy, & Maloney, 2006; Trommershäuser, Maloney, & Landy, 2008a, 2008b; Wu, Delgado, & Maloney, 2009) — or, in other words, by Type 1 processes. In a similar vein, there is an extensive literature showing that the human perceptual system routinely solves complex Bayesian inference problems (Kersten, Mamassian, & Yuille, 2004; Roach, Heron, & McGraw, 2006; Wolpert & Ghahramani, 2005) considerably more complex than those even highly numerate individuals struggle to solve explicitly (e.g. Barbey & Sloman, 2007; McNair & Feeney, 2013).

I predict that changes of mind should emerge in mouse tracking studies of processes where Type 2 processes, involving maintaining and manipulating information in working memory, are required to produce at least one of the possible re-

sponses. That being said, this is not the only situation that can give rise to changes of mind. Changes of mind have been demonstrated, for instance, when participants move initially to categorise pseudo-words as real words (Barca & Pezzulo, 2015), to categorise short-haired women or long-haired men as members of the opposite sex (in contexts where such styles are rare; Freeman, 2014), to respond that scalar implicatures (“some elephants are grey”) are true, before deciding them to be false (Tomlinson et al., 2013), to ignore the “not” in a negated sentence (Dale & Duran, 2011), as well as when switching between tasks (Hindly & Spivey, 2008). Of course, while arguments could be made that some of these tasks can be interpreted from a dual process perspective, it is also possible that these changes of mind arose from the interaction of two simple processes, activated one after another, or even from a single process accompanied by a shift in attention. My claim therefore is not that changes of mind are diagnostic of dual processes phenomena and the engagement of working memory, but that continuous attraction, and the absence of changes of mind, are indicative of tasks based only on Type 1 processes.

7.5.2 CONFLICT IN THE HAND/CONFLICT IN THE MIND

One possible criticism of my findings is that cognitive conflict may not always have a measurable effect on mouse cursor trajectories. For instance, the failure of these mouse tracking data to support the intuitive logic account may simply mean that although participants do implicitly experience conflict, this conflict does not have a detectable influence on their motor output.

However, a great deal of previous work has demonstrated that even subtle, implicit cognitive processes produce measurable effects in participants’ motor output. Simply seeing objects, for instance, primes us to perform the actions afforded by them (their *microaffordances*; Ellis & Tucker, 2000; Tucker & Ellis, 1998, 2004). Subtle priming influences have also been demonstrated using the mouse tracking paradigm, including on simple judgement tasks (Finkbeiner, Song, Nakayama, & Caramazza, 2008; Xiao & Yamauchi, 2014), spatial congruency effects (Tower-Richardi, Brunye, Gagnon, Mahoney, & Taylor, 2012) and even temporal congruency effects (i.e. the past is on the left; Miles, Betka, et al., 2010). These effects show that specific, actionable information passes from the simple cognitive and

perceptual processes to the motor system.

The question remains, however, as to why the current results are not in accord with the predictions of the intuitive logic theory. From previous work, there is much evidence that people are sensitive to conflict when giving heuristic responses during reasoning. These data are mostly epiphenomenal, in that they reflect the by-products of conflict in reasoning, such as ACC activation (De Neys et al., 2008), the galvanic skin response (De Neys & Glumicic, 2008), affective appraisal (Morsanyi & Handley, 2012), or metacognitive confidence (De Neys et al., 2011). Given the diversity of evidence for the intuitive logic model, however (see De Neys, 2012, for a review), I do not argue that the two null effects reported here serve to falsify the account. Why, then, when so much information passes continuously from cognitive and perceptual process to the motor system, does no evidence of the kind of conflict in reasoning predicted by an intuitive logic account feed into cursor trajectories on these tasks?

I cannot answer this question based on the evidence at hand. However, I would suggest a few possibilities. First, as I noted in Chapter 1, it may be the case that participants are implicitly aware that their responses and normative principles, as proposed by what I labelled the dual process conflict monitoring theory (De Neys & Glumicic, 2008; De Neys et al., 2008), but contrary to the intuitive logic accounts (De Neys, 2012, 2014), this is not because Type 1 processes are attempting to cue the correct response. At this point, however, it is not clear what other mechanisms could underlie these conflict detection effects.

Second, it could be that the kind of implicit, cognitive monitoring invoked by the intuitive logic theory (De Neys, 2012), as well as by conflict monitoring accounts more broadly (M. M. Botvinick et al., 2004; De Neys & Glumicic, 2008) — what Shea et al. (2014) label *Type 1* metacognition — does not feed information directly to the motor system, whereas non-metacognitive implicit processes of the kind discussed above do. Again, this raises further questions about the mechanisms that underlie logical intuition.

Thirdly, it is possible that conflict detection in reasoning, when it does occur, is not an entirely Type 1 processes. Rather, it may be mediated in some way by Type 2 processes — although not impinging enough on working memory that it is

affected by secondary load manipulations (Franssens & De Neys, 2009), or available to conscious introspection (De Neys & Glumicic, 2008). Above, I argued that Type 2 processes dependent on working memory impose a symbolic bottleneck on cognition. If this conflict detection relies on such processes, it may be that this bottleneck prevents it from reaching the motor system. Again, while I do not believe these questions can be resolved using the current data alone, this discussion highlights some of the value of analysing more than simply participants' ultimate responses.

7.5.3 MOUSE TRACKING, FAST AND SLOW

Since its inception (Spivey et al., 2005), the mouse tracking paradigm has for the most part been applied to the study of simple processes that unfold over at most two or three seconds. As discussed above, these tasks have largely yielded continuous attraction effects, as participants' cursor trajectories curve slightly more on conflict trials than no-conflict trials. Here, I have applied this method to tasks that typically take considerably longer than this to perform. Previous experiments, for example using the base rate neglect task as used here, minus the time pressure constraint, typically find response times of ~18 seconds on conflict trials (and ~14 seconds on control trials; De Neys & Franssens, 2009; De Neys & Glumicic, 2008; Franssens & De Neys, 2009). Similarly, while no work has collected response times for the inductive triad task pitting associative against structured knowledge (Bright & Feeney, submitted), I would assume that participants rarely responded within two or three seconds when completing the task normally. However, time pressure was necessary for the mouse tracking paradigm to work: in pilot studies where participants were not placed under time pressure on these tasks, they did not move the mouse until they had made their decision, and so the mouse data revealed little. Fortunately, the triad task pitting perceptual cues against perceptual knowledge (Chapter 3), being adapted from a task used with children, was relatively easy for participants to perform, and so they responded quickly, and moved the mouse while deciding, without the need for extrinsic time pressure.

As should be obvious to readers familiar with speeded response paradigms, requiring participants to reason under time pressure can influence the nature of the

reasoning they engage in. In particular, in the dual process literature, participants are often required to respond quickly in order to record their *Type 1* responses (e.g. De Neys, 2006; Markovits & Doyon, 2004; Thompson et al., 2011; Villejoubert, 2009). Similarly, Bright and Feeney (2014b, Experiment 1) demonstrated that the influence of structured knowledge in induction, but not of associative knowledge, was reduced when participants were placed under time pressure. Therefore, we can expect my data on these tasks to be less influenced by Type 2 processes, or structured knowledge, than would be the case if participants were not under time pressure. Fortunately, this does not affect my conclusions. In the induction tasks (Experiments 3 and 4), placing participants under time pressure may have rendered them less likely to draw on structured knowledge, and more likely to draw on associative knowledge. This possibly explains the poor structured knowledge displayed by participants in Experiment 3, as well as the fact that even on control trials participants did not universally give the structurally-cued response. However, while perhaps amplified by it, the key finding here can not be caused by time pressure alone: both associative and structured knowledge are drawn on in induction.

In the same vein, when descriptions and base rates conflict in Experiment 5 participants only gave the base rate-cued response 20% of the time, considerably less than the norm in such experiments (see Table 5.8). Similarly, in contrast to previous work, manipulating the base rate did not influence participants' response times when they gave the description-cued response, again contrary to previous findings (e.g. De Neys & Glumicic, 2008, see again Table 5.8). While all of this is consistent with a default-interventionist account, it differs from previous results in favour of the intuitive logic account. It is tempting to ascribe these differences to the speed of participants' responses in Experiment 5, which were even faster than those in previous speeded tests of the accounts (De Neys et al., 2008; Pennycook et al., 2014, Experiment 2). However, a core aspect of the intuitive logic theory is that this conflict occurs because both responses are cued by fast, automatic Type 1 processes, rather than requiring the engagement of slower Type 2 processes, as would be the case in the parallel-competitive theory (Sloman, 1996).

A few possibilities are consistent with this pattern of results. One is that, as dis-

cussed above, conflict detection and the cuing of the correct response do require the engagement of Type 2 processes, consistent with a parallel-competitive account. An alternative is that while Type 1 processes do cue the correct response, doing so still requires more time than is available here. However, Handley and Trippas (2015, see also Handley, Newstead, & Trippas, 2011; Pennycook et al., 2014) have recently argued that speed of processing does not provide a good means of differentiating between Type 1 and Type 2 processes, as some Type 1 processes can operate relatively slowly, and some Type 2 processes relatively quickly. For this reason, it is difficult to draw strong conclusions here.

7.5.4 FUTURE DIRECTIONS

At this point, I hope to have demonstrated some of the value of the mouse tracking paradigm adopted in this thesis. I will conclude this thesis by discussing what role mouse tracking may play in future research.

THE DEVELOPMENT OF REASONING

Although these experiments focused on adults' reasoning, children's inductive reasoning (Bright & Feeney, 2014a; S. A. Gelman & Davidson, 2013; Sloutsky et al., 2007), as well as their reasoning more broadly (Handley, Capon, Beveridge, Dennis, & Evans, 2004; O'Connor, McCormack, & Feeney, 2012; Steegen & De Neys, 2012), remains an active topic of research. In particular, Experiments 1 and 2, pitting perceptual similarity against conceptual knowledge, took their inspiration from developmental studies seeking to reveal which of these sources of information four-year-old children rely on (S. A. Gelman & Davidson, 2013; S. A. Gelman & Markman, 1986; Sloutsky et al., 2007). In this literature, it is typically taken for granted that adult reasoning is based on conceptual knowledge, usually on the basis of adult control groups that complete conflict problems, and rely on conceptual knowledge almost all of the time (e.g. S. A. Gelman & Davidson, 2013). Recall that while my (adult) participants did indeed give the conceptually-cued response the majority of the time on conflict trials, they did so significantly more often on no-conflict control trials, showing that they are also driven by perceptual cues.

An interesting next step here would be to adapt the mouse tracking paradigm to collect data with children. Of course, it remains to be seen if young children are capable of projecting their unfolding cognitive states onto the position of the mouse cursor in the way that came naturally to adult participants extremely adept at interacting with a computer in this way. One possibility we are currently exploring is the viability of low-cost motion tracking systems to tracking children's finger movements in pointing versions of this task. Such finger tracking has been used in the past in analyses almost identical to those used in mouse tracking studies (e.g. Song & Nakayama, 2006, 2008a, 2008b), and may provide a new window into children's reasoning.

INDIVIDUAL DIFFERENCES

There are, of course, individual differences in reasoning (see, e.g., De Neys, Schaeken, & d'Ydewalle, 2005; Feeney, Shafto, & Dunning, 2007; Frederick, 2005; Stanovich & West, 2000). Variability in participants' responses has been attributed to individual differences in general intelligence (Stanovich & West, 2000), working memory capacity (De Neys et al., 2005), personality traits (Cacioppo & Petty, 1982; Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012) including cognitive reflection (Frederick, 2005; Stanovich, 2009a, 2009b), inhibitory control (Bright & Feeney, submitted; Markovits & Doyon, 2004) and even differences in circadian cycles (i.e. *morning people* are more rational in the morning; Wieth & Zacks, 2011).

These individual differences were not the focus of this thesis, and the designs used rarely provided enough data from each participant to reliably estimate individual differences parameters. However, the novel data generated by the mouse tracking paradigm used here may prove useful in future individual differences work. In other words, it would be useful for future work to investigate individual differences in not only participants' responses, but also the more subtle analyses reported here. One particular candidate is the transition probability analysis. It would be informative, for instance, to discover what factors predict whether participants will initially move towards the foil response option (parameter $1 - \alpha$), whether they select a correct response after moving towards it (parameter β), and whether they will override initial movements towards a heuristically-appealing foil

(parameter γ). Although not directly used in this way here, this type of analysis would also be useful in evaluating frameworks for individual differences in reasoning discussed in Chapter 1.

In Chapter 1, I discussed three such frameworks for individual differences in reasoning. Stanovich and West (2008) constructed their framework by correlating individual differences in variety of predictive factors (fluid intelligence, formal knowledge, inhibitory control, thinking dispositions) with differences in participants' *responses* on reasoning tasks. De Neys and Bonnefon (2013) go somewhat further by considering the question of *when* these individual differences arise. The mouse tracking paradigm, and analyses of temporal dynamics more broadly, may provide a new means of investigating these "when" questions. Lastly, Pennycook, Fugelsang, and Koehler (2015) synthesise the theory and findings in support of the intuitive logic theory, discussed in Chapters 1, 5, and 6, many of which involve subtle behavioural and biological measures. Mouse tracking data, however, could go further than this. In particular, along with participants' responses and response latencies, the paradigm yields individual differences in the cursor parameters α , β , and γ , above. Clearly, much further work is required to establish a coherent account of how individual differences, and differences in task characteristics, relate to variance between participants in these parameters.

7.6 CONCLUSIONS

To recapitulate, in this thesis I used the mouse tracking paradigm to study conflict in reasoning. Using it to study inductive reasoning, I have shown that a) perceptual cues have a major influence on the early stages of inductive reasoning, with conceptual knowledge being retrieved and used later; b) perceptually-driven movements are more likely to be overridden by conceptual knowledge when participants reason about properties conceptually related to that knowledge; c) associative and structured knowledge also conflict during reasoning, but d) compared to perceptual cues, associative knowledge is less likely to drive early movements, but also less likely to be overridden when it does drive them, and I propose that e) the latter effect is the result of conflict between competing mental representations attempting to enter working memory, while the former may or may not be mediated

by working memory at all.

Likewise, studying reasoning from a dual process perspective, I have shown that f) participants reasoning based on stereotypical descriptions experience less conflict than those reasoning based on statistical base rates when the two cues conflict; g) descriptions influence the reasoning process both early on and throughout, while base rates are only influential later in reasoning; h) participants responding according to the base rate are affected by manipulations of the description, showing signs of conflict when the description disagrees with the base rate; i) participants responding according to the description were minimally affected by manipulations of the base rate — participants either ignored the base rate altogether, or attended to it, and gave the response it cued; j) heuristic responses on the CRT are approached more quickly than correct ones; k) heuristic responses that are also correct are not approached or chosen any more quickly than those that are incorrect; l) participants selecting the correct response option on the CRT are drawn towards the heuristic option before doing so; but, finally, m) participants selecting the heuristic option are not drawn towards the correct one.

7.6.1 CLOSING STATEMENT

This is a thesis on conflict in reasoning. Using the mouse tracking paradigm as a continuous measure of participants' attraction towards competing response options during reasoning, I have demonstrated where this conflict does, and does not occur. Beyond this, I have shown something of the nature of this conflict; it is a discrete, all-or-nothing effect, as a number of distinct cognitive processes — including both associative, autonomous Type 1 processes, and Type 2 processes dependent on working memory — vie for control over cognition, and over the motor system. Outside of the specific tasks I have studied, however, hopefully this thesis will provide a useful set of practical and conceptual tools for making sense of the processes that underlie human reasoning. Ultimately, while I hope I have helped to answer some questions here, my real hope for this thesis is that it encourages readers to ask, and to answer, new questions about how we reason about the world.

References

- Albert, A., & Anderson, J. A. (1984). On the Existence of Maximum Likelihood Estimates in Logistic Regression Models. *Biometrika*, 71(1), 1–10. doi: 10.2307/2336390
- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of memory and language*, 38(4), 419–439.
- Anderson, J. R. (1996). *The Architecture of Cognition*. Psychology Press.
- Anderson, J. R., & Fincham, J. M. (2014). Discovering the Sequential Structure of Thought. *Cognitive Science*, 38(2), 322–352. doi: 10.1111/cogs.12068
- Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, 8(4), 170–177. doi: 10.1016/j.tics.2004.02.010
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. doi: 10.1016/j.jml.2007.12.005
- Baddeley, A. D. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, 4(10), 829–839. doi: 10.1038/nrn1201
- Baddeley, A. D. (2007). *Working Memory, Thought, and Action*. Oxford, England: Oxford University Press.
- Badger, J. R., & Shapiro, L. R. (2012). Evidence of a transition from perceptual to category induction in 3- to 9-year-old children. *Journal of Experimental Child Psychology*, 113(1), 131–146. doi: 10.1016/j.jecp.2012.03.004

Ball, L. J. (2014). Eye-tracking and reasoning: What your eyes tell about your inferences. In W. De Neys & M. Osman (Eds.), *New Approaches in Reasoning Research*. Hove, UK: Psychology Press.

Ball, L. J., Lucas, E. J., Miles, J. N. V., & Gale, A. G. (2003). Inspection times and the selection task: What do eye-movements reveal about relevance effects? *The Quarterly Journal of Experimental Psychology Section A*, 56(6), 1053–1077. doi: 10.1080/02724980244000729

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44(3), 211–233. doi: 10.1016/0001-6918(80)90046-3

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences*, 30(03). doi: 10.1017/S0140525X07001653

Barca, L., & Pezzulo, G. (2015). Tracking Second Thoughts: Continuous and Discrete Revision Processes during Visual Lexical Decision. *PLoS ONE*, 10(2), e0116193. doi: 10.1371/journal.pone.0116193

Bargh, J. A. (2006). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology*, 36(2), 147.

Barr, D. J. (2008). Analyzing ‘visual world’ eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4), 457–474. doi: 10.1016/j.jml.2007.09.002

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. doi: 10.1016/j.jml.2012.11.001

Barsalou, L. W., Breazeal, C., & Smith, L. B. (2007). Cognition as coordinated non-cognition. *Cognitive Processing*, 8(2), 79–91. doi: 10.1007/s10339-007-0163-1

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious Mixed Models. *arXiv:1506.04967 [stat]*.

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi: 10.18637/jss.v067.i01
- Beer, R. D. (1995). On the dynamics of small continuous-time recurrent neural networks. *Adaptive Behavior*, 3(4), 469–509.
- Böckenholt, U. (2012). The Cognitive-Miser Response Model: Testing for Intuitive and Deliberate Reasoning. *Psychometrika*, 77(2), 388–399. doi: 10.1007/s11336-012-9251-y
- Botvinick, M., & Plaut, D. C. (2004). Doing without schema hierarchies: a recurrent connectionist approach to normal and impaired routine sequential action. *Psychological review*, 111(2), 395–429. doi: 10.1037/0033-295X.111.2.395
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624–652. doi: 10.1037/0033-295X.108.3.624
- Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences*, 8(12), 539–546. doi: 10.1016/j.tics.2004.10.003
- Bright, A. K., & Feeney, A. (2014a). Causal knowledge and the development of inductive reasoning. *Journal of Experimental Child Psychology*, 122, 48–61. doi: 10.1016/j.jecp.2013.11.015
- Bright, A. K., & Feeney, A. (2014b). The engine of thought is a hybrid: roles of associative and structured knowledge in reasoning. *Journal of Experimental Psychology. General*, 143(6), 2082–2102. doi: 10.1037/a0037653
- Bright, A. K., & Feeney, A. (submitted). Conflict resolution in selective category-based induction: The roles of knowledge inhibition and retrieval.
- Bruner, J. S. (1973). *Beyond the Information Given: Studies in the Psychology of Knowing*. W W Norton & Company Incorporated.
- Burgess, P. W., & Shallice, T. (1996). Response suppression, initiation and strategy use following frontal lobe lesions. *Neuropsychologia*, 34(4), 263–272. doi: 10.1016/0028-3932(95)00104-2

- Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton Tests*. Thurston, Suffolk, UK: Thames Valley Test Company.
- Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3), 432–459. doi: 10.1037/0033-295X.100.3.432
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116–131. doi: 10.1037/0022-3514.42.1.116
- Campitelli, G., & Gerrans, P. (2013). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition*, 42(3), 434–447. doi: 10.3758/s13421-013-0367-9
- Campitelli, G., & Labollita, M. (2010). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making*, 5(3), 182–191.
- Carey, S. (2009). *The origin of concepts*. Oxford University Press.
- Chaiken, S. (1987). The heuristic model of persuasion. In *Social influence: The ontario symposium* (Vol. 5, pp. 3–39).
- Chen, M. C., Anderson, J. R., & Sohn, M. H. (2001). What Can a Mouse Cursor Tell Us More?: Correlation of Eye/Mouse Movements on Web Browsing. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems* (pp. 281–282). New York, NY, USA: ACM. doi: 10.1145/634067.634234
- Chen, S. Y., Ross, B. H., & Murphy, G. L. (2013). Implicit and Explicit Processes in Category-Based Induction: Is Induction Best When We Don't Think? *Journal of Experimental Psychology: General*. doi: 10.1037/a0032064
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12(4), 335–359. doi: 10.1016/S0022-5371(73)80014-3
- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making*, 7(1), 25–47.

- Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, 4(1), 20–33.
- Coltheart, M. (2013). How Can Functional Neuroimaging Inform Cognitive Theories? *Perspectives on Psychological Science*, 8(1), 98–103. doi: 10.1177/1745691612469208
- Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In *Visual information processing* (pp. xiv, 555). Oxford, England: Academic.
- Cosmides, L., & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 85–116). New York, NY, US: Cambridge University Press.
- Cox, A. L., & Silva, M. M. (2006). The role of mouse movements in interactive search. In *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (pp. 1156–1161).
- Crisp-Bright, A. K. (2010). *Knowledge Selection in Category-Based Inductive Reasoning* (Unpublished doctoral dissertation). Durham University.
- Dale, R., & Duran, N. D. (2011). The Cognitive Dynamics of Negated Sentence Verification. *Cognitive Science*, 35(5), 983–996. doi: 10.1111/j.1551-6709.2010.01164.x
- Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, 35(1), 15–28. doi: 10.3758/BF03195938
- Dale, R., Roche, J., Snyder, K., & McCall, R. (2008). Exploring action dynamics as an index of paired-associate learning. *PLoS One*, 3(3), e1728.
- Dale, R., & Spivey, M. J. (2005). From apples and oranges to symbolic dynamics: a framework for conciliating notions of cognitive representation. *Journal of Experimental & Theoretical Artificial Intelligence*, 17(4), 317–342. doi: 10.1080/09528130500283766
- Darlow, A. L., & Sloman, S. A. (2010). Two systems of reasoning: architecture and relation to emotion. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(3), 382–392. doi: 10.1002/wcs.34

- De Neys, W. (2006). Automatic–heuristic and executive–analytic processing during reasoning: Chronometric and dual-task considerations. *The Quarterly Journal of Experimental Psychology*, 59(6), 1070–1100. doi: 10.1080/02724980543000123
- De Neys, W. (2012). Bias and Conflict: A Case for Logical Intuitions. *Perspectives on Psychological Science*, 7(1), 28–38. doi: 10.1177/1745691611429354
- De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning*, 20(2), 169–187. doi: 10.1080/13546783.2013.854725
- De Neys, W., & Bonnefon, J.-F. (2013). The ‘whys’ and ‘whens’ of individual differences in thinking biases. *Trends in Cognitive Sciences*, 17(4), 172–178. doi: 10.1016/j.tics.2013.02.001
- De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in Doubt: Conflict and Decision Confidence. *PLoS ONE*, 6(1), e15954. doi: 10.1371/journal.pone.0015954
- De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition*, 113(1), 45–61. doi: 10.1016/j.cognition.2009.07.009
- De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, 106(3), 1248–1299. doi: 10.1016/j.cognition.2007.06.002
- De Neys, W., Moyens, E., & Vansteenwegen, D. (2010). Feeling we’re biased: autonomic arousal and reasoning conflict. *Cognitive, affective & behavioral neuroscience*, 10(2), 208–216. doi: 10.3758/CABN.10.2.208
- De Neys, W., Rossi, S., & Houdé, O. (2013). Bats, balls, and substitution sensitivity: cognitive misers are no happy fools. *Psychonomic Bulletin & Review*, 20(2), 269–273. doi: 10.3758/s13423-013-0384-5
- De Neys, W., Schaeken, W., & d’Ydewalle, G. (2005). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning*, 11(4), 349–381.
- De Neys, W., Vartanian, O., & Goel, V. (2008). Smarter Than We Think When Our Brains Detect That We Are Biased. *Psychological Science*, 19(5), 483–489.

- Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, 64(1), 135–168. doi: 10.1146/annurev-psych-113011-143750
- Dietrich, E., & Markman, A. B. (2003). Discrete Thoughts: Why Cognition Must Use Discrete Representations. *Mind & Language*, 18(1), 95–119. doi: 10.1111/1468-0017.00216
- Donders, F. C. (1868/1969). On the speed of mental processes. *Acta Psychologica*, 30, 412–431. doi: 10.1016/0001-6918(69)90065-1
- Duchowski, A. (2007). *Eye Tracking Methodology: Theory and Practice*. Springer Science & Business Media.
- Duran, N. D., Dale, R., & McNamara, D. S. (2010). The action dynamics of overcoming the truth. *Psychonomic Bulletin & Review*, 17(4), 486–491. doi: 10.3758/PBR.17.4.486
- Ellis, R., & Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British journal of psychology*, 91(4), 451–471.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215–251. doi: 10.1037/0033-295X.87.3.215
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16(1), 143–149. doi: 10.3758/BF03203267
- Evans, J. S. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, 75(4), 451–468. doi: 10.1111/j.2044-8295.1984.tb01915.x
- Evans, J. S. B. T. (1996). Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology*, 87(2), 223–240. doi: 10.1111/j.2044-8295.1996.tb02587.x
- Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, 13(3), 378–395. doi: 10.3758/BF03193858
- Evans, J. S. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning*, 13(4), 321–339. doi: 10.1080/13546780601008825

- Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, 59(1), 255–278. doi: 10.1146/annurev.psych.59.103006.093629
- Evans, J. S. B. T., & Ball, L. J. (2010). Do people reason on the Wason selection task? A new look at the data of Ball et al. (2003). *The Quarterly Journal of Experimental Psychology*, 63(3), 434–441. doi: 10.1080/17470210903398147
- Evans, J. S. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, 11(3), 295–306. doi: 10.3758/BF03196976
- Evans, J. S. B. T., & Elqayam, S. (2007). Dual-processing explains base-rate neglect, but which dual-process theory and how? *Behavioral and Brain Sciences*, 30(03), 261–262. doi: 10.1017/S0140525X07001720
- Evans, J. S. B. T., & Over, D. (2004). *IF*. Oxford; New York: Oxford University Press.
- Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and Reasoning*. Psychology Press.
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, 8(3), 223–241. doi: 10.1177/1745691612460685
- Falke, J., Winter, B., & Spivey, M. J. (2013). Candle-Candle-Candle-Candy: Continuous Attraction Toward Previously Seen Phonological Competitors. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society*.
- Farmer, T. A., Cargill, S. A., Hindy, N. C., Dale, R., & Spivey, M. J. (2007). Tracking the continuity of language comprehension: Computer mouse trajectories suggest parallel syntactic processing. *Cognitive Science*, 31(5), 889–909.
- Feeney, A., & Heit, E. (2007). *Inductive Reasoning: Experimental, Developmental, and Computational Approaches*. Cambridge University Press.
- Feeney, A., Shafto, P., & Dunning, D. (2007). Who is susceptible to conjunction fallacies in category-based induction? *Psychonomic bulletin & review*, 14(5), 884–889.

- Fernbach, P. M., Rogers, T., Fox, C. R., & Sloman, S. A. (2013). Political Extremism Is Supported by an Illusion of Understanding. *Psychological Science*, 24(6), 939–946. doi: 10.1177/0956797612464058
- Fiedler, S., & Glockner, A. (2012). The Dynamics of Decision Making in Risky Choice: An Eye-Tracking Analysis. *Frontiers in Psychology*, 3. doi: 10.3389/fpsyg.2012.00335
- Figner, B., & Murphy, R. O. (2010). Using skin conductance in judgment and decision making research. *A handbook of process tracing methods for decision research: A critical review and user's guide*, 163–184.
- Finkbeiner, M., Song, J.-H., Nakayama, K., & Caramazza, A. (2008). Engaging the motor system with masked orthographic primes: A kinematic analysis. *Visual Cognition*, 16(1), 11–22. doi: 10.1080/13506280701203838
- Fisher, A. V., Godwin, K. E., & Matlen, B. J. (2015). Development of inductive generalization with familiar categories. *Psychonomic Bulletin & Review*, 1–25. doi: 10.3758/s13423-015-0816-5
- Fiske, S. T., & Taylor, S. E. (1991). *Social Cognition*. McGraw-Hill.
- Fodor, J. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. A Bradford Book / MIT Press.
- Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking & Reasoning*, 15(2), 105–128. doi: 10.1080/13546780802711185
- Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, 19(4), 25–42. doi: 10.1257/089533005775196732
- Freeman, J. B. (2014). Abrupt category shifts during real-time person perception. *Psychonomic Bulletin & Review*, 21(1), 85–92. doi: 10.3758/s13423-013-0470-8
- Freeman, J. B., & Ambady, N. (2009). Motions of the hand expose the partial and parallel activation of stereotypes. *Psychological Science*, 20(10), 1183–1188.
- Freeman, J. B., & Ambady, N. (2011a). A dynamic interactive theory of person construal. *Psychological Review*, 118(2), 247–279. doi: 10.1037/a0022327
- Freeman, J. B., & Ambady, N. (2011b). Hand movements reveal the time-course of shape and pigmentation processing in face categorization. *Psychonomic Bulletin & Review*, 18(4), 705–712. doi: 10.3758/s13423-011-0097-6

- Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General*, 137(4), 673–690. doi: 10.1037/a0013875
- Freeman, J. B., & Dale, R. (2012). Assessing bimodality to detect the presence of a dual cognitive process. *Behavior Research Methods*, 45(1), 83–97. doi: 10.3758/s13428-012-0225-x
- Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in Motion Reveals Mind in Motion. *Frontiers in Psychology*, 2. doi: 10.3389/fpsyg.2011.00059
- Freeman, J. B., Pauker, K., Apfelbaum, E. P., & Ambady, N. (2010). Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology*, 46(1), 179–185. doi: 10.1016/j.jesp.2009.10.002
- Freeman, J. B., Stolier, R. M., Ingbretnsen, Z. A., & Hehman, E. A. (2014). Amygdala Responsivity to High-Level Social Information from Unseen Faces. *The Journal of Neuroscience*, 34(32), 10573–10581. doi: 10.1523/JNEUROSCI.5063-13.2014
- French, R. M., Mareschal, D., Mermilliod, M., & Quinn, P. C. (2004). The Role of Bottom-Up Processing in Perceptual Categorization by 3- to 4-Month-Old Infants: Simulations and Data. *Journal of Experimental Psychology: General*, 133(3), 382–397. doi: 10.1037/0096-3445.133.3.382
- Gangemi, A., Bourgeois-Gironde, S., & Mancini, F. (2015). Feelings of error in reasoning—in search of a phenomenon. *Thinking & Reasoning*, 21(4), 383–396. doi: 10.1080/13546783.2014.980755
- Garrett, H. E. (1922). A study of the relation of accuracy to speed. *Archives of Psychology*, 8(56), 104.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis*, 1(3), 515–534.
- Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multi-level/Hierarchical Models*. Cambridge University Press.
- Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, 8(9), 404–409. doi: 10.1016/j.tics.2004.07.001

- Gelman, S. A., & Davidson, N. S. (2013). Conceptual influences on category-based induction. *Cognitive Psychology*, 66(3), 327–353. doi: 10.1016/j.cogpsych.2013.02.001
- Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, 23(3), 183–209. doi: 10.1016/0010-0277(86)90034-X
- Gelman, S. A., & Meyer, M. (2011). Child categorization. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(1), 95–105. doi: 10.1002/wcs.96
- Gelman, S. A., & Waxman, S. R. (2007). Looking Beyond Looks Comments on Slout-sky, Kloos, and Fisher (2007). *Psychological Science*, 18(6), 554–555.
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, 62(1), 451–482. doi: 10.1146/annurev-psych-120709-145346
- Gigerenzer, G., & Regier, T. (1996). How do we tell an association from a rule? Comment on Sloman (1996). *Psychological Bulletin*, 119(1), 23–26. doi: 10.1037/0033-2909.119.1.23
- Gilbert, D. T. (1991). How mental systems believe. *American psychologist*, 46(2), 107.
- Gilbert, D. T. (1999). What the Mind's Not. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology*. New York: Guilford.
- Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press.
- Glaser, C., Trommershauser, J., Mamassian, P., & Maloney, L. T. (2012). Comparison of the Distortion of Probability Information in Decision Under Risk and an Equivalent Visual Task. *Psychological Science*, 23(4), 419–426. doi: 10.1177/0956797611429798
- Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: an adaptive network model. *Journal of Experimental Psychology: General*, 117(3), 227.
- Goodale, M. A., Westwood, D. A., & Milner, A. D. (2004). Two distinct modes of control for object-directed action. *Progress in brain research*, 144, 131–144.
- Gopnik, A. (2003). The Theory Theory as an Alternative to the Innateness Hypothesis. In L. M. Antony & N. Hornstein (Eds.), *Chomsky and His Critics* (pp. 238–254). Blackwell Publishing Ltd.

- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4–27. doi: 10.1037/0033-295X.102.1.4
- Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, 51(4), 334–384. doi: 10.1016/j.cogpsych.2005.05.004
- Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, 116(4), 661–716. doi: 10.1037/a0017201
- Haigh, M., Ferguson, H. J., & Stewart, A. J. (2014). An eye-tracking investigation into readers' sensitivity to actual versus expected utility in the comprehension of conditionals. *The Quarterly Journal of Experimental Psychology*, 67(1), 166–185. doi: 10.1080/17470218.2013.797475
- Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2), 413–497. doi: 10.1103/RevModPhys.65.413
- Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. S. B. T. (2004). Working memory, inhibitory control and the development of children's reasoning. *Thinking & Reasoning*, 10(2), 175–195. doi: 10.1080/13546780442000051
- Handley, S. J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 28.
- Handley, S. J., & Trippas, D. (2015). Dual Processes and the Interplay between Knowledge and Structure: A New Parallel Processing Model. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 62, pp. 33–58). Academic Press.
- Hanley, M., Riby, D. M., McCormack, T., Carty, C., Coyle, L., Crozier, N., ... McPhillips, M. (2014). Attention during social interaction in children with autism: Comparison to specific language impairment, typical development, and links to social cognition. *Research in Autism Spectrum Disorders*, 8(7), 908–924. doi: 10.1016/j.rasd.2014.03.020
- Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, 70–84.

- Hawkins, G. E., Hayes, B. K., & Heit, E. (2015). A dynamic model of reasoning and memory. *Journal of Experimental Psychology. General*. doi: 10.1037/xge0000113
- Hayes, B. K., Heit, E., & Swendsen, H. (2010). Inductive reasoning. *Wiley Interdisciplinary Reviews: Cognitive Science*, n/a–n/a. doi: 10.1002/wcs.44
- Hehman, E., Carpinella, C. M., Johnson, K. L., Leitner, J. B., & Freeman, J. B. (2014). Early Processing of Gendered Facial Cues Predicts the Electoral Success of Female Politicians. *Social Psychological and Personality Science*, 1948550614534701. doi: 10.1177/1948550614534701
- Hehman, E., Ingbretnsen, Z. A., & Freeman, J. B. (2014). The neural basis of stereotypical impact on multiple social categorization. *NeuroImage*, 101, 704–711. doi: 10.1016/j.neuroimage.2014.07.056
- Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. *Rational models of cognition*, 248–274.
- Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(2), 411–422. doi: 10.1037/0278-7393.20.2.411
- Hindly, N. C., & Spivey, M. J. (2008). Motor dynamics of task switching. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society, Washington, DC, July 23–26, 2008* (pp. 2474–2479).
- Hinton, G. E., & Anderson, J. A. (2014). *Parallel Models of Associative Memory: Updated Edition*. Psychology Press.
- Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18(7), 1527–1554. doi: 10.1162/neco.2006.18.7.1527
- Huang, J., White, R. W., & Dumais, S. (2011). No Clicks, No Problem: Using Cursor Movements to Understand and Improve Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1225–1234). New York, NY, USA: ACM. doi: 10.1145/1978942.1979125
- Huette, S., & Spivey, M. (2012). Fuzzy Consciousness. In S. Edelman, T. Fekete, & N. Zach (Eds.), *Being in Time: Dynamical Models of Phenomenal Experience* (pp. 88–149). John Benjamins Pub. Co.

- Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional magnetic resonance imaging* (Vol. 1). Sinauer Associates Sunderland.
- Huetting, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta psychologica*, 137(2), 151–171.
- Huey, E. B. (1908). *The Psychology and Pedagogy of Reading*. Oxford, England: Macmillan.
- Jackson, R. L., Hoffman, P., Pobric, G., & Ralph, M. A. L. (2015). The Nature and Neural Correlates of Semantic Association versus Conceptual Similarity. *Cerebral Cortex*, bhv003. doi: 10.1093/cercor/bhv003
- Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Harvard University Press.
- Johnson-Laird, P. N., & Byrne, R. M. (1991). *Deduction*. Lawrence Erlbaum Associates, Inc.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY, US: Farrar, Straus and Giroux.
- Kahneman, D., & Beatty, J. (1966). Pupil Diameter and Load on Memory. *Science*, 154(3756), 1583–1585. doi: 10.1126/science.154.3756.1583
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. D. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York, NY, US: Cambridge University Press.
- Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267–293). New York, NY, US: Cambridge University Pres.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237–251. doi: 10.1037/h0034747

- Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263. doi: 10.2307/1914185
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience*, 17(11), 4302–4311.
- Kemp, C., & Jern, A. (2014). A taxonomy of inductive problems. *Psychonomic Bulletin & Review*, 21(1), 23–46. doi: 10.3758/s13423-013-0467-3
- Kemp, C., Shafto, P., & Tenenbaum, J. B. (2012). An integrated account of generalization across objects and features. *Cognitive Psychology*, 64(1-2), 35–73. doi: 10.1016/j.cogpsych.2011.10.001
- Kemp, C., & Tenenbaum, J. B. (2003). Theory-based induction. In *Proceedings of the 25th annual conference of the cognitive science society* (pp. 658–663).
- Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological review*, 116(1), 20.
- Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annu. Rev. Psychol.*, 55, 271–304.
- Kieslich, P. J., & Hilbig, B. E. (2014). Cognitive conflict in social dilemmas: An analysis of response dynamics. *Judgment and Decision Making*, 9(6), 510–522.
- Klaczynski, P. A. (2000). Motivated Scientific Reasoning Biases, Epistemological Beliefs, and Theory Polarization: A Two-Process Approach to Adolescent Cognition. *Child Development*, 71(5), 1347–1366. doi: 10.1111/1467-8624.00232
- Klaczynski, P. A., & Cottrell, J. M. (2004). A dual-process approach to cognitive development: The case of children's understanding of sunk cost decisions. *Thinking & Reasoning*, 10(2), 147–174. doi: 10.1080/13546780442000042
- Klauer, K. C., & Singmann, H. (2013). Does logic feel good? Testing for intuitive detection of logicality in syllogistic reasoning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 39(4), 1265–1273. doi: 10.1037/a0030530
- Koehler, D. J., & James, G. (2009). Probability matching in choice under uncertainty: Intuition versus deliberation. *Cognition*, 113(1), 123–127. doi: 10.1016/j.cognition.2009.07.003

- Koop, G. J., & Johnson, J. G. (2011). Response dynamics: A new window on the decision process. *Judgment and Decision Making*, 6(8), 750–758.
- Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive Psychology*, 67(4), 151–185. doi: 10.1016/j.cogpsych.2013.09.001
- Kording, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, 427(6971), 244–247.
- 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. doi: 10.1073/pnas.1101328108
- Kruglanski, A. W. (2013). Only one? The default interventionist perspective as a unimodel—Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science*, 8(3), 242–247.
- Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review*, 118(1), 97–109. doi: 10.1037/a0020762
- Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological review*, 99(1), 22–44.
- Kruschke, J. K. (2011). Bayesian Assessment of Null Values Via Parameter Estimation and Model Comparison. *Perspectives on Psychological Science*, 6(3), 299–312. doi: 10.1177/1745691611406925
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). *lmerTest: Tests in Linear Mixed Effects Models*. (R package version 2.0-29)
- Lappin, J. S., & Eriksen, C. W. (1966). Use of a delayed signal to stop a visual reaction-time response. *Journal of Experimental Psychology*, 72(6), 805–811. doi: 10.1037/h0021266
- LaRiccia, M. V. (2005). *Black mane*. Suffolk, UK: One Time Press. (LaRiccia, Michael V. (2005). Black mane. Suffolk, UK: One Time Press.)
- Legrenzi, P., Girotto, V., & Johnson-Laird, P. N. (1993). Focussing in reasoning and decision making. *Cognition*, 49(1–2), 37–66. doi: 10.1016/0010-0277(93)90035-T

Leite, F. P., & Ratcliff, R. (2010). Modeling reaction time and accuracy of multiple-alternative decisions. *Attention, Perception, & Psychophysics*, 72(1), 246–273. doi: 10.3758/APP.72.1.246

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual Differences in Numeracy and Cognitive Reflection, with Implications for Biases and Fallacies in Probability Judgment. *Journal of Behavioral Decision Making*, 25(4), 361–381. doi: 10.1002/bdm.752

Malt, B. C., Ross, B. H., & Murphy, G. L. (1995). Predicting features for members of natural categories when categorization is uncertain. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(3), 646.

Marghetis, T., Núñez, R., & Bergen, B. K. (2014). Doing arithmetic by hand: Hand movements during exact arithmetic reveal systematic, dynamic spatial processing. *The Quarterly Journal of Experimental Psychology*, 67(8), 1579–1596. doi: 10.1080/17470218.2014.897359

Markovits, H., & Doyon, C. (2004). Information processing and reasoning with premises that are empirically false: Interference, working memory, and processing speed. *Memory & cognition*, 32, 592–601.

Mata, A., Fiedler, K., Ferreira, M. B., & Almeida, T. (2013). Reasoning about others' reasoning. *Journal of Experimental Social Psychology*, 49(3), 486–491. doi: 10.1016/j.jesp.2013.01.010

Mathôt, S., & Theeuwes, J. (2011). Mantra: An open method for object and movement tracking. *Behavior research methods*, 43(4), 1182–1193.

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, 7(7), 293–299. doi: 10.1016/S1364-6613(03)00134-7

McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). Interactive activation and mutual constraint satisfaction in perception and cognition. *Cognitive Science*, 38(6), 1139–1189. doi: 10.1111/cogs.12146

McClelland, J. L., Rumelhart, D. E., & PDP Research Group, C. (Eds.). (1986). *Parallel Distributed Processing: Explorations in the Microstructure, Vol. 1: Foundations*. Cambridge, MA, USA: MIT Press.

- McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, 19(1), 22–24.
- McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance*, 34(6), 1609.
- McNair, S., & Feeney, A. (2013). When does information about causal structure improve statistical reasoning? *The Quarterly Journal of Experimental Psychology*, 1–21. doi: 10.1080/17470218.2013.821709
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100(2), 254–278. doi: 10.1037/0033-295X.100.2.254
- Mele, M. L., & Federici, S. (2012). Gaze and eye-tracking solutions for psychological research. *Cognitive Processing*, 13, S261–S265. (WOS:000308824700295) doi: 10.1007/s10339-012-0499-z
- Mevel, K., Poirel, N., Rossi, S., Cassotti, M., Simon, G., Houdé, O., & De Neys, W. (2014). Bias detection: Response confidence evidence for conflict sensitivity in the ratio bias task. *Journal of Cognitive Psychology*, 27(ahead-of-print), 1–11. doi: 10.1080/20445911.2014.986487
- Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology*, 26(1–3), 3–67. doi: 10.1016/0301-0511(88)90013-0
- Miles, L. K., Betka, E., Pendry, L. F., & Macrae, C. N. (2010). Mapping temporal constructs: Actions reveal that time is a place. *The Quarterly Journal of Experimental Psychology*, 63(11), 2113–2119. doi: 10.1080/17470218.2010.524932
- Miles, L. K., Karpinska, K., Lumsden, J., & Macrae, C. N. (2010). The meandering mind: Vection and mental time travel. *PLoS One*, 5(5), e10825.
- Mill, J. S. (1856). *A System of Logic, Ratiocinative and Inductive*. Parker.
- Miller, E. K., & Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annual Review of Neuroscience*, 24(1), 167–202. doi: 10.1146/annurev.neuro.24.1.167

- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97. doi: 10.1037/h0043158
- Mirman, D. (2014a). *Growth Curve Analysis and Visualization Using R*. CRC Press.
- Mirman, D. (2014b). *Growth Curve Analysis and Visualization Using R*. Taylor & Francis.
- Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59(4), 475–494. doi: 10.1016/j.jml.2007.11.006
- Mishra, A., & Mishra, H. (2010). Border Bias: The Belief That State Borders Can Protect Against Disasters. *Psychological Science*, 21(11), 1582–1586. doi: 10.1177/0956797610385950
- Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing Atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, 7(3), 134–140. doi: 10.1016/S1364-6613(03)00028-7
- Morsanyi, K., Busdraghi, C., & Primi, C. (2014). Mathematical anxiety is linked to reduced cognitive reflection: a potential road from discomfort in the mathematics classroom to susceptibility to biases. *Behavioral and Brain Functions*, 10(1), 31.
- Morsanyi, K., & Handley, S. J. (2012). Logic feels so good—I like it! Evidence for intuitive detection of logicality in syllogistic reasoning. *Journal of experimental psychology. Learning, memory, and cognition*, 38(3), 596–616. doi: 10.1037/a0026099
- Murphy, G. L. (2004). *The Big Book of Concepts*. MIT Press.
- Murphy, G. L., Chen, S. Y., & Ross, B. H. (2012). Reasoning with uncertain categories. *Thinking & Reasoning*, 18(1), 81–117. doi: 10.1080/13546783.2011.650506
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92(3), 289–316. doi: 10.1037/0033-295X.92.3.289

- Murphy, G. L., & Ross, B. H. (1994). Predictions from uncertain categorizations. *Cognitive psychology*, 27(2), 148–193. doi: 10.1006/cogp.1994.1015
- Murphy, G. L., & Ross, B. H. (2010). Uncertainty in category-based induction: When do people integrate across categories? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(2), 263–276. doi: 10.1037/a0018685
- Mynatt, C. R., Doherty, M. E., & Dragan, W. (1993). Information relevance, working memory, and the consideration of alternatives. *The Quarterly Journal of Experimental Psychology Section A*, 46(4), 759–778. doi: 10.1080/14640749308401038
- Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. xiv). Oxford, England: Prentice-Hall.
- Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences*, 38(01), 1–19. doi: 10.1017/S0140525X12003214
- Niedermeyer, E., & da Silva, F. H. L. (2005). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231–259. doi: 10.1037/0033-295X.84.3.231
- Oakes, P. J., Alexander, S., & Turner, J. C. (1994). *Stereotyping and social reality* (Vol. xiv). Malden: Blackwell Publishing.
- Oaksford, M., & Chater, N. (2010). Conditional inference and constraint satisfaction: Reconciling mental models and the probabilistic approach. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thinking* (pp. 309–333). Oxford, England: Oxford University Press.
- Oaksford, M., & Chater, N. (2012). Dual processes, probabilities, and cognitive architecture. *Mind & Society*, 11(1), 15–26. doi: 10.1007/s11299-011-0096-3
- O'Connor, E., McCormack, T., & Feeney, A. (2012). The development of regret. *Journal of Experimental Child Psychology*, 111(1), 120–127. doi: 10.1016/j.jecp.2011.07.002

- Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, 72(1), 147–152. doi: 10.1016/j.jebo.2009.04.018
- O'Hora, D., Dale, R., Piiroinen, P. T., & Connolly, F. (2013). Local dynamics in decision making: The evolution of preference within and across decisions. *Scientific Reports*, 3, 2210. doi: 10.1038/srep02210
- Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological review*, 97(2), 185.
- Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, 11(6), 988–1010.
- Papesh, M. H., & Goldinger, S. D. (2012). Memory in motion: Movement dynamics reveal memory strength. *Psychonomic Bulletin & Review*, 19(5), 906–913. doi: 10.3758/s13423-012-0281-3
- Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2015). On the reception and detection of pseudo-profound bullshit. *Judgment and Decision Making*, 10(6), 549–563.
- Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition*, 123(3), 335–346. doi: 10.1016/j.cognition.2012.03.003
- Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2012). Are we good at detecting conflict during reasoning? *Cognition*, 124(1), 101–106. doi: 10.1016/j.cognition.2012.04.004
- Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, 80, 34–72. doi: 10.1016/j.cogpsych.2015.05.001
- Pennycook, G., & Thompson, V. A. (2012). Reasoning with base rates is routine, relatively effortless, and context dependent. *Psychonomic bulletin & review*, 19(3), 528–534. doi: 10.3758/s13423-012-0249-3

- Pennycook, G., Trippas, D., Handley, S. J., & Thompson, V. A. (2014). Base rates: Both neglected and intuitive. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(2), 544–554. doi: 10.1037/a0034887
- Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 19, pp. 123–205). Academic Press.
- Pfister, R., Schwarz, K. A., Janczyk, M., Dale, R., & Freeman, J. B. (2013). Good things peak in pairs: a note on the bimodality coefficient. *Frontiers in Psychology*, 4. doi: 10.3389/fpsyg.2013.00700
- Port, R. F., & Gelder, T. V. (1995). *Mind as Motion: Explorations in the Dynamics of Cognition*. MIT Press.
- Posner, M. I. (1978). *Chronometric explorations of mind* (Vol. xiii). Oxford, England: Lawrence Erlbaum.
- Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2015). The Development and Testing of a New Version of the Cognitive Reflection Test Applying Item Response Theory (IRT). *Journal of Behavioral Decision Making*, n/a–n/a. doi: 10.1002/bdm.1883
- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59(4), 413–425. doi: 10.1016/j.jml.2008.02.002
- Quinton, J.-C., Volpi, N., Barca, L., & Pezzulo, G. (2013). The Cat Is On the Mat. Or Is It a Dog? Dynamic Competition in Perceptual Decision Making. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, Early Access Online*. doi: 10.1109/TSMC.2013.2279664
- R Core Team. (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ranyard, R., & Ola, S. (2010). Verbal Data and Decision Protocol Analysis. In M. Schulte-Mecklenbeck, A. Kuehberger, & R. Ranyard (Eds.), *A Handbook of Process Tracing Methods for Decision Research: A Critical Review and User's Guide*. Psychology Press.

- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, 85(2), 59.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, 20(4), 873–922. doi: 10.1162/neco.2008.12-06-420
- Ratcliff, R., & Smith, P. L. (2010). Perceptual discrimination in static and dynamic noise: The temporal relation between perceptual encoding and decision making. *Journal of Experimental Psychology: General*, 139(1), 70–94. doi: 10.1037/a0018128
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422.
- Rescorla, R. A., & Furrow, D. R. (1977). Stimulus similarity as a determinant of Pavlovian conditioning. *Journal of Experimental Psychology. Animal Behavior Processes*, 3(3), 203–215.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (Vol. 2, pp. 64–99). Appleton-Century-Crofts.
- Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, 461(7261), 263–266. doi: 10.1038/nature08275
- Rhodes, M., & Gelman, S. A. (2009). Five-year-olds' beliefs about the discreteness of category boundaries for animals and artifacts. *Psychonomic Bulletin & Review*, 16(5), 920–924. doi: 10.3758/PBR.16.5.920
- Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: looking at things that aren't there anymore. *Cognition*, 76(3), 269–295.
- Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American journal of physiology. Heart and circulatory physiology*, 278(6), H2039–2049.
- Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 665–681. doi: 10.1016/S0022-5371(75)80055-7
- Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proceedings*

of the Royal Society of London B: Biological Sciences, 273(1598), 2159–2168. doi: 10.1098/rspb.2006.3578

Rodden, K., Fu, X., Aula, A., & Spiro, I. (2008). Eye-mouse Coordination Patterns on Web Search Results Pages. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems* (pp. 2997–3002). New York, NY, USA: ACM. doi: 10.1145/1358628.1358797

Rogers, T. T., & McClelland, J. L. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach*. MIT Press.

Rogers, T. T., & McClelland, J. L. (2008). Précis of Semantic Cognition: A Parallel Distributed Processing Approach. *Behavioral and Brain Sciences*, 31(06), 689. doi: 10.1017/S0140525X0800589X

Rosch, E. (1988). Principles of Categorization. In A. Collins & E. E. Smith (Eds.), *Readings in Cognitive Science, a Perspective From Psychology and Artificial Intelligence* (pp. 312–22). Morgan Kaufmann Publishers.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3), 382–439. doi: 10.1016/0010-0285(76)90013-X

Ross, B., & Murphy, G. (1999). Food for Thought: Cross-Classification and Category Organization in a Complex Real-World Domain. *Cognitive Psychology*, 38, 495–553.

Rumelhart, D. E. (1986). *Parallel distributed processing: explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

Salti, M., Monto, S., Charles, L., King, J.-R., Parkkonen, L., & Dehaene, S. (2015). Distinct cortical codes and temporal dynamics for conscious and unconscious percepts. *eLife*, e05652. doi: 10.7554/eLife.05652

SAS Institute. (1990). *SAS/STAT user's guide: version 6* (Vol. 2). Sas Inst.

Satterthwaite, F. E. (1946). An Approximate Distribution of Estimates of Variance Components. *Biometrics Bulletin*, 2(6), 110–114. doi: 10.2307/3002019

Scherbaum, S., Dshemuchadse, M., Fischer, R., & Goschke, T. (2010). How decisions evolve: The temporal dynamics of action selection. *Cognition*, 115(3), 407–416. doi: 10.1016/j.cognition.2010.02.004

- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84(1), 1–66. doi: 10.1037/0033-295X.84.1.1
- Schulte-Mecklenbeck, M., Kühberger, A., & Ranyard, R. (2011). The role of process data in the development and testing of process models of judgment and decision making. *Judgment and Decision Making*, 6(8), 733–739.
- Shafto, P., Coley, J. D., & Baldwin, D. (2007). Effects of time pressure on context-sensitive property induction. *Psychonomic bulletin & review*, 14(5), 890–894.
- Shafto, P., Kemp, C., Baraff, E., Coley, J. D., & Tenenbaum, J. B. (2005). Context-sensitive induction. In *Proceedings of the 27th annual conference of the cognitive science society* (pp. 2003–2008).
- Shafto, P., Kemp, C., Bonawitz, E. B., Coley, J. D., & Tenenbaum, J. B. (2008). Inductive reasoning about causally transmitted properties. *Cognition*, 109(2), 175–192. doi: 10.1016/j.cognition.2008.07.006
- Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*, 18(4), 186–193. doi: 10.1016/j.tics.2014.01.006
- Shenhav, A., Rand, D. G., & Greene, J. D. (2012). Divine intuition: Cognitive style influences belief in God. *Journal of Experimental Psychology: General*, 141(3), 423–428. doi: 10.1037/a0025391
- Shipley, E. F. (1993). Categories, hierarchies, and induction. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 30, pp. 265–301). San Diego, CA: Academic Press.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, 63(2), 129.
- Simon, J. R., & Wolf, J. D. (1963). Choice reaction time as a function of angular stimulus-response correspondence and age. *Ergonomics*, 6(1), 99–105.
- Sloman, S. A. (1993). Feature-based induction. *Cognitive psychology*, 25(2), 231–280.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological bulletin*, 119(1), 3. doi: 10.1037/0033-295X.119.1.3

- Sloman, S. A. (2002). Two systems of reasoning. In T. D. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 379–396). New York, NY, US: Cambridge University Press.
- Sloman, S. A. (2014a). Comments on Quantum Probability Theory. *Topics in Cognitive Science*, 6(1), 47–52. doi: 10.1111/tops.12072
- Sloman, S. A. (2014b). Two Systems of Reasoning: An Update. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the Social Mind* (pp. 69–79). New York, NY, US: Guilford Press.
- Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, 7(6), 246–251. doi: 10.1016/S1364-6613(03)00109-8
- Sloutsky, V. M. (2010). From Perceptual Categories to Concepts: What Develops? *Cognitive Science*, 34(7), 1244–1286. doi: 10.1111/j.1551-6709.2010.01129.x
- Sloutsky, V. M., Deng, W. S., Fisher, A. V., & Kloos, H. (2015). Conceptual influences on induction: A case for a late onset. *Cognitive Psychology*, 82, 1–31. doi: 10.1016/j.cogpsych.2015.08.005
- Sloutsky, V. M., & Fisher, A. V. (2004a). Induction and categorization in young children: a similarity-based model. *Journal of Experimental Psychology: General*, 133(2), 166.
- Sloutsky, V. M., & Fisher, A. V. (2004b). When Development and Learning Decrease Memory Evidence Against Category-Based Induction in Children. *Psychological Science*, 15(8), 553–558.
- Sloutsky, V. M., & Fisher, A. V. (2008). Attentional learning and flexible induction: How mundane mechanisms give rise to smart behaviors. *Child Development*, 79(3), 639–651.
- Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007). When Looks Are Everything Appearance Similarity Versus Kind Information in Early Induction. *Psychological Science*, 18(2), 179–185.
- Song, J.-H., & Nakayama, K. (2006). Role of focal attention on latencies and trajectories of visually guided manual pointing. *Journal of Vision*, 6(9).

- Song, J.-H., & Nakayama, K. (2008a). Numeric comparison in a visually-guided manual reaching task. *Cognition*, 106(2), 994–1003. doi: 10.1016/j.cognition.2007.03.014
- Song, J.-H., & Nakayama, K. (2008b). Target selection in visual search as revealed by movement trajectories. *Vision Research*, 48(7), 853–861. doi: 10.1016/j.visres.2007.12.015
- Spivey, M. J. (2007). *The continuity of mind* (Vol. xv). New York, NY, US: Oxford University Press.
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, 102(29), 10393–10398. doi: 10.1073/pnas.0503903102
- Stan Development Team. (2015). *Stan: A C++ Library for Probability and Sampling, Version 2.8.0*.
- Stanovich, K. E. (2005). *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago: University of Chicago Press.
- Stanovich, K. E. (2009a). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory. In *In two minds: Dual processes and beyond* (pp. 55–88).
- Stanovich, K. E. (2009b). Rational and Irrational Thought: The Thinking That IQ Tests Miss. *Scientific American Mind*, 20(6), 34–39. doi: 10.1038/scientificamericanmind1109-34
- Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind & Society*, 11(1), 3–13. doi: 10.1007/s11299-011-0093-6
- Stanovich, K. E., & West, R. F. (1999). Discrepancies between normative and descriptive models of decision making and the understanding/acceptance principle. *Cognitive psychology*, 38(3), 349–385. doi: 10.1006/cogp.1998.0700
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and brain sciences*, 23(5), 645–665.

- Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, 94(4), 672–695. doi: 10.1037/0022-3514.94.4.672
- Steegen, S., & De Neys, W. (2012). Belief inhibition in children's reasoning: Memory-based evidence. *Journal of Experimental Child Psychology*, 112(2), 231–242. doi: 10.1016/j.jecp.2012.01.006
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643–662. doi: 10.1037/h0054651
- Stupple, E. J. N., & Ball, L. J. (2008). Belief-logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel-process model. *Thinking & Reasoning*, 14(2), 168–181.
- Sun, R. (1995). Robust reasoning: integrating rule-based and similarity-based reasoning. *Artificial Intelligence*, 75(2), 241–295.
- Sun, R., & Zhang, X. (2006). Accounting for a variety of reasoning data within a cognitive architecture. *Journal of Experimental & Theoretical Artificial Intelligence*, 18(2), 169–191. doi: 10.1080/09528130600557713
- Sussman, A. B., Khemlani, S. S., & Oppenheimer, D. M. (2014). Latent scope bias in categorization. *Journal of Experimental Social Psychology*, 52, 1–8. doi: 10.1016/j.jesp.2013.11.010
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634. doi: 10.1126/science.7777863
- Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, 10(7), 309–318. doi: 10.1016/j.tics.2006.05.009
- Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning*, 20(2), 215–244. doi: 10.1080/13546783.2013.869763
- Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, 63(3), 107–140. doi: 10.1016/j.cogpsych.2011.06.001

- Tomlinson, J. M., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language*, 69(1), 18–35. doi: 10.1016/j.jml.2013.02.003
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, 39(7), 1275–1289. doi: 10.3758/s13421-011-0104-1
- Tower-Richard, S. M., Brunye, T. T., Gagnon, S. A., Mahoney, C. R., & Taylor, H. A. (2012). Abstract spatial concept priming dynamically influences real-world actions. *Cognitive Science*, 3, 361. doi: 10.3389/fpsyg.2012.00361
- Townsend, J. T., & Busemeyer, J. (1995). Dynamic representation of decision-making. In R. F. Port & T. van (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 101–120). Cambridge, MA, US: The MIT Press.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136. doi: 10.1016/0010-0285(80)90005-5
- Trommershäuser, J., Landy, M. S., & Maloney, L. T. (2006). Humans rapidly estimate expected gain in movement planning. *Psychological Science*, 17(11), 981–988.
- Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008a). Decision making, movement planning and statistical decision theory. *Trends in Cognitive Sciences*, 12(8), 291–297. doi: 10.1016/j.tics.2008.04.010
- Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008b). The expected utility of movement. *Neuroeconomics: Decision Making and the Brain*, PW Glimcher, CF Camerer, E. Fehr, and RA Poldrack, eds. (Elsevier, 2009), 95–111.
- Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human perception and performance*, 24(3), 830.
- Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica*, 116(2), 185–203. doi: 10.1016/j.actpsy.2004.01.004
- Tunney, R. J., & Shanks, D. R. (2003). Subjective measures of awareness and implicit cognition. *Memory & Cognition*, 31(7), 1060–1071.

- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207–232. doi: 10.1016/0010-0285(73)90033-9
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131. doi: 10.1126/science.185.4157.1124
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science (New York, N.Y.)*, 211(4481), 453–458.
- Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In *Judgment under uncertainty*. Cambridge University Press.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review*, 108(3), 550–592.
- van der Wel, R. P., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others' beliefs? Evidence from a continuous measure. *Cognition*, 130(1), 128–133. doi: 10.1016/j.cognition.2013.10.004
- Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain sciences*, 21(05), 615–628.
- van Veen, V., & Carter, C. S. (2006). Conflict and Cognitive Control in the Brain. *Current Directions in Psychological Science*, 15(5), 237–240. doi: 10.1111/j.1467-8721.2006.00443.x
- Villejoubert, G. (2009). Are representativeness judgments automatic and rapid? The effect of time pressure on the conjunction fallacy. In *Proceedings of the Annual Meeting of the Cognitive Science society* (Vol. 30, pp. 2980–2985). Amsterdam, Netherlands: Cognitive Science Society.
- Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Harvard university press.
- Wang, J. e. (2011). Pupil dilation and eye tracking. *A handbook of process tracing methods for decision research: A critical review and user's guide*, 185–204.
- Wang, Q., Taylor, H. A., & Brunyé, T. T. (2012). When going the right way is hard to do: Distinct phases of action compatibility in spatial knowledge development. *Acta Psychologica*, 139(3), 449–457. doi: 10.1016/j.actpsy.2012.01.006

- Warton, D. I., & Hui, F. K. C. (2010). The arcsine is asinine: the analysis of proportions in ecology. *Ecology*, 92(1), 3–10. doi: 10.1890/10-0340.1
- Wason, P. C. (1968). Reasoning about a rule. *The Quarterly journal of experimental psychology*, 20(3), 273–281.
- Wason, P. C., & Evans, J. S. B. T. (1975). Dual processes in reasoning? *Cognition*, 3(2), 141–154. doi: 10.1016/0010-0277(74)90017-1
- Weiskrantz, L. (1986). *Blindsight: A Case Study and Implications*. Oxford University Press.
- Westermann, G., & Mareschal, D. (2013). From perceptual to language-mediated categorization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634), 20120391–20120391. doi: 10.1098/rstb.2012.0391
- Wieth, M. B., & Zacks, R. T. (2011). Time of day effects on problem solving: When the non-optimal is optimal. *Thinking & Reasoning*, 17(4), 387–401. doi: 10.1080/13546783.2011.625663
- Willemse, M. C., & Johnson, E. J. (2011). Visiting the decision factory: observing cognition with MouselabWEB and other information acquisition methods. *A handbook of process tracing methods for decision research: A critical review and user's guide*, 21–42.
- Wojnowicz, M. T., Ferguson, M. J., Dale, R., & Spivey, M. J. (2009). The Self-Organization of Explicit Attitudes. *Psychological Science*, 20(11), 1428–1435. doi: 10.1111/j.1467-9280.2009.02448.x
- Wolpert, D. M., & Ghahramani, Z. (2005). Bayes rule in perception, action and cognition.
- Wu, S.-W., Delgado, M. R., & Maloney, L. T. (2009). Economic decision-making compared with an equivalent motor task. *Proceedings of the National Academy of Sciences*, 106(15), 6088–6093.
- Xiao, K., & Yamauchi, T. (2014). Semantic priming revealed by mouse movement trajectories. *Consciousness and Cognition*, 27, 42–52. doi: 10.1016/j.concog.2014.04.004

Yamauchi, T., Kohn, N., & Yu, N.-Y. (2007). Tracking mouse movement in feature inference: Category labels are different from feature labels. *Memory & Cognition*, 35(5), 852–863.

Yarbus, A. L., Haigh, B., & Riggs, L. A. (1967). *Eye movements and vision* (Vol. 2) (No. 5.10). Plenum press New York.

Yu, Z., Wang, F., Wang, D., & Bastin, M. (2012). Beyond Reaction Times: Incorporating Mouse-Tracking Measures into the Implicit Association Test to Examine its Underlying Process. *Social Cognition*, 30(3), 289–306. doi: 10.1521/soco.2012.30.3.289

Zorn, C. (2005). A Solution to Separation in Binary Response Models. *Political Analysis*, 13(2), 157–170. doi: 10.1093/pan/mpi009

Appendices

APPENDIX A

Continuous Attraction and Changes of Mind in Mouse Tracking Studies

Table A.1: Continuous and discrete (reversal/change of mind) conflict effects in mouse tracking studies. While most studies of simple tasks revealed continuous conflict effects, some investigating higher-level processes revealed discrete reversals, where participants move initially towards one option, and then change direction.

Note: Some studies do not explicitly report the type of effect found. Where possible, I have inferred the type of effect from the results reported. Such studies are marked with a ?.

Continuous attraction effects	
Study	Domain
Koop and Johnson (2013)	JDM (preferential choice)
Spivey et al. (2005)	Language (word recognition)
Farmer, Cargill, Hindy, Dale, and Spivey (2007)	Language (sentence parsing)
Freeman et al. (2008)	Categorisation (faces)
Papesh and Goldinger (2012)	Recognition memory
Dale, Roche, Snyder, and McCall (2008)	Learning (paired-associates task)
Q. Wang, Taylor, and Brunyé (2012)	Spatial judgement

Miles, Karpinska, Lumsden, and Macrae (2010)	Temporal judgement (thinking about past/future)
Falke et al. (2013)	Language (A-not-B task)
Hehman, Carpinella, Johnson, Leitner, and Freeman (2014)	Categorisation (faces)
Hehman, Ingbretnsen, and Freeman (2014)	Categorisation (faces)
Kieslich and Hilbig (2014)	Reasoning (Social dilemmas)
Marghetis, Núñez, and Bergen (2014)	Maths (Arithmetic)
O'Hora, Dale, Piiroinen, and Connolly (2013)	Learning (stimuli paired with rewards)
Quinton, Volpi, Barca, and Pezzulo (2013)	Simple judgement (categorising line drawings)
Scherbaum et al. (2010)	Simple judgement (Simon Task)
Yu, Wang, Wang, and Bastin (2012)	Simple judgement (Implicit Association Task)
van der Wel, Sebanz, and Knoblich (2014)	Simple judgement (False Belief Task)
Dale et al. (2007)?	Categorisation (animals)
Koop and Johnson (2011)?	JDM (Iowa Gambling Task)
Xiao and Yamauchi (2014)?	Primed judgement

Discrete attraction effects

Dale and Duran (2011)	Language (negated sentences)
Tomlinson et al. (2013)	Language (scalar implicature)
Hindy and Spivey (2008)	Simple judgement (task switching — Wisconsin Card Sorting Task)
Xiao and Yamauchi (2014)?	Primed judgement
Duran, Dale, and McNamara (2010)?	High-order cognition (Deception)

McKinstry et al. (2008)	High-order cognition (True/false judgements)
Barca and Pezzulo (2015) ¹	Lexical decision
Koop and Johnson (2013) ²	JDM (Lottery choice)
Freeman, Stolier, Ingbretnsen, and Hehman (2014)	Categorisation (faces)

¹Discrete for pseudowords, continuous otherwise

²Discrete when choosing gambles

APPENDIX B

Classifying Cursor Trajectories

In Experiments 1–5, I classified cursor trajectories as either direct movements to a response, or reversals/changes of mind. To do this, I fit two-sample finite mixture models to the Maximum Deviation data from each experiment, and from these derived Maximum Deviation cut-offs, above which trials could be classed as reversals.

EXPERIMENT 1

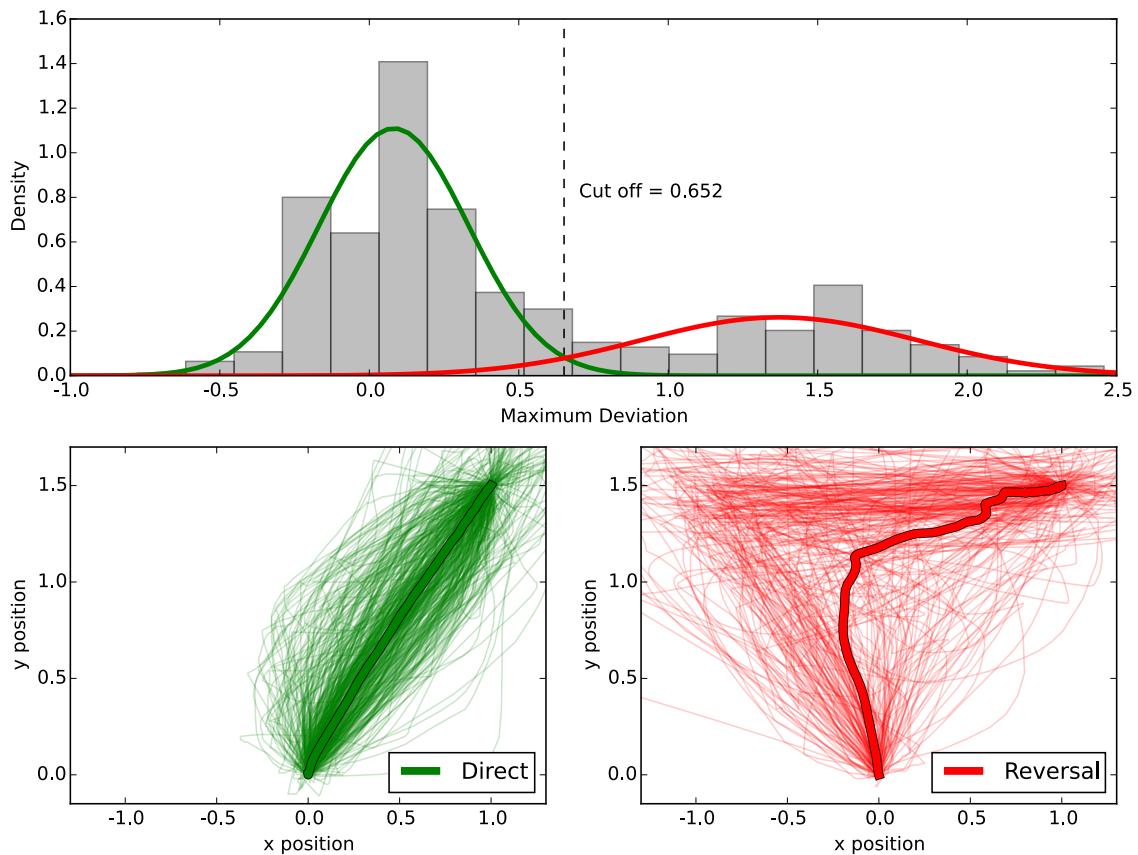


Figure B.1: Trajectories classed as direct and reversals in Experiment 1.

Bimodality Coefficient = 0.689. Hartigan's' D = 0.027, N = 580, p = 0.006.

Table B.1: Mean and SD, and relative proportions, of the Maximum Deviation for direct and reversal trials, in Experiment 1

	Direct	Reversal
Mean	0.08	1.37
Standard deviation	0.25	0.46
Proportion	70%	30%

EXPERIMENT 2

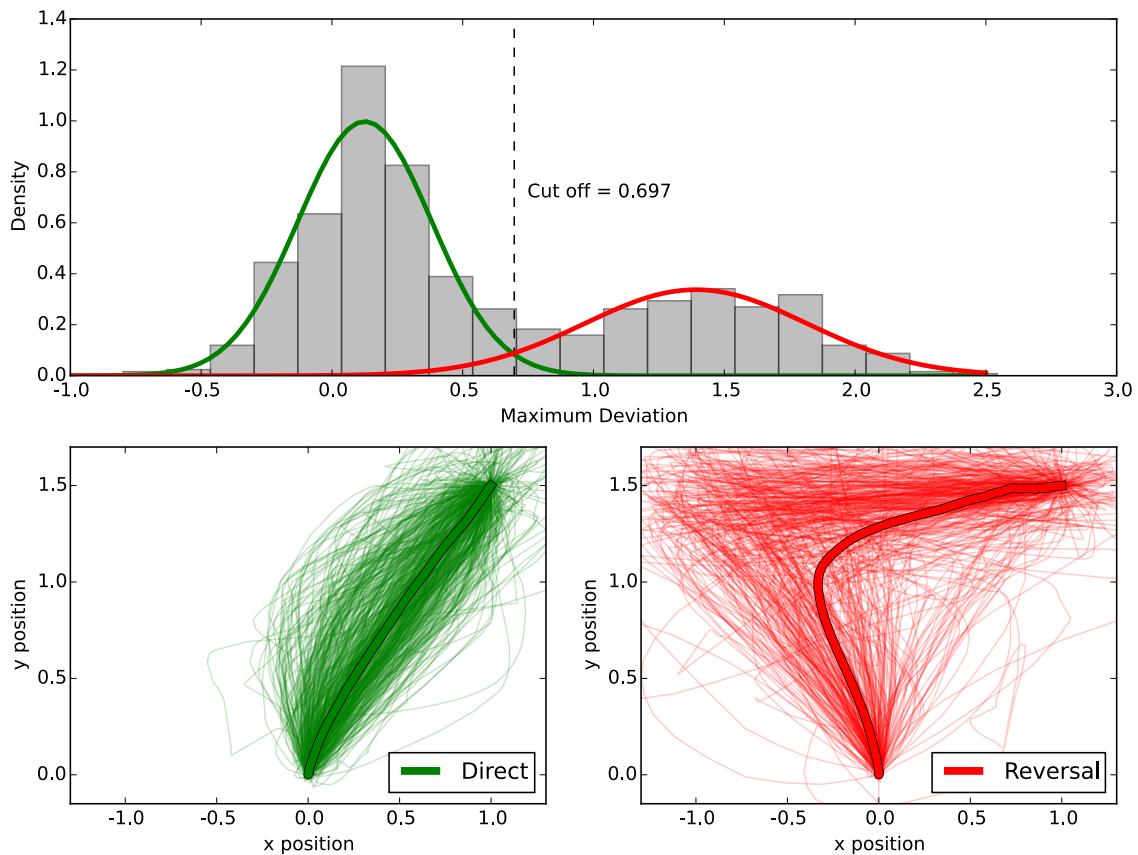


Figure B.2: Trajectories classed as direct and reversals in Experiment 2.

Bimodality Coefficient = 0.65. Hartigan's' D = 0.021, N = 754, p = .0182.

Table B.2: Mean and SD, and relative proportions, of the Maximum Deviation for direct and reversal trials, in Experiment 2

	Direct	Reversal
Mean	0.13	1.39
Standard deviation	0.26	0.42
Proportion	64%	36%

EXPERIMENT 3

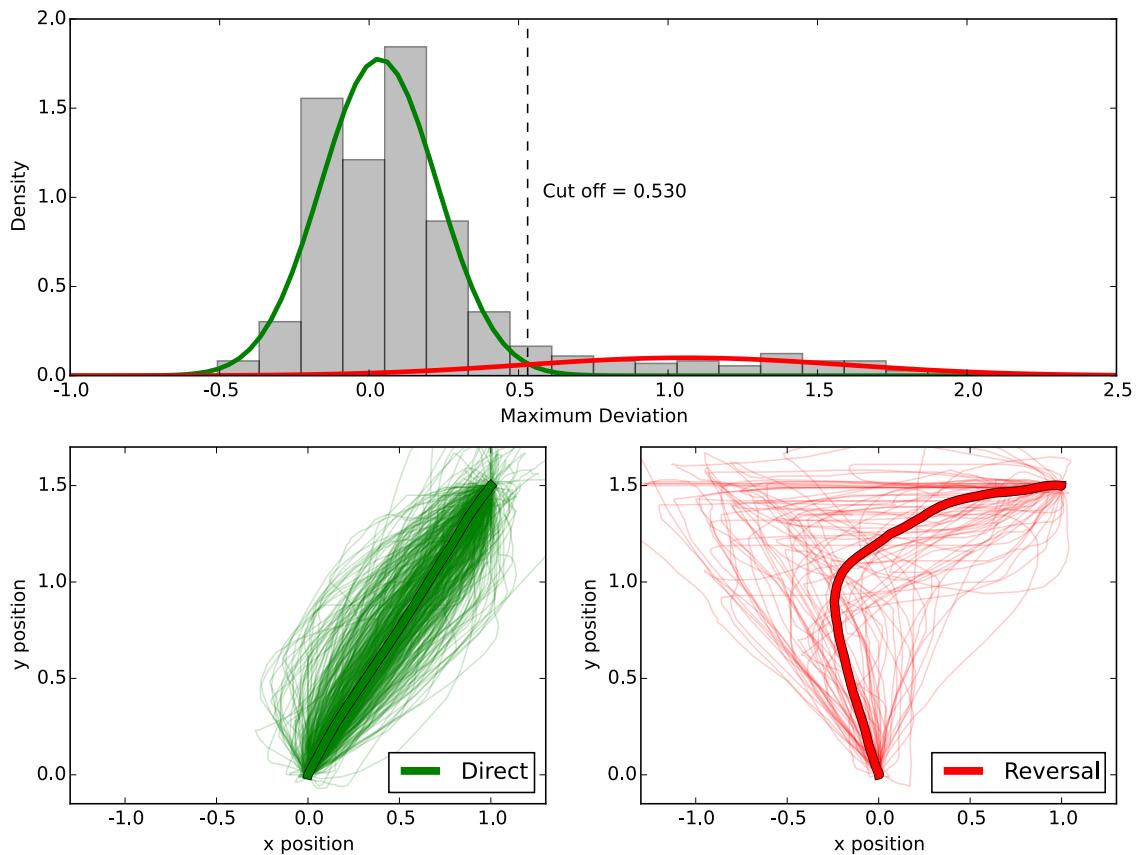


Figure B.3: Trajectories classed as direct and reversals in Experiment 3 (after data was excluded from stimulus sets for which participants performed poorly on the post-test check; see Chapter 4).

Bimodality Coefficient = 0.705. Hartigan's' D = 0.042, N = 520, p < .0001.

Table B.3: Mean and SD, and relative proportions, of the Maximum Deviation for direct and reversal trials, in Experiment 3

	Direct	Reversal
Mean	0.03	1.06
Standard deviation	0.19	0.54
Proportion	86%	14%

EXPERIMENT 4

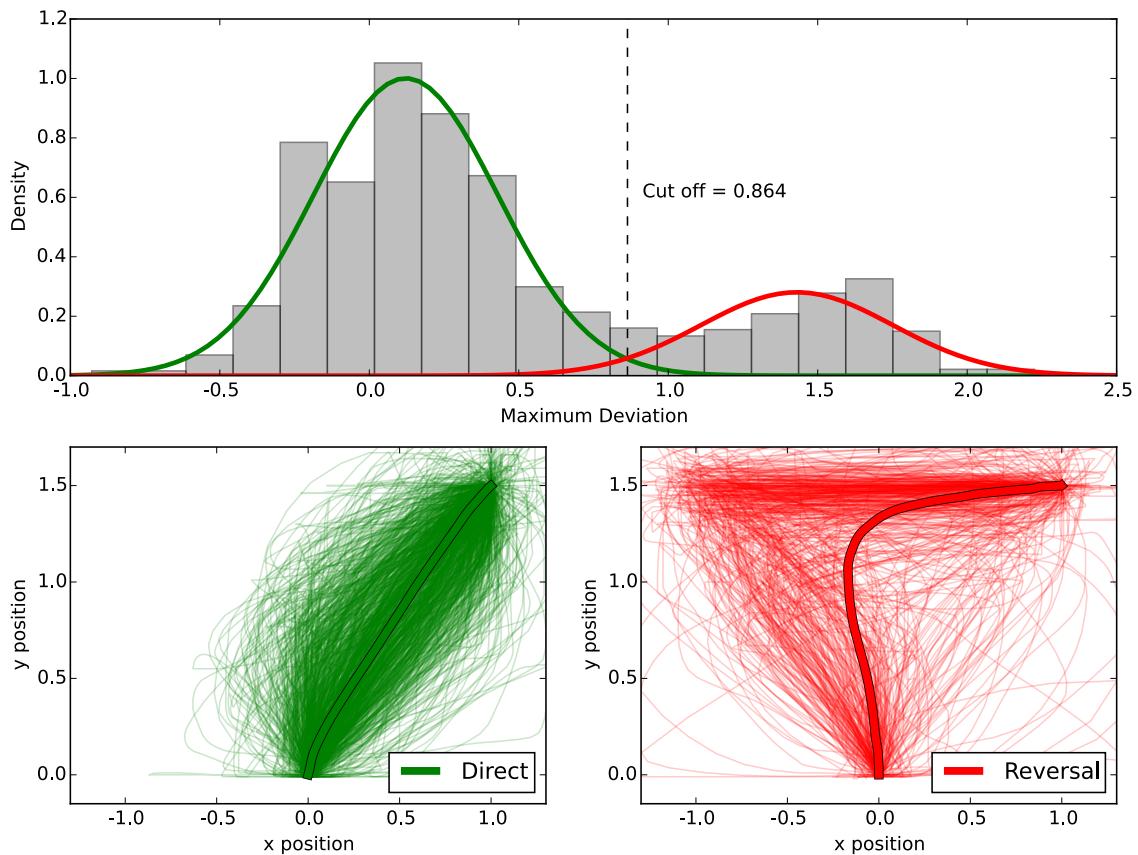


Figure B.4: Trajectories classed as direct and reversals in Experiment 4.

Bimodality Coefficient = 0.636. Hartigan's' D = 0.025, N = 1188, p < .0001.

Table B.4: Mean and SD, and relative proportions, of the Maximum Deviation for direct and reversal trials, in Experiment 4

	Direct	Reversal
Mean	0.12	1.43
Standard deviation	0.31	0.32
Proportion	77%	23%

EXPERIMENT 5

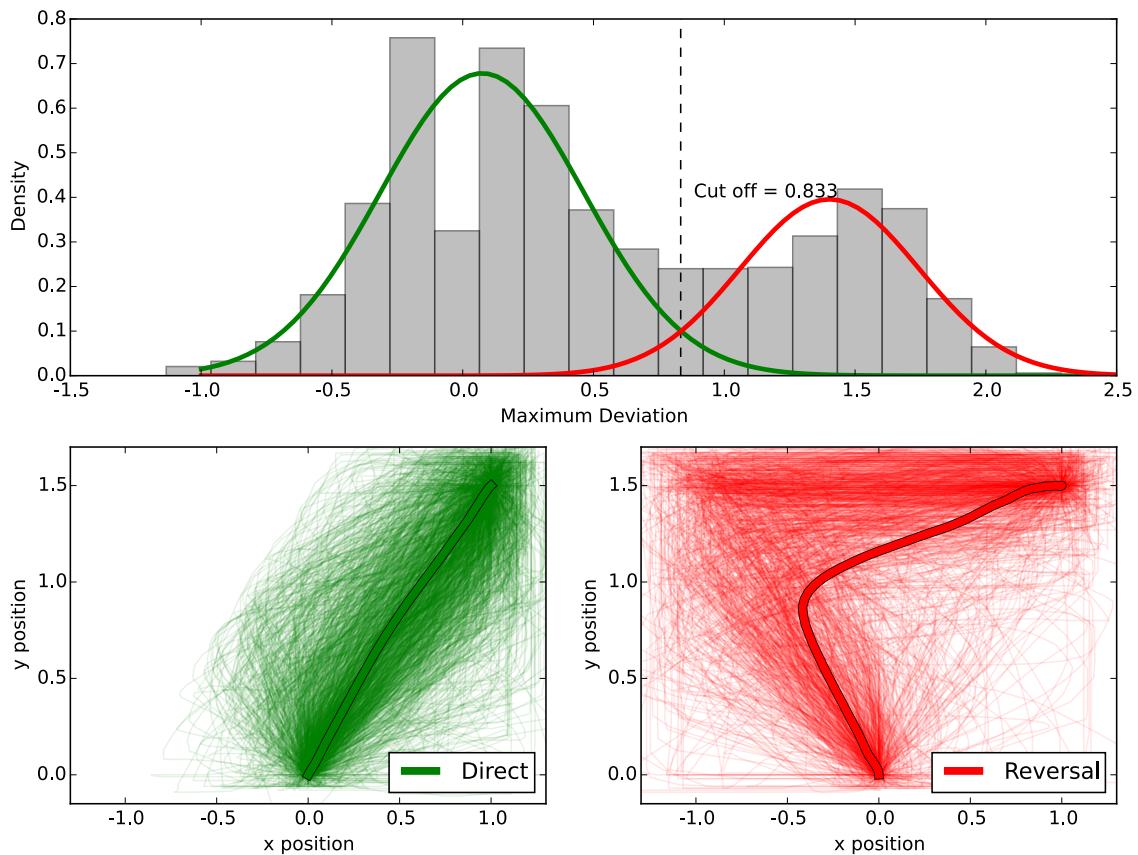


Figure B.5: Trajectories classed as direct and reversals in Experiment 5.

Bimodality Coefficient = 0.562. Hartigan's' D = 0.024, N = 1998, p < .0001.

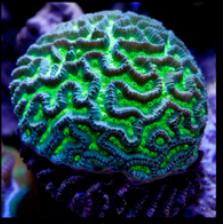
Table B.5: Mean and SD, and relative proportions, of the Maximum Deviation for direct and reversal trials, in Experiment 1

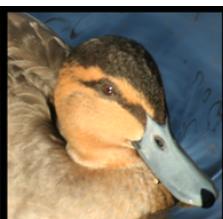
	Direct	Reversal
Mean	0.07	1.4
Standard deviation	0.39	0.34
Proportion	66%	34%

APPENDIX C

Stimuli from Experiment 1

Table C.1: Stimulus images used in Experiment 1.

Set	Base species	Correct option	Conflict foil	Control foil
1				
2				
3				
4				

Set	Base species	Correct option	Conflict foil	Control foil
5				
6				
7				
8				
9				
10				

APPENDIX D

Post test scores for stimuli used in Experiment 1

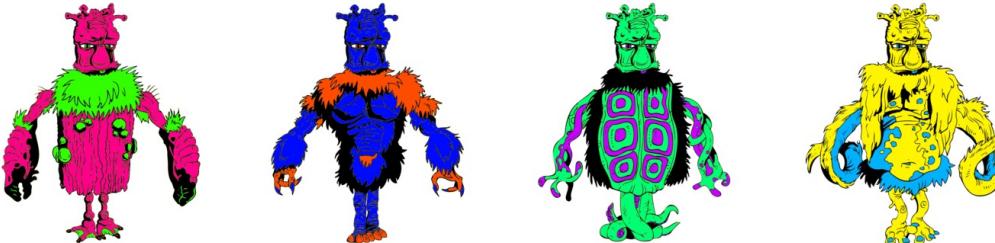
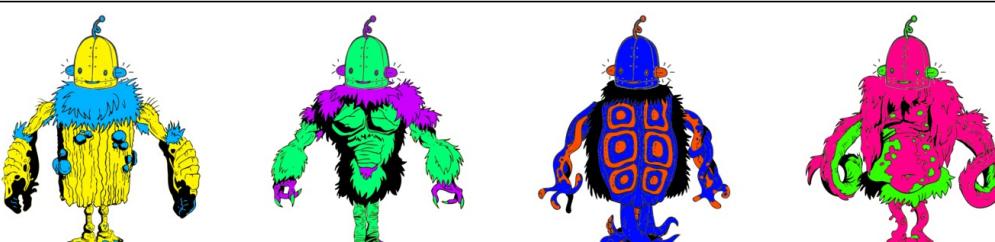
Table D.1: Accuracy on the post-test for stimuli pairs used in Experiment 1. Scores greater than 67.8% are significant at the $p < .01$ level.

Base	Correct response (Accuracy)	Conflict foil (Accuracy)
Bird	Flamingo (92%)	Bat (85%)
Dandelions	Roses (98%)	Coral (flower-like) (82%)
Insect (leaf-like)	Insect (Ladybird) (93%)	Plant (leafy) (100%)
Hedgehog	Dog (87%)	Pinecone (100%)
Dinosaur (Triceratops)	Dinosaur (Sauropod) (91%)	Rhino (68%)
Tuna	Fish (yellow) (95%)	Dolphin (71%)
Lime	Strawberry (100%)	Fish (green) (100%)
Duck	Eagle (92%)	Platypus (73%)
Leopard	Deer (95%)	Leopard Gecko (97%)
Fish (stripey)	Goldfish (100%)	Zebra (95%)

APPENDIX E

Stimuli from Experiment 2

Table E.1: Stimuli used in Experiment 3: animals called Flurps, and robots called Floobits. There were four possible colourations, and four possible body shapes, and each colouration and body shape was equally likely to occur in each category. Categories could be recognised by inspecting their heads.

Flurps

Floobits


APPENDIX F

Properties used in Experiment 2

Properties were presented in the form “This one X. Who else do you think X”.

Properties specific to Flurps (animals)

- This one breathes.
- This one can have babies.
- This one has a heart inside.
- This one has a mummy.
- This one has bones inside.
- This one needs water.
- This one can climb trees.
- This one tries to stay warm.

Properties specific to Floobits (robots)

- This one can be turned off.
- This one can break.
- This one has batteries inside.
- This one has wires inside.
- This one was made by people.
- This one was sold in a store.
- This one is cold touch.

- This one doesn't sleep.

Generic properties

- This one can make a zevy sound.
- This one has a very sticky toma.
- This one has zimmer inside.
- This one is used for derriping.
- This one needs tiddles to make it move.
- This one uses danner.
- This one can help yippets.
- This one goes outside in the winter.
- This one has a part inside called a cece.
- This one has blickets inside of it.
- This one has grumpets that make it strong.
- This one is good for kertling.
- This one is found on farms.
- This one can be very old.
- This one feels yinty.
- This one lacks ombelots.

APPENDIX G

Association ratings for stimuli used in Experiment 3

Table G.1: Experimental stimulus sets from Experiment 3. Values in parentheses denote mean strength of association between that species and the base species, on a scale from 1 to 9, as collected by Crisp-Bright (2010).

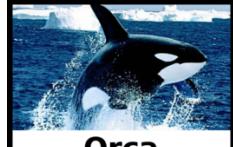
Base	Correct response	Conflict foil	Control foil
Acorns	Lychees (3.7)	Squirrels (7.9)	Seals
Butterflies	Locusts (5.5)	Flowers (7.6)	Seaweed
Ants	Dragonflies (4.5)	Anteaters (7.6)	Moose
Grass	Palm Trees (5.0)	Sheep (7.5)	Hyenas
Carrots	Bamboo (3.1)	Rabbits (7.7)	Tigers
Shepherds	Bus Drivers (3.3)	Sheep (8.2)	Porpoises
Orca Whales	Cows (2.2)	Cod (5.3)	Pigeons
Snails	Octopuses (2.8)	Hedgehogs (4.3)	Sloths
Salmon	Goldfish (5.0)	Grizzly Bears (7.5)	Hedgehogs
Monkeys	Seals (2.5)	Peanuts (6.8)	Almonds
Bananas	Tulips (2.4)	Monkeys (7.5)	Sea Lions
Mice	Goats (2.7)	Wheat (4.4)	Bamboo
Penguins	Chickens (3.0)	Orca Whales (5.9)	Dogs
Dolphins	Llamas (2.0)	Cod (7.3)	Parrots

APPENDIX H

Stimuli used in Experiment 3

Table H.1: Stimulus images used in Experiment 3.

Set	Base species	Correct option	Conflict foil	Control foil
1	 Acorns	 Lychees	 Squirrels	 Seals
2	 Butterflies	 Locusts	 Flowers	 Seaweed
3	 Ants	 Dragonflies	 Anteaters	 Moose
4	 Grass	 Palm Trees	 Sheep	 Hyenas
5	 Carrots	 Bamboo	 Rabbits	 Tigers

Set	Base species	Correct option	Conflict foil	Control foil
6	 Shepherds	 Bus Drivers	 Sheep	 Porpoises
7	 Orca	 Cows	 Cod	 Pigeons
8	 Snails	 Octopuses	 Hedgehogs	 Sloths
9	 Salmon	 Goldfish	 Bears	 Hedgehogs
10	 Monkeys	 Seals	 Peanuts	 Almonds
11	 Bananas	 Tulips	 Monkeys	 Sealions
12	 Mice	 Goats	 Wheat	 Bamboo
13	 Penguins	 Chickens	 Orca	 Dogs
14	 Dolphins	 Llamas	 Cod	 Parrots

APPENDIX I

Post test performance for stimuli used in Experiment 3

Table I.1: Participants' accuracy on the post-test check in Experiment 3, for each stimulus set. Participants were significantly above chance at correctly identifying that the base and correct response species belonged to the same taxonomic group for seven of the fourteen stimulus sets (left column), and at correctly identifying that the base and foil species did not belong to the same group for eleven of fourteen sets.

Base	Correct species	Foil species
Dolphins	40%	65%
Monkeys	40.3%	97.2%**
Snails	42.6%	89.7%**
Bananas	52.9%	97.1%**
Penguins	53%	57.6%
Mice	56.5%	97.1%**
Orca	60%	66.2%
Acorns	81.1%**	98.6%**
Butterflies	85.9%**	90.1%**
Carrots	90.1%**	100%**
Ants	93%**	93%**
Grass	100%**	97.1%**
Salmon	100%**	98.6%**
Shepherds	100%**	82.6%**

Note: ** $p < .01$.

APPENDIX J

Stimuli used in Experiment 4

Table J.1: Stimulus images for experimental trials, Experiment 4.

	Base species	Correct option	Weak foil	Moderate foil	Strong foil
1	 Acorns	 Strawberries	 Cats	 Chipmunks	 Squirrels
2	 Ants	 Butterflies	 Dogs	 Aardvark	 Anteaters
3	 Bears	 Hamsters	 Clownfish	 Trout	 Salmon
4	 Carrots	 Bamboo	 Foxes	 Horses	 Rabbits

5



Grass



Oranges



Mosquitos

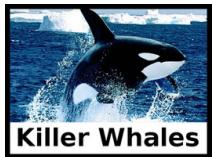


Beetles



Crickets

6



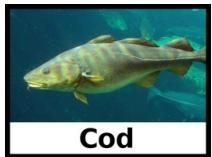
Killer Whales



Hippos



Goldfish



Cod



Sharks

7



Mice



Camels



Emus



Kestrels

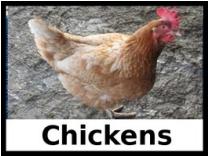


Owls

8



Owls



Chickens



Llamas



Foxes



Bats

9



Penguins



Robins



Goats



Arctic Wolves



Polar Bears

Table J.2: Stimulus sets for experimental trials, Experiment 4.

Base	Correct species	Weak foil	Moderate foil	Strong foil
Acorns	Strawberries	Cats	Chipmunks	Squirrels
Ants	Butterflies	Dogs	Aardvarks	Anteaters
Bears	Hamsters	Clownfish	Trout	Salmon
Carrots	Bamboo	Foxes	Horses	Rabbits
Grass	Oranges	Mosquitoes	Beetles	Crickets
Killer Whales	Hippos	Goldfish	Cod	Sharks
Mice	Camels	Emus	Kestrels	Owls
Owls	Chickens	Llamas	Foxes	Bats
Penguins	Robins	Goats	Arctic Wolves	Polar Bears

APPENDIX K

Post test performance for stimuli used in Experiment 4

Table K.1: Performance on the post-test for each stimulus set from Experiment 4. Participants successfully identified that the correct response species belonged to the same taxonomic group as the base, and that the two foil species presented did not, on at least 70% of trials ($p's < .01$).

Base	Correct Species	Foil #1	Foil #2
Acorns	Strawberries (95%)	Chipmunks (98%)	Squirrels (95%)
Ants	Butterflies (93%)	Aardvarks (91%)	Anteaters (86%)
Bears	Hamsters (91%)	Salmon (89%)	Trout (93%)
Carrots	Bamboo (89%)	Horses (84%)	Rabbits (95%)
Grass	Oranges (91%)	Crickets (98%)	Beetles (95%)
Killer Whales	Hippos (70%)	Cod (91%)	Sharks (100%)
Mice	Camels (86%)	Kestrels (95%)	Owls (95%)
Owls	Chickens (84%)	Bats (93%)	Foxes (75%)
Penguins	Robins (84%)	Polar bears (91%)	Arctic wolves (91%)

Note: * $p < .01$; All other $p's < .0001$.

APPENDIX L

Bayesian model priors from Experiment 5

$$\begin{aligned} \log(rt_{(p, d)}) &= (A + \alpha_p + \alpha_d) + Condition * B + \epsilon \\ A &\sim Uniform(0, 8.5) \\ B &\sim Normal(-.69, .69) \\ \alpha_p &\sim Normal(0, \sigma_{\alpha_p}) \\ \alpha_d &\sim Normal(0, \sigma_{\alpha_d}) \\ \sigma_{\alpha_p}^2 &\sim Cauchy(0, 10); \sigma_{\alpha_p}^2 > 0 \\ \sigma_{\alpha_D}^2 &\sim Cauchy(0, 10); \sigma_{\alpha_D}^2 > 0 \end{aligned} \tag{L.1}$$

This model predicts log response time (rt) for participants p , reasoning about description d . The model parameters are the overall intercept term A , representing the log of the average response time when the base rates agreed with the description, participant p 's offset from this intercept, α_p , and description d 's offset, α_d — both of which were normally distributed with a mean of 0, the regression weight B , representing the log of the percentage increase in rt when the base rate disagrees, and the error term ϵ . *Condition* is 0 for trials where the cues agree, and 1 where they disagree. A uniform prior was set on A so that all baseline *rts* from 1 to 5,000 msec were equally likely. B was given a broad normal prior with mean 0 and SD .69, meaning that, a priori, I was 95% certain that response times would be between 5 times slower and 5 times faster than those when the cues agreed. The variance parameters for the by-participant and by-description differences in the intercept had uninformative half-Cauchy distributions, indicating that I could not specify in advance how much variance there would be (A. Gelman, 2006).

APPENDIX M

CRT Reasoning Problems from Experiment 6

Table M.1: Stimuli used in Experiment 6. Items 1–3 were adapted from Frederick (2005). Items 4–8 were adapted from Primi et al. (2015). *Note.* Percentages in parentheses show the proportion of participants who gave each response.

	Conflict	No-conflict
1	A bat and a ball together costs £1.10. A bat costs £1 more than a ball. How much does a ball cost?	A bat and a ball together costs £1.05. A bat costs £1. How much does a ball cost?
	Correct response: 5p (15%) Heuristic response: 10p (83%) Foil response: 15p (0%) Foil response: 90p (2%)	Correct response: 5p (97%) Foil response: 10p (0%) Foil response: 15p (1%) Foil response: 90p (1%)
2	It takes 5 machines 5 minutes to make 5 widgets. How many minutes would it take 100 machines to make 100 widgets?	It takes a machine 5 minutes to make 5 widgets. How many minutes would it take the machines to make 100 widgets?
	Correct response: 5 (24%) Heuristic response: 100 (69%) Foil response: 50 (4%) Foil response: 10 (3%)	Correct response: 100 (83%) Foil response: 5 (2%) Foil response: 50 (13%) Foil response: 10 (2%)

Conflict	No-conflict
<p>3 In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how many days would it take for the patch to cover half of the lake?</p> <p>Correct response: 47 (25%) Heuristic response: 24 (59%) Foil response: 12 (15%) Foil response: 2 (2%)</p>	<p>In a lake, there is a patch of lily pads. Every day, the patch grows by $10m^2$. If it takes 48 days for the patch to cover the $150m^2$, how many days would it take for the patch to cover $140m^2$?</p> <p>Correct response: 47 (79%) Foil response: 24 (17%) Foil response: 12 (4%) Foil response: 2 (0%)</p>
<p>4 If you flipped a fair coin twice, what is the probability that it would land 'Heads' at least once?</p> <p>Correct response: 75% (4%) Heuristic response: 50% (84%) Foil response: 25% (11%) Foil response: 100% (1%)</p>	<p>If you flipped a fair coin twice, what is the probability that it would land 'Heads' exactly once?</p> <p>Correct response: 25% (68%) Foil response: 50% (26%) Foil response: 75% (6%) Foil response: 100% (0%)</p>
<p>5 If 3 elves can wrap 3 toys in 1 hour, how many elves are needed to wrap 6 toys in 2 hours?</p> <p>Correct response: 3 (73%) Heuristic response: 6 (22%) Foil response: 1 (2%) Foil response: 12 (4%)</p>	<p>If 3 elves can wrap 3 toys in 1 hour, how many toys could 6 elves wrap in half an hour?</p> <p>Correct response: 3 (71%) Foil response: 6 (20%) Foil response: 1 (1%) Foil response: 12 (8%)</p>
<p>6 Ellen and Kim are running around a track. They run equally fast but Ellen started later. When Ellen has run 5 laps, Kim has run 10 laps. When Ellen has run 10 laps, how many has Kim run?</p> <p>Correct response: 15 (73%) Heuristic response: 20 (27%) Foil response: 5 (0%) Foil response: 19 (0%)</p>	<p>Ellen and Kim are running around a track. They started at the same time, but Kim is twice as fast as Ellen. When Ellen has run 5 laps, Kim has run 10 laps. When Ellen has run 10 laps, how many has Kim run?</p> <p>Correct response: 20 (98%) Foil response: 15 (2%) Foil response: 5 (0%) Foil response: 19 (0%)</p>

	Conflict	No-conflict
7	<p>Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are there in the class?</p> <p>Correct response: 29 (26%) Heuristic response: 30 (72%) Foil response: 40 (2%) Foil response: 5 (0%)</p>	<p>Jerry received both the 2nd highest and the 2nd lowest mark in the class. How many students are there in the class?</p> <p>Correct response: 3 (79%) Foil response: 2 (13%) Foil response: 5 (8%) Foil response: 10 (0%)</p>
8	<p>In an athletics team tall members tend to win three times as many medals than short members. This year the team has won 60 medals so far. How many of these have been won by short athletes?</p> <p>Correct response: 15 (44%) Heuristic response: 20 (52%) Foil response: 30 (1%) Foil response: 50 (3%)</p>	<p>In an athletics team tall members tend to win twice as many medals than short members. This year the team has won 60 medals so far. How many of these have been won by short athletes?</p> <p>Correct response: 20 (58%) Foil response: 15 (12%) Foil response: 30 (27%) Foil response: 50 (4%)</p>